Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Contributions to Time Series Factor Modeling: Model Averaging and Bias Correction

Autor:

Guadalupe Bastos

Directores:

Dr. Andrés M. Alonso

Dra. Carolina García-Martos

DEPARTAMENTO DE ESTADISTICA

Getafe, marzo 2017

Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Contributions to Time Series Factor Modeling: Model Averaging and Bias Correction

Autor: *Guadalupe Bastos*

Directores: **Dr. Andrés M. Alonso**

Dra. Carolina García-Martos

Firma del Tribunal Calificador:

Firma

Presidente: Daniel Peña Sánchez de Rivera

Vocal: Julio Rodríguez Puerta

Secretario: Juan Miguel Morales González

Calificación:

Getafe, de de

UNIVERSIDAD CARLOS III DE MADRID

DEPARTMENT OF STATISTICS

PHD IN BUSINESS ADMINISTRATION AND QUANTITATIVE METHODS

DOCTORAL THESIS

Contributions to Time Series Factor Modeling: Model Averaging and Bias Correction

Author:
Guadalupe Bastos

Advisors:
Dr. Andrés M. Alonso
Dra. Carolina García-Martos

March 24, 2017

Contents

To my grandparents: Tatá, Nonni and Vilma.

Acknowledgements

I would like to extend my gratitude to my advisors Andrés Alonso and Carolina García-Martos. Thank you for teaching me, for your patience, inspiration, and for your invaluable support and guidance during these years. I am in your debt.

I would also like to thank the members of the Business Administration and Quantitative Methods program. I have learned from my teachers as well as from my peers. In particular, thanks to Professor Michael Wiper for reviewing the use of English in my first published paper; and thanks to my fellow Matias for studying with me and helping with paperwork.

I am most thankful to my husband for his encouragement through this journey and for his unconditional support. Thanks to my parents for their love, cheering and support in the path to better myself. And thanks to my brothers for always believing in me.

Last, I am grateful for the financial support provided for the Master by the Statistics Department of the University, without which none of this would have been possible.

List of Figures

List of Tables

List of Acronyms

AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
ANOVA	Analysis of Variance
AR	AutoRegressive
ARIMA	AutoRegressive Integrated Moving Average
ARX	AutoRegressive with Exogenous Variable
BC	Bias Corrected (AutoRegressive coefficients)
BIC	Bayesian Information Criterion
BMA	Bayesian Model Averaging
CSS	Conditional Sum of Squares
DFA	Dynamic Factor Analysis
DFM	Dynamic Factor Model
EM	Expectation Maximization (algorithm)
FM	Factor Model
GARCH	Generalized AutoRegressive Conditional Hetereoskedastic
GME	Gestore dei Mercati Energetici (The Energy Market Operator)
GWh	Gigawatts per hour
$\boldsymbol{IC_3}$	(Bai and Ng) Information Criterion 3
ICA	Independent Component Analysis
IPI	Industrial Production Index
MA	Moving Average

MAE	Mean Absolute Error
MAPE	Mean Average Percentage Error
MedAE	Median Absolute Error
MIBEL	Mercado Ibérico de la Electricidad (The Iberian Electricity Market)
MSE	Mean Squared Error
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
MWh	Megawatts per hour
OLS	Ordinary Least Squares
OMIE	Operador del Mercado Ibérico de Energía (The Iberian Market Operator)
PC	Principal Components
PCA	Principal Components Analysis
RelMAE	Relative Mean Absolute Error
RF	Roy Fuller (AutoRegressive coefficients)
sARIMA	Seasonal AutoRegressive Integrated Moving Average
SeaDFA	Seasonal Dynamic Factor Analysis
SVD	Singular Value Decomposition
VAR	Vector AutoRegressive
VARIMA	Vector AutoRegressive Integrated Moving Average

List of Notations (Symbols)

$\boldsymbol{A}^T$	Transpose of a matrix $\boldsymbol{A}$
a^j_{Fi}	$j-th$ Root for the characteristic equation of factor F_i
C_m	Average prediction interval coverage rate
CQ_m	C_m and L_m combined measure
χ^2	Chi Square distribution
$\boldsymbol{\varepsilon}_t$	Vector of idiosyncratic or specific factors at time t
$\boldsymbol{e}$	eigenvectors
$\boldsymbol{F}_t$	Vector of underlying common factors at time t
$\hat{\boldsymbol{F}}$	$T \times r$ matrix of estimated common factors
$\boldsymbol{f}_t$	Estimated vector of underlying common factors at time t
h	Forecasting horizon
$\boldsymbol{I}$	Identity matrix
K	Number of alternative models
L	Lag operator
L_m	Average prediction interval length
L_t	Theoretical prediction interval length
λ	eigenvalues
N	Number of time series or variables
$\boldsymbol{\Omega}$	$N \times R$ Matrix of true weights $\boldsymbol{\Omega} = [\omega_{ij}]$, where $\omega_{ij} \in \mathbb{R}$ for $1 \leq i, j \leq R$
p	AutoRegressive order

$\hat{p}$	Estimated Order of AutoRegressive model
R	True Number of common factors
r	Estimated Number of common factors
$\mathbf{\Sigma}_Y$	Covariance matrix for a matrix $\boldsymbol{Y}$, $\mathbf{\Sigma} = [\sigma_{ij}]_{1 \leq i,j \leq N}$
T	Number of observations in the time dimension, length of historic dataset
$\hat{\tau}$	Statistic for unit root test
$\boldsymbol{y}_t$	Vector of N centred observed time series
$\boldsymbol{Y}$	Sample data matrix $\boldsymbol{Y} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_T)^T$, where $y_{it} \in \mathbb{R}$
$\boldsymbol{S}$	Sample covariance matrix $\boldsymbol{S} = [s_{ij}]_{1 \leq i,j \leq p}$

Summary

Given the increasing availability of data and the evolution of computation, there is a growing body of theory and applications taking advantage of multivariate datasets. By including many variables in the analysis (even hundreds), we can exploit more complete information as well as improve the robustness of the estimators obtained (Stock and Watson, 2006).

In this dissertation, we work with multivariate time series. With the aim of forecasting vectors of time series, well known approaches in time series literature are AutoRegressive Integrated Moving Average (ARIMA, working with each variable independently) and Vector AutoRegressive Integrated Moving Average (VARIMA, working with a few variables at a time) models.

However, when there are many interrelated series, these approaches either fail to include interconnections, or rapidly present methodological constraints when more than few series are considered simultaneously. ARIMA models fail to account for the variables' mutual influence; while VARIMA models can present an overwhelming complexity and possibly unfeasibility when the number of time series is large. As a consequence of these limitations, a large portion of research has focused on dimensionality reduction techniques. These allow to exploit the relation between the series, as well as their dynamic nature, and have the virtue of employing a reduced number of

parameters, thus circumventing the "curse of dimensionality" often associated with multivariate data. In particular, in this thesis we focus on Factor Models (FM).

The purpose of this dissertation is to improve the forecasts of high-dimensional vectors of time series. Even with the expansion of research in this area, many issues are still open. We explore some of the questions that arise with the use of FM. In particular, we take an alternative approach for decisions regarding the number of underlying common factors and what models these factors follow (Chapter 2). On the other hand, even if the factors are accurately estimated, and their estimation taken as observations in posterior calculations, it is not unusual to deal with bias of the estimates of the parameters for the model of the common factors, especially when the sample size is small. Therefore, in Chapter 3 we work with techniques to correct this bias and deal strictly with the effect of the time dimension, T.

Our discussion focuses on statistical and econometrical developments that have been employed to address questions in the context of economics, business, and demographics. For empirical examples we work with electricity prices and industrial production indexes of European countries.

The rest of the dissertation is organized as follows. In Chapter 1 we introduce the theory and challenges related to the estimation of factor models. We address the reasons for employing dimensionality reduction, the techniques that may be employed in the estimation of FM, the alternative criteria for selecting the number of unobservable common factors, and what models are usually employed for the common factors.

In Chapter 2 we work with the combination of forecasts, motivated by the unsolved issues of selecting a number of common factors and selecting a model for each of them. Instead of applying a particular criterion, we estimate several specifications, with alternative numbers of common factors and alternative models for them. Afterwards, we evaluate the performance of five easy to apply combination techniques in an application to electricity prices of the Iberian and Italian markets. Even though the improvements that result from the combinations are not big, it must be acknowledged that they are maintained during a long period of time and are statistically significant for some of the combinations considered, according to an Analysis of Variance (ANOVA).

In Chapter 3 we propose two alternative techniques to correct the bias in AR models for the estimated common factors, specifically when these are highly persistent and the sample has a small time dimension (T is small). These are the Bootstrap Bias-Correction methodology (Clements and Kim, 2007) and Roy-Fuller's methodology (Roy and Fuller, 2001). Though not originally intended for factor models, these techniques contribute to reduce the bias of AR coefficients, and by employing Monte Carlo simulations we show that the improvement in the factors' coefficients produces more accurate forecasts. We obtain forecasting intervals, and present results in terms of coverage and interval length. We apply these extensions to data of the Industrial Production Index (IPI) for a group of European countries.

Finally, in Chapter 4 conclusions and further lines of research are summarized.

Chapter 1

Introduction to the Factor Model

1.1 The Factor Model

Multivariate time series are datasets containing several interrelated time series. When the number of series, N, is small, standard approaches are VAR or VARIMA modeling. However, nowadays the large amount of information available results in a larger value of N, which implies different modeling approaches. One of them, the Factor Model (FM), is a dimensionality reduction technique that works on the premise that a large portion of the N series' variation can be explained by a small number, R, of unobserved factors. One advantage of these factors is that they could be employed to forecast variables of interest, using a parsimonious model. Also, a one factor model has been shown to outperform univariate ARIMA as well as pooled forecasts under reasonable assumptions (Peña and Poncela, 2004, they also extend the conclusion to multifactor models).

In other words, dimensionality reduction techniques capture complex relations between the time series included in the study, while keeping the number of parameters to estimate manageable[1]. On the contrary, depending on the number of variables involved, from the computational perspective estimation of typical time series models such as VARIMA[2] or ARIMA[3] may turn out to be inefficient or even unfeasible, not to mention other problems such as highly correlated coefficients (Peña and Box, 1987).

Consider an N-dimensional vector of centred observed time series, $\boldsymbol{y}_t$, where $t = 1, ..., T$ and T is the historical length. According to the theory of factor models, the variation of $\boldsymbol{y}_t$ can be decomposed into variation due to a few common latent factors, $\boldsymbol{F}_t$ and variation due to specific or idiosyncratic components, $\boldsymbol{\varepsilon}_t$, both unobservable. The model in vector form is the following

$$\boldsymbol{y}_t = \boldsymbol{\Omega}\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \tag{1.1}$$

where $\boldsymbol{F}_t$ has dimension $R \times 1$, $R \ll N$, and it captures the part of the behavior shared by the series. Alternatively, we may want only to represent the system more simply (a particular rotation of the series) without dimensionality reduction, $R = N$ (Peña and Box, 1987). $\boldsymbol{\varepsilon}_t$, on the other hand, is a vector of dimension $N \times 1$ and it captures the variation that is a specific characteristic of each time series. $\boldsymbol{\Omega}$ is a matrix of unknown loads or weights with dimension $N \times R$ and rank R. Each element of $\boldsymbol{\Omega}$ associates a factor

[1] Dimensionality reduction has also been successfully applied to functional data (Hörmann et al., 2015).

[2] Even though VARIMA models allow to capture the relations between the series of the dataset, the "curse of dimensionality" can easily become a problem in estimation. For example, take $N = 25$, which for many situations is not a very large number of series to include. A simple VAR(1), $\boldsymbol{y}_t = \boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{w}_t$, would require the estimation of a matrix of coefficients $\boldsymbol{\Phi}$ of size 25×25 relating the value of each variable in time t to the values of all the variables (itself and the others) at $t - 1$.

[3] Using an ARIMA model for each series would ignore all together any cross-correlation between them. ARIMAX models on the other hand, which incorporate exogenous variables as predictors, most likely will present multicollinearity and would be unfeasible when T is not large enough.

of vector $\boldsymbol{F}_t$ with a variable in vector $\boldsymbol{y}_t$, thus indicating the variable's sensitivity to changes in the underlying factors. The term $\boldsymbol{\Omega F}_t$ is usually known as the common component. As for notation, notice that, even though one is written in capital letters, both $\boldsymbol{y}_t$ and $\boldsymbol{F}_t$ are vectors; $\boldsymbol{F}_t$ is written in capital letter in order to indicate that it represents the actual (unobserved) factors, while the estimated factors will be denoted by $\boldsymbol{f}_t$.

In this context, alternative assumptions may be made. In the so-called "classical factor analysis", the number of series included, N, is considered fixed, and $N \ll T$ (Bai, 2003). Additionally, $\boldsymbol{\varepsilon}_t$ is white-noise, independently and identically distributed usually normal, with full-rank diagonal covariance matrix (Peña and Box, 1987). This last requirement describes the so called Exact FM: the specific factors are mutually uncorrelated for all leads and lags (Geweke and Singleton, 1981; Chamberlain and Rothschild, 1983, though they call it *strict* rather than *exact*). One way to interpret this assumption is that any correlation between the studied time series is originated in the common factors, either because they have heavy weights in the same factors, or because the common factors are correlated between them (Geweke and Singleton, 1981).

Alternatively, in the "approximate factor analysis", some correlation in vector $\boldsymbol{\varepsilon}_t$ is allowed. For example, Stock and Watson (2010) allow for cross and serial correlation, while Chamberlain and Rothschild (1983) and Forni et al. (2005) allow some cross correlation. In this case, the series in $\boldsymbol{\varepsilon}_t$ may be modeled as dynamic processes themselves. Additionally, there has been a trend in the literature to include a large number of series (for practical purposes, $N > 100$), more than has been usual in classical applications (Boivin and Ng, 2006).

A further assumption both in approximate and classical factor analysis is that specific factors are uncorrelated with common factors. However, it is possible for the common factors to present cross-correlation and it is also possible for any of the latent variables to be serially correlated (Geweke and Singleton, 1981).

Notice that, because both $\boldsymbol{\Omega}$ and $\boldsymbol{F}_t$ are unknown, their product is undefined under rotations. This means there are infinite configurations of $\boldsymbol{\Omega}$ and $\boldsymbol{F}_t$ that would fit in equation (1.1) since, for instance, $\boldsymbol{\Omega}^* \boldsymbol{F}_t^* = \boldsymbol{\Omega} \boldsymbol{V} \boldsymbol{V}^{-1} \boldsymbol{F}_t$ verifies the model (for any non-singular matrix, $\boldsymbol{V}$). The literature proposes two alternatives to deal with the indeterminacy. On the one hand, we can make assumptions for the matrix of weights. Most often, it is assumed that $\boldsymbol{\Omega}\boldsymbol{\Omega}' = \boldsymbol{I}_{N \times N}$, where $\boldsymbol{I}$ stands for the identity matrix (for example, see Peña and Box, 1987). Alternatively, it can be assumed that $\boldsymbol{F}_t' \boldsymbol{F}_t = \boldsymbol{I}_{T \times T}$ (for example, see Chamberlain and Rothschild, 1983). Notice that, whatever the restriction imposed, there are no consequences for the commonality $\boldsymbol{\Omega} \boldsymbol{F}_t$.

Equation (1.1) is also known as the "static factor model", given that the common factors enter the equation only contemporaneously. In other words, this specification excludes lags of $\boldsymbol{F}_t$ and a lagged polynomial loading matrix $\boldsymbol{\Omega}(\boldsymbol{L})$ which constitute the Dynamic Factor Model (DFM); the same approach is followed by Bai (2003). Notwithstanding, this does not mean that the common factors do not exhibit some sort of dynamic behavior; on the contrary, their evolution could be explained by a model of the following type

$$\boldsymbol{\Phi}(\boldsymbol{L}) \boldsymbol{F}_t = \boldsymbol{\eta}_t, \tag{1.2}$$

where $\boldsymbol{\Phi}(\boldsymbol{L})$ represents a polynomial containing the lag operator L. In this dissertation we will consider seasonal ARIMA and AR models for the common factors, depending on the feature we study. Other options are, for example, using Vector Autorregressive (for example Poncela et al., 2011), or even Factor Augmented Vector Autorregressive (Poncela et al., 2014, in which a common factor is related to fundamentals, financial and uncertainty variables) models for the common factors.

The static factor model could be considered a particular case of the more general dynamic factor model (Geweke and Singleton, 1981), in which the effect of lagged factors is included as well. In that case, $\mathbf{\Omega}$ in Equation (1.1) is replaced by a lagged polynomial matrix $\mathbf{\Omega}(\boldsymbol{L})$ or some other sort of time dependence. For instance, Stock and Watson (2002) present the loads' time-dependence with an equation like $\mathbf{\Omega}_{r,t} = \mathbf{\Omega}_{r,t-1} + g_r \boldsymbol{\varsigma}_{r,t}$, where g_r is a scalar and $\boldsymbol{\varsigma}_{r,t}$ a vector of variables. These authors show that in order for principal components to continue to be consistent under this approach, some assumptions are needed: the change between consecutive periods (g_r) and the cross-sectional dependence in $\boldsymbol{\varsigma}_{r,t}$ should both be small.

We focus on the static factor model because its properties have received more attention than those of dynamic models and because, while the static approach is easier to understand and estimate, the two approaches produce similar forecasts (Bai and Ng, 2008).

Last, notice that the matrix of variance-covariance for $\boldsymbol{y}_t$, $\mathbf{\Sigma}_y$, can be expressed as the sum of the variation due to the commonality and the variation that is specific to each series (Stock and Watson, 2006). As explained before, we are working under the assumption that common and specific factors are uncorrelated (for all leads and lags). Then,

$$\mathbf{\Sigma}_y = \mathbf{\Omega}\mathbf{\Sigma}_F\mathbf{\Omega} + \mathbf{\Sigma}_\varepsilon. \tag{1.3}$$

Additionally, the variance of the specific components, $\mathbf{\Sigma}_\varepsilon$, should be much smaller than the one for the common factors, $\mathbf{\Sigma}_F$; otherwise, variability that is specific to each series may take part of the estimated common factors (Mardia et al., 1979, pp.276). Similarly, the lagged covariance matrix for $\boldsymbol{y}_t$ is a function of the lagged covariance for the common factors. As the factors are assumed to be uncorrelated, the eigenvectors of the lagged covariance matrix for $\boldsymbol{y}_t$ are the columns of the matrix of weights (Peña, 2009).

1.2 Estimation of the Common Factors

Some of the alternatives available to estimate FM are: Maximum Likelihood, Principal Components Analysis, hybrid techniques, and Bayesian techniques.

Gaussian Maximum Likelihood Estimation

A widely extended technique is Gaussian Maximum Likelihood Estimation (MLE). In this case, the Kalman filter is employed to obtain estimates for the factors[4], innovations, and their variance-covariance matrix. To estimate the matrix of loads, the constants of the model (if any), the variance-covariance matrix of the specific factors, and the parameters of the factors' model, direct maximization of the log likelihood function can be attempted, for instance, employing Newton-Raphson's algorithm. Or, other techniques can be employed, such as the Expectation Maximization (EM) algorithm. Since these techniques are usually employed for factors following VARIMA models, and we will be modeling the common factors as ARIMA models instead, we refer to Stock and Watson (2010) or Alonso et al. (2011) for detailed explanations and applications.

An advantageous property of this technique is that the estimates for the factors are efficient (Stock and Watson, 2010) and misspecification[5] has a negligible effect in the factors' estimates when N and T are large (Doz et al., 2012). Additionally, it can be employed even when there is missing data (Stock and Watson, 2010; Doz et al., 2012), which is useful for nowcasting.

[4] Interval estimates for the common factors can also be obtained. See Ruiz and Poncela (2015a).

[5] Ignoring serial correlation of the observed variables, or wrongly assuming that the specific factors are not cross correlated.

Ruiz and Poncela (2015b) analyze the uncertainty of estimating the common factors with the Kalman Filter for finite N and T.

Given that the number of parameters to estimate grows with the number of series N, a drawback of this approach is that estimation involves non linear optimization. Therefore, for practical purposes N should not be too large (Stock and Watson, 2006, 2010).

Principal Components Analysis

Among non-parametric estimation techniques (Stock and Watson, 2010), a popular option is to employ a dynamic adaptation of Principal Components Analysis (PCA). The Principal Components (PC) act as estimates of the common factors. This allows to work with a large N while maintaining the number of parameters low, and computation is easier than for parametric techniques.

Peña and Box (1987) presented the PCA extension for stationary series. They use a canonical transformation to obtain a structure for the underlying factors. They decompose vector $\boldsymbol{y}_t$ into an ARMA(1,1) process for R factors; and a white noise process (with dimension $N - R$). However, as Stock and Watson (2010) point out, $\boldsymbol{F}_t$ is not estimated directly, which is a drawback for obtaining forecasts.

Lee and Carter (1992), working with mortality rates, proposed to estimate the matrix of loads, $\hat{\boldsymbol{\Omega}}$, by Singular Value Decomposition (SVD) of the variance-covariance matrix of the centred dataset, $\hat{\boldsymbol{\Sigma}}_y$. Normalization conditions for weights and factors are imposed to obtain a unique solution. See the appendix of García-Martos et al. (2012) for a derivation from the optimization process to finding the eigenvectors.

Following Lee and Carter (1992) we use the matrix of centred data $\boldsymbol{Y}_{N\times T}$ and calculate the sample variance-covariance $\hat{\boldsymbol{\Sigma}}_y$, of dimension $N \times N$ which is assumed to be consistent (Bai, 2003). The SVD of this matrix renders N eigenvectors. Each eigenvector is associated to an eigenvalue. The greater the eigenvalue, the greater the percentage of variability of the data explained by the associated factor. Thus, as Stock and Watson (2010) explain, $\hat{\boldsymbol{\Omega}}_{N\times R}$ consists of the R eigenvectors corresponding to the highest R eigenvalues. The matrix of common factors, $\hat{\boldsymbol{F}}$, is therefore estimated as a linear combination of the time series: $\hat{\boldsymbol{F}}_{T\times R} = \boldsymbol{Y}_{T\times N}\hat{\boldsymbol{\Omega}}_{N\times R}$.[6] The specific or idiosyncratic factors are the portion of the data not captured by the R common factors included, $\boldsymbol{\varepsilon}_{N\times T} = \boldsymbol{Y}_{N\times T} - \hat{\boldsymbol{\Omega}}_{N\times R}\hat{\boldsymbol{F}}'_{R\times T}$.

Equivalently, this is the solution to the problem

$$\begin{aligned} min_{\boldsymbol{F}_t,\boldsymbol{\Omega}} \quad (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \omega_i F_{it})^2 \\ s.t. \quad \boldsymbol{\Omega}'\boldsymbol{\Omega}/N = \boldsymbol{I}_{R\times R}. \end{aligned} \tag{1.4}$$

Instead, when there are more series than time periods ($N > T$), it is computationally advantageous to perform the optimization with respect to the matrix of loads, replacing the condition in (1.4) by $\boldsymbol{F}\boldsymbol{F}'/T = \boldsymbol{I}_{R\times R}$ (Stock and Watson, 2002). In this case the estimated factors are equal to the R eigenvectors associated to the R highest eigenvalues of $\boldsymbol{Y}\boldsymbol{Y}'$ (instead of $\boldsymbol{Y}'\boldsymbol{Y} \propto \hat{\boldsymbol{\Sigma}}_y$). However relevant in many current applications, we will not study this specification. See Bai (2003) for theoretical developments and empirical applications in this context. In any case, an $N > T$ situation may be transformed into $N < T$. Boivin and Ng (2006) explain that, when there exists correlation between the idiosyncratic factors, pre-screening the series to include in the factor model may result in better forecasts than using a

[6]Note on notation: $\boldsymbol{F}_t$ represents the $R \times 1$ vector of true factors at time t (unknown), $\hat{\boldsymbol{F}}$ represents the $R \times T$ matrix of estimated factors, and $\boldsymbol{f}_t$ represents the $R \times 1$ vector of estimated factors at time t. We are assuming R is known, though in practice this is estimated as well.

very large N. In their empirical example they reduce the number of series from 147 to 40.

In the static form, Stock and Watson (2002) establish that the estimator of principal components is consistent for the space spanned by the factors when $N, T \to \infty$.[7] This is true as long as the specific errors are stationary and the loads of the factors are not trivial. This conclusion is maintained even when the specific factors exhibit weak cross-sectional or serial correlation (Stock and Watson, 2002, investigate this issue in simulations). Additionally, the estimator is robust to small and idiosyncratic changes of the factors' weights (Stock and Watson, 2002).

Even if we do not have N and T very large in real data applications, research achieves a high degree of accuracy for the factors' estimates for values of N, T as small as 30 (Bai and Ng, 2008) or 40 (Boivin and Ng, 2006, for N). This result is key to support a second step estimation in which the factors are taken as observed; in other words, the estimation error is considered to be negligible (see Bai, 2003, for an assessment of the conditions in which this assumption is reasonable).

In the classical approach (large T, N fixed), the estimated PC is asymptotically normal (Bai, 2003). In practice, this is the case even if some of the assumptions are altered, such as allowing some correlation and heteroscedasticity in the series (Bai, 2003).

When the variables are normally distributed, the PC estimator is asymptotically equivalent to the maximum likelihood estimator (Bai, 2003, also considering large values of N and T).

Another interesting property is studied by Bai and Ng (2008) for non-stationary data: "even if each cross-section equation is a spurious regression,

[7] Stock and Watson (2002) indicate that there are no restrictions regarding the rates of growth of N vs. T.

the common stochastic trends are well defined and can be consistently estimated, if they exist". In this case the factors are estimated by PC after differenciating the data. This approach considers that non stationary data may be caused by $\boldsymbol{F}_t \sim I(1)$, or $\boldsymbol{\varepsilon}_t \sim I(1)$, or both. It is different from Lee and Carter (1992) in that the latter estimates $\boldsymbol{F}_t$ as a random walk with drift (so, non-stationary factor) but ignores the specific error implicitly assuming $\boldsymbol{\varepsilon}_t \sim I(0)$.

There are many variations and approaches for particular situations and for dropping assumptions. For instance, in the case in which the specific components are correlated, feasible generalized PC can be employed (feasibility has to do with the fact that $\boldsymbol{\Sigma}_\varepsilon$ is unknown) (Stock and Watson, 2010). When the model is dynamic (incorporating lags of the factors in (1.1)), it may be re-written as static and PCA may still be employed; while dynamic confirmatory factor models work with the frequency domain (Geweke and Singleton, 1981). Forni et al. (2005) also work with frequency domain techniques and propose a predictor that takes advantage of the data's dynamic covariance structure and which weights the variables by an estimation of the signal-to-noise ratio. Weighted PCA allows to gain efficiency when there are non spherical specific factors. For instance, Boivin and Ng (2006) propose alternative weighing schemes, including the option of ruling out series that do not contribute relevant information. Last, as a generalization of PCA for non-Gaussian data and independent components, Independent Component Analysis can be used (García-Ferrer et al., 2011, 2012).

Hybrid Techniques

Stock and Watson (2010) indicate the possibility of employing hybrid techniques that combine state space models and principal components for the estimation of the factors. For static factors, the common factors are estimated by PC, and then the matrix of weights is obtained by regressing $\boldsymbol{y}_t$

on the estimated factors. The $\Phi(L)$ coefficients for the model of the common factors are obtained from the estimated factors as well. See Stock and Watson (2010) for details on the estimation when the factors are taken to be dynamic.

Doz et al. (2012) summarize the advantages of this approach: possible improvements in efficiency, it allows handling missing data, and it has benefits characteristic of parametric models such as the possibility of imposing constraints in the weights.

Bayesian Techniques

Bayesian techniques may also be employed. They incorporate prior distributions that allow to include the researcher's a-priori beliefs and are specially helpful for models that involve non Gaussian errors (Stock and Watson, 2010). Markov Chain Monte Carlo (MCMC) techniques are employed to estimate the parameters of the FM and the factors themselves.

Some examples of the implementation of this approach are: Otrok and Whiteman (1998), Kim and Nelson (1998), and Peña and Safadi (2008). The first one presents an empirical application for local economic data in the United States. The authors estimate one underlying factor using data augmentation and MCMC to sample from the posterior distribution of the factor. The second one is a work to model the business cycle, producing inference for a dynamic factor model with regime switching and obtaining Bayesian estimates for a common unobserved factor ('growth component') and a regime switching variable, also with United States data. Last, Peña and Safadi (2008) use Bayesian estimation to perform an estimation of a dynamic factor model to associate air pollution with mortality in Brazil.

1.3 Selection of the Number of Factors

A source of uncertainty in the estimation of factor models is how many common factors to include, since the actual number, R, is unknown. There are several criteria to obtain an estimate (r), though there is no definite assessment as to which one is the most accurate for forecasting (Poncela et al., 2011, obtain that more components may not result in better forecasts). Fortunately, the asymptotic distributions of the estimated factors and their loadings are unaffected by the estimation of the number of factors (Bai, 2003). Furthermore, in a model with specific factors either mutually uncorrelated or correlated, Boivin and Ng (2006) find that estimating $r < R$ implies a loss in efficiency, while $r > R$ only slightly affects the forecasts. As previously mentioned, we focus in the static case, in which the factors enter equation (1.1) only contemporaneously.

A popular, somewhat arbitrary, criterion for deciding the number of common factors to include in the model is selecting r such that the explained data variability is at least 80%. The percentage of variability explained by each factor can be calculated by $\lambda_i / \sum_{i=1}^{N} \lambda_i$, where λ represents the eigenvalues resulting from the SVD. This approach is employed, for instance, in Forni et al. (1999) and García-Martos et al. (2012).

Other approach to obtain the estimate r is the scree plot. Scree plots are a visual diagnostic tool which starts by arranging the eigenvalues of $\hat{\boldsymbol{\Sigma}}_y$ in descending order. Some tests based on them consist of calculating the ratio of adjoining eigenvalues and selecting the maximum (Ahn and Horenstein, 2013), the maximum ratio of adjoining growth rates of residual variances (Ahn and Horenstein, 2013), and a test to identify changes in the curvature of the scree plot for factor models with correlated normal specific errors (Onatski, 2006).

Alternatively, Peña and Box (1987) obtain r from the rank of the lagged covariance matrices (stationary factors), while Peña and Poncela (2006) propose a test based on the χ^2 which is valid for the static and the dynamic factor model.

The problem of determining r can also be seen as a problem of model selection. Bai and Ng (2002) indicate that there is a trade off between fit and efficiency: including more common factors in the model tends to improve fit, but at the cost of lower efficiency because more factor weights need to be estimated. Because the factors are not observable, the Bayesian Information Criterion (BIC) may not provide with a consistent estimator of R; while Akaike Information Criterion (AIC) tends to over-estimate R (Bai and Ng, 2002). See Bai and Ng (2002) for details on the conditions under which it is still useful to employ the well known AIC or BIC.

Bai and Ng (2002) and Bai and Ng (2008), among others, developed alternative information criteria. In their review, Bai and Ng (2008) discuss two criteria with the important feature that the penalty function, $g(N,T)$, depends on both dimensions: the number of series as well as the number of observations[8]. Under some conditions that control over and under fitting, the proposed criteria render r (estimated number of factors) that converges in probability to the true number of factors R, taking $N,T \to \infty$. These criteria are (Bai and Ng, 2002)[9]:

$$\hat{r}_{PCP} = \underset{0 \le r \le r_{max}}{\operatorname{argmin}} PCP(r), \tag{1.5}$$

[8] Bai and Ng (2002) show that a consistent estimation of R requires a penalty that depends on N,T instead of including only one of these dimensions.

[9] We adapted their notation to provide continuity with the one employed throughout this dissertation.

where $PCP(r) = S(r) + r\bar{\sigma}^2 g(N,T)$, $\bar{\sigma}^2 = S(r_{max})$ (r_{max} is set in advance), and

$$\hat{r}_{IC} = \underset{0 \leq r \leq r_{max}}{\operatorname{argmin}} IC(r), \tag{1.6}$$

where $IC(r) = \ln(S(r)) + rg(N,T)$. Both criteria are loss functions and in both cases $S(r) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{\omega}_i^{r\prime}\hat{F}_t^r)^2$, with common factors estimated by a principal components procedure. Some alternatives for $g(N,T)$ are (Bai and Ng, 2002):

$$g_1(N,T) = \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right), \tag{1.7}$$

$$g_2(N,T) = \frac{N+T}{NT} \ln(C_{NT}^2), \tag{1.8}$$

and

$$g_3(N,T) = \frac{\ln(C_{NT}^2)}{C_{NT}^2}, \tag{1.9}$$

where $C_{NT}^2 = \min(N,T)$.

Some drawbacks of these criteria are that their behavior in small samples may diverge (see Bai and Ng, 2002, for their performance in different scenarios) and that the penalty factor is arbitrary. For other scenarios such as DFM or the general dynamic case, other criteria are appropriate (for instance Hallin and Liska, 2007).

A further possibility consists of working with a set of possible specifications and then combining their forecasts. Caggiano et al. (2011) employ forecast combination of models with alternative number of lags, with r fixed. Alternatively, in Chapter 2 we employ this approach to circumvent the need to select a particular number of common factors. We estimate models with different number of common factors and weight their forecasts in alternative ways.

1.4 Models for the Factors

We may employ alternative approaches to compute forecasts of the variables of interest. Stock and Watson (2010) consider 'direct multistep $(t+h)$ forecasts', obtained by regressing $\boldsymbol{y}_{t+h}$ on $\boldsymbol{f}_t$, $\boldsymbol{y}_t$, and their lags, and 'iterated multistep forecasts', which consists of using a one-step forecast equation for $\boldsymbol{y}_{t+1}$ and the model for the factors to iterate and obtain $\boldsymbol{y}_{t+2}, ..., \boldsymbol{y}_{t+h}$. A theoretical pronunciation regarding the best technique would require knowledge of the true generating process (Marcellino et al., 2006) and empirical results are inconclusive for deciding which one is best (Stock and Watson, 2010).

Our approach, thus, follows part of the literature on DFM, including García-Martos et al. (2012) and Alonso et al. (2011). It is a case of 'iterated multistep forecasts' in which $\boldsymbol{y}_{t+h}$ depends only on estimates of $\boldsymbol{F}_{t+h}$ (obtained by iteration of the model for the common factors), and in which instead of estimating a forecasting equation for $\boldsymbol{y}_{t+h}$ we use the weight estimates of the FM in equation (1.1), implicitly assuming that the matrix of loads is time invariant.

To forecast $\boldsymbol{F}_{t+h}$ alternative models can be employed. We will present the cases of factors that follow ARIMA or VARIMA models. Either way the factors are modeled, alternative assumptions can also be considered regarding variances, correlations, and the behavior of the errors $\boldsymbol{\eta}_t$. These assumptions could prove reasonable or not depending on empirical circumstances.

Authors that model the common factors as ARIMA are the following. Peña and Box (1987) assume $\boldsymbol{F}_t$ which follows an ARMA process and generalizes to factors with dynamic dependence and allowing the noise matrix to have contemporaneous dependency. García-Martos et al. (2012) use common factors that follow univariate seasonal ARIMA models. The same for García-Ferrer et al. (2011) except they estimate the common factors with an

Independent Component Analysis (ICA) algorithm. Lee and Carter (1992) have a model with a single factor which follows a random walk with drift (an ARIMA(0,1,0)). Peña and Poncela (2004) work mainly with models of one common factor following ARIMA specifications, but also generalize their conclusions for a model with three common factors, each specified as an ARI model.

Notice that in this case the factors are taken as independent (any cross-correlation between them is ignored). Expanding equation (1.2) for detail, the ARIMA model for each estimated factor $f_{i,t}, i = 1, \ldots, r$ would be the following

$$(1-L)^d(1-L^s)^D\phi_i(L)\Phi_i(L^s)f_{i,t} = c_i + \theta_i(L)\Theta_i(L^s)\eta_{i,t}, \tag{1.10}$$

where $\phi_i(L) = (1 - \phi_{i,1}L - \phi_{i,2}L^2 - \ldots - \phi_{i,p_i}L^{p_i})$, $\Phi_i(L^s) = (1 - \Phi_{i,1}L^s - \Phi_{i,2}L^{2s} - \ldots - \Phi_{i,P_i}L^{P_i s})$, $\theta_i(L) = (1 - \theta_{i,1}L - \theta_{i,2}L^2 - \ldots - \theta_{i,q_i}L^{q_i})$, and $\Theta_i(L^s) = (1 - \Theta_{i,1}L - \Theta_{i,2}L^{2s} - \ldots - \Theta_{i,Q_i}L^{Q_i s})$. L is the lag operator such that $Lf_{i,t} = f_{i,t-1}$. The roots of the polynomials are outside the unit circle, which translates in stationary (for the AR part) and invertible (for the MA part) processes. The innovations $\eta_{i,t}$ are assumed to be white noise; it is customary to assume that they are uncorrelated for all leads and lags $E(\eta_{i,t}\eta_{i,t+h}) = 0$ $(h \neq 0)$, as well as uncorrelated with the specific factors $E(\eta_{i,t}\varepsilon_{j,t}) = 0, j = 1, \ldots, N$. d and D are the number of regular and seasonal differences, respectively, needed to make the series of the common factor stationary.

When the factors are expected to be cross-correlated at different lags of time they are modeled as VARIMA. Among others, this is the case of Stock and Watson (2010), who consider a DFM with factors following a VAR model and Alonso et al. (2011), who work with seasonal common factors.

In this case, the polynomials in equation (1.10) are replaced by $r \times r$ polynomial matrices $\boldsymbol{\phi}(L), \boldsymbol{\Phi}(L^s), \boldsymbol{\theta}(L), \boldsymbol{\Theta}(L^s)$. We would as well replace $f_{i,t}$ and $\eta_{i,t}$ by vectors $\boldsymbol{f}_t$ and $\boldsymbol{\eta}_t$, respectively. The same assumptions for the errors are usually maintained (ex. Stock and Watson, 2010).

Notice that the parameters of the factors' model may be estimated simultaneously with the common component (this is usually the case of a factor model in which $\boldsymbol{F_t}$ is modeled as a VARIMA, which can be estimated by ML) or in a second step after the common factors are obtained (this is usually the case when each $f_{i,t}$ is modeled as an ARIMA).

In Chapter 2 we consider common factors that follow seasonal ARIMA models. This specification will allow rapid computation of several models (for alternative parameters). We use a more parsimonious version, AR factors, in Chapter 3 in order to assess a possible bias in the estimation of the AR coefficients when the sample size (T) is small and the factors are highly persistent.

1.5 Outline of the Dissertation

The techniques and applications contributed in this thesis are presented in chapters 2 and 3. Their objective is to improve the forecasting results for multivariate time series datasets when working with factor models. We address two problems that we find have not been studied in depth in the existing literature. On the one hand, we consider the rigidity involved in selecting a 'best' model to produce forecasts. This implies selecting a number of common factors and model them in a particular way. To introduce some flexibility and improve forecasts, we use forecast combinations. On the other hand, we study the problem of biased estimates for highly persistent AR common factors, specially when the dataset consist of a small sample in the time dimension (small T).

In Chapter 2 we study whether forecast combination techniques have the ability to outperform a benchmark FM selected with a popular information criterion. Our motivations are decisions the researcher must make when employing FM to reduce the dimensionality of a dataset and produce forecasts. In particular, we consider estimating several specifications for the common factors and employing models with alternative numbers of those factors, in lieu of a criteria to select a number of common factors and one model for them. Having a group of specifications allows to apply combination techniques to produce forecasts. In particular, we evaluate the performance of five combination techniques in applications to electricity prices of the Iberian and Italian markets.

In Chapter 3 we work with two techniques to correct the bias of AR estimates: Bootstrap Bias-Correction (Clements and Kim, 2007) and Roy-Fuller's (Roy and Fuller, 2001). These techniques are particularly successful when the series are highly persistent and the sample size (T) is small. We innovate by employing them in a context of AR common factors. We use Monte Carlo simulations to show that the improvement in the factors' coefficients translates into more accurate forecasts, under diverse conditions. Additionally, we illustrate the performance of these techniques in an empirical setting consisting of data for the Industrial Production Index (IPI) for a group of European countries.

Finally, in Chapter 4 we summarize the conclusions and present further lines of research that could derive from our work.

Chapter 2

Electricity Prices Forecasting by Averaging Factor Models

2.1 Introduction

Nowadays, electricity trading is liberalized in most countries of the Western world. Due to the particular characteristics of supply and demand, prediction of electricity prices in this context is complex. Notwithstanding the difficulties, forecasts are necessary for several reasons:

- this is a strategic sector of the economy,
- there are financial implications due to the trading of forwards and options,
- forecasts help optimize and plan consumption and production.

As with other commodities, there are various ways to operate in this market (see Weron, 2014; Conejo et al., 2010a, for detailed market descriptions). We focus on prices that result from a *pool*[1] in which there is a central auction. In this *pool*, prices could be settled for each hour of the day, or every half hour, depending on the market.

In the first case, the 24 hourly prices for day t are cleared at the same instant in day $t-1$, with the same common information for all the hours. Therefore, for each day, a 24-dimensional vector is generated $(p_{1,t}, p_{2,t}, \ldots, p_{24,t})$; where $p_{hour,t}$ represents the price of $hour = 1,\ 2, \ldots,\ 24$ at day t. Consequently, prices can be presented in a $T \times 24$ dimensional matrix, where T is the number of days in the sample, and modeling should be multivariate (as in Huisman et al., 2007; García-Martos et al., 2007; Panagiotelis and Smith, 2008; Alonso et al., 2011). Even more, modeling could be attempted by transforming hourly (discrete) prices into functional data (see Hörmann et al., 2015, for a methodology of dimension reduction for time series functional data).

In several fields, there has been an increasing interest in the development of methodology to deal with multivariate time series or a high dimensional vector of series like the ones in electricity markets. By the end of the 1970s, Sargent and Sims (1977) (these authors presented a factor model for stationary time series vectors) and Geweke (1977) were the first to propose a Dynamic Factor Model. Later, Lee and Carter (1992) contributed by extending the idea of Principal Components to the dynamic case. More recently, dimensionality reduction techniques have gained popularity, in particular since the work by Stock and Watson (2002). For example, Peña and Poncela (2004) and Peña and Poncela (2006) extended Sargent and Sims (1977)'s model for the non-stationary case.

Regarding applications in electricity markets, García-Martos et al. (2012) extended Lee and Carter (1992) and Peña and Box (1987) to prices with seasonality. Working with data for the Iberian market for 2007-2009, they

[1]Among others, a reference including balancing settlement markets is Morales et al. (2014).

propose extracting common factors from the 24-dimensional price vector and modeling such factors as univariate seasonal AutoRegressive Integrated Moving Average (ARIMA) processes. Another example is Alonso et al. (2011), who propose a technique called Seasonal Dynamic Factor Analysis (SeaDFA), which involves the estimation of a Vector AutoRegressive Integrated Moving Average (VARIMA) model for unobserved common factors having seasonal patterns. The work in Maciejowska and Weron (2015) also uses a Factor Model, including hours and locations.

In an independent path, Forecast Combination or Model Averaging has been developed as a technique to take advantage of the availability of alternative forecasting approaches. This methodology consists of weighting a set of forecasts corresponding to alternative models and combining them to obtain a single forecast. In this way, model selection uncertainty is incorporated. According to Clemen (1989), 'the idea of combining forecasts implicitly assumed that one could not identify the underlying process, but that different forecasting models were able to capture different aspects of the information available for prediction'. Other justifications for model averaging are: doubts of the existence of a 'best model' (Sánchez, 2006), 'portfolio diversification', a better adaptation to structural breaks, or to average out omitted variables bias (Bjørnland et al., 2010).

Applications of model averaging in electricity markets are given by Bordignon et al. (2013), for the British Market and Nowotarskia et al. (2014), for European and USA markets. Furthermore, Raviv et al. (2015) obtain forecasts for the daily average price employing dimensionality reduction techniques as well as Forecast Combination of several models for hourly prices. Other references are Monteiro et al. (2015), who use averaging to obtain wind speed, solar irradiation, and temperature forecasts, which are then employed to estimate prices; and García-Martos et al. (2015), who forecast hourly electricity prices for the Spanish market by weighting seasonal ARIMA (with exogenous variables) and seasonal Dynamic Factor Models of similar performance.

In spite of its advantages, a major drawback of dimensionality reduction techniques is the uncertainty concerning the 'correct' model: how many factors to include and what models they follow. The literature is not definite in regards to the best technique for estimating the number of underlying factors that would contain enough information to make accurate predictions, and it should be considered that, as the number of factors included increases, so does estimation complexity and computational burden. As previously indicated, there is not either a unique model for the factors that outperforms all other models, in all circumstances (Weron, 2014).

In this work it is assumed that the major decisions involved in forecasting by using dimensionality reduction techniques may be resolved in a less arbitrary way if we rely on Forecast Combination. In order to follow this line of thought, alternative models, including different numbers of common factors, are estimated. Forecasts for prices are obtained by transforming the factors' forecasts back to the data units, according to the relations established in the dimensionality technique employed. Subsequently, Forecast Combination approaches are used to weight each of the forecasts obtained and thus provide a single prediction.

Summing up, factor models extract information *ex ante* (before any forecast is obtained) while Forecast Combination works *ex post* (after forecasts are available). The contribution of this work is to amalgamate both techniques. A reduced number of latent unobserved variables is estimated and their forecasts are combined in order to obtain a single prediction.

We apply these techniques to one-day to two-month-ahead electricity prices for the Iberian spot Market for a period of five years (2008-2012); and for the Italian spot Market for a period of three and a half years (mid 2009-2012). Several ARIMA specifications[2] are estimated for the factors and used to obtain forecasts of the prices for each hour, which makes the task computationally intensive. Next, these forecasts are combined. We study alternative ways to combine forecasts because their performance may vary depending on the data-set. The predictions concern mainly the medium- and long-term (one and two months).

The rest of the chapter is organized as follows. Fundamentals containing a mathematical description of the proposed methodology are presented in Section 2.2, which includes definitions on Factor Models, classical techniques for Forecast Combination, and Bayesian Model Averaging. Section 2.3 describes the methodology for this chapter. In section 2.4, we present the results of the empirical applications. This section is divided into sub-sections presenting the data, an Analysis of Variance (ANOVA) comparing specifications, and forecasting results. Finally, section 2.5 concludes with remarks, limitations and possible extensions.

2.2 Fundamentals

An outline of the methodology used in this proposal is presented below as well as the drawbacks of other approaches, which we attempt to resolve.

2.2.1 Factor Model

Factor Models (FM) are a widely applied dimensionality reduction technique. It is employed when the researcher believes there are fundamental factors

[2]For each one of the common factors included in the analysis 36 choices of parameters are available: p = 1,2,3, $d = 0$, q = 1,2,3, P = 0,1, $D = 1$, Q = 0,1, $s = 7$. These predefined models are all automatically estimated with the software TRAMO, by its Matlab interface, intervening outliers.

driving several variables in a data-set. These factors, like the variables, evolve through time, and allow to obtain information about the larger data-set with a simpler model. The explanation here follows García-Martos et al. (2012). As there, once the common factors are obtained, univariate seasonal ARIMA models are fitted to them. The forecasts of these models are then combined to obtain one improved forecast.

Let $\boldsymbol{y}_t$ be an N-dimensional observed time series vector, generated by an R-dimensional vector of unobserved common factors with $R \ll N$. In the Iberian and Italian electricity markets $N = 24$ and the matrix of observed series has as many rows as days are considered in the historic data-set. As in Lee and Carter (1992), it is assumed that vector $\boldsymbol{y}_t$ can be written as a linear combination of the unobserved common factors $\boldsymbol{F}_t$, plus a vector of specific components or factors $\boldsymbol{\varepsilon}_t$:

$$\boldsymbol{y}_t = \boldsymbol{\Omega}\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \tag{2.1}$$

where $\boldsymbol{\Omega}$ is an $N \times R$ matrix of loads relating the set of R common unobserved factors with the vector of observed series $\boldsymbol{y}_t$ (the vector of the 24 hourly prices for our application) and $\boldsymbol{\varepsilon}_t$ is an N-dimensional vector of specific components.

To estimate the factors $\boldsymbol{F}_t$, singular value decomposition (SVD) is used (as in Lee and Carter, 1992) for the covariance of the 24 dimensional vector of centred prices (García-Martos et al., 2012). This consists in calculating the eigevalues and their associated eigenvectors, for the sample covariance matrix, and thereupon calculating the matrix of common factors, $\boldsymbol{f}$, as a linear combination of the time series: $\boldsymbol{f}_{T\times R} = \boldsymbol{Y}_{T\times N}\hat{\boldsymbol{\Omega}}_{N\times R}$.

The common factors $\boldsymbol{F}_t$ may be non-stationary, including regular or seasonal unit roots in addition to auto-regressive and moving average (regular and seasonal) components. These ARIMA$(p, d, q) \times (P, D, Q)s$ models are used to obtain factors' forecasts, and from them prices' forecasts. For instance, the i-th factor at date t, F_{it}, would be modeled by

$$(1-L)^d(1-L^s)^D\phi_i(L)\Phi_i(L^s)F_{it} = c_i + \theta_i(L)\Theta_i(L^s)\eta_{it}, \tag{2.2}$$

where $i = 1,2,\ldots,R$ is the i-th factor, $\phi_i(L) = (1 - \phi_{i1}L - \phi_{i2}L^2 - \ldots - \phi_{ip_i}L^{p_i})$, $\Phi_i(L^s) = (1 - \Phi_{i1}L^s - \Phi_{i2}L^{2s} - \ldots - \Phi_{iP_i}L^{P_i s})$, $\theta_i(L) = (1 - \theta_{i1}L - \theta_{i2}L^2 - \ldots - \theta_{iq_i}L^{q_i})$, and $\Theta_i(L^s) = (1 - \Theta_{i1}L - \Theta_{i2}L^{2s} - \ldots - \Theta_{iQ_i}L^{Q_i s})$ are polynomials, L is the lag operator such that $LF_{i,t} = F_{i,t-1}$. The roots of $|\phi_i(L)| = 0$, $|\Phi_i(L^s)| = 0$, $|\theta_i(L)| = 0$, $|\Theta_i(L^s)| = 0$, satisfy the usual stationarity and invertibility conditions, and $\eta_{it}{\sim}N(0, W_i)$ are uncorrelated $E(\eta_{it}\eta'_{it-h}) = 0, h \neq 0$. It is also assumed that the error term of the common factors η_{it} is uncorrelated with the specific components $E(\eta_{it}\varepsilon'_{t-h}) = 0, \forall h$. c_i is the constant of the model for the common factors and its inclusion in the common factors model (2.2) can be particularly relevant to calculate long-term forecasts in the non-stationary case, which is the case of electricity prices. Furthermore, in this work the specific components are assumed to be independent and have no dynamic structure along them (e.g. Peña and Poncela, 2006).

It should be noticed that we work in two consecutive steps: firstly we estimate the factors $\boldsymbol{f}$ and secondly we estimate the ARIMA models like (2.2) ($\hat{\phi}, \hat{\Phi}, \hat{\theta}, \hat{\Theta}$ are estimated for each common factor f_i). Morever, the estimation of the first factor is the same when $R = 1$ or $R > 1$, which is natural consequence of the SVD procedure. Nevertheless, the selection of R affects the forecasting errors for the series. The more factors are included (greater r), the greater the variability of the data explained by the model. On the other hand, the cost of incorporating more factors is an increase in the number of parameters to estimate.

To summarize, a key stage when estimating this kind of models is the selection of the number of common factors, R, as well as the model they follow, which implies selecting the orders: p, d, q, P, D, Q. r could be obtained by using existing tests such as the ones proposed in Peña and Poncela (2006) or Bai and Ng (2002), and could also be selected such that diagnostic checking results[3] are reasonable (Alonso et al., 2011). However, alternative values could satisfy these criteria. Because selecting one value for R and the other parameters will likely not render the best results in every scenario, we will instead keep the alternatives and combine their forecasts. Forecast Combination is presented in the following subsection.

2.2.2 Forecast Combination

Empirically, the improvements of using Forecast Combination instead of a "best single model" have been shown for different types of models (for instance see Poncela et al., 2011; Kuzin et al., 2012; Martínez-Álvarez et al., 2015), and in various research areas (Clemen, 1989; Stock and Watson, 2004). However, Weron (2014) points out that Forecast Combination techniques have not been fully exploited for electricity prices.

In general, we can think of the combination equation as follows:

$$\hat{y}^{C}_{t+h|t} = \sum_{i=1}^{K} w^{(h)}_{t,i} \hat{y}^{(i)}_{t+h|t}, \tag{2.3}$$

where $w^{(h)}_{t,i}$ is the *i-th* model weight at time t for the forecast horizon h, K the total number of models considered, and $\hat{y}^{(i)}_{t+h|t}$ the forecast obtained with the *i-th* model for the forecast horizon h.

Combinations will vary depending on the weights they use and the set of models they include. There are classical and Bayesian techniques. In the

[3]Specific factors and errors of the observation equation must be uncorrelated between them, and specific factors without any cross correlation.

next subsections we briefly summarize the literature on both classical approach and an approximation to Bayesian combination, mentioning their drawbacks and advantages. This will help us justify our methodological proposal presented in Section 2.3.

2.2.2.1 Classical techniques for Forecast Combination

One easy way to obtain a Forecast Combination is the simple average, in which all alternative forecasts are given the same weight. This approach often works very well in comparison with more complex ones. One possible reason for this is that 'complicated combining methods pursuing "optimal" behavior often lead to unstable weights and the combined forecast even performs significantly worse than the individual forecasts' (Yang, 2004). Alternatively, a simple combination method outperforming more complex ones might be explained by a larger variability of the latter (Yang, 2004). In this regard, Bjørnland et al. (2010) advice to use a simple average when the alternative models to combine have similar forecast error variance.

A different approach to assign weights consists of estimating weights that minimize a loss function with the forecast error of the models to combine as explanatory variables (Elliott et al., 2006).

A further option is a combination using only a subset d_m of the best models. Possible advantages of this approach are: to reduce the variability of the combination (Yang, 2004) and to avoid under-weighting independent information when the models are correlated (Bjørnland et al., 2010). The set d_m could change through time depending on the most recent performance of the models (Bjørnland et al., 2010)[4], or it could be fixed (Swanson and Zeng, 2001).

[4]These authors evaluate model performance based on the sum squared errors. The results they obtain with a time-varying subgroup of models outperform those of the simple average of all the models.

Another way to combine forecasts would be to employ the median prediction (Kuzin et al., 2012). Alternatively, some authors employ a combination regression of the form

$$y_{t+h} = \alpha_0 + \sum_{i=1}^{P} \alpha_i p_{i,t} + \epsilon_{t+1},$$

where y_{t+h} is the forecast resulting from the combination and $p_{i,t}$ are the predictions of the alternative models. Swanson and Zeng (2001) use the Bayesian Information Criterion (BIC)(Schwarz, 1978) or the Akaike Information Criterion (AIC) to select the best combination. There are also some drawbacks to this regression based approach. Swanson and Zeng (2001) indicate collinearity in the competing forecasts and over-fitting due to outliers; Wright (2008) adds that while in-sample fit is improved, out-of-sample prediction tends to be worse than using the average to combine. Poncela et al. (2011) also outperform this type of model with combination techniques that involve dimension reduction.

Even using complex combinations, empirical findings in Swanson and Zeng (2001) suggest that, in some cases, the difference between alternative combination methods is not significant, a result that will also be obtained at points in this work.

2.2.2.2 Bayesian techniques for Forecast Combination

With this approach, the predictive distribution of a new observation is obtained by averaging with different weights the predictive distribution of each model considered. The idea was initially introduced by Leamer (1978) and allows to incorporate the uncertainty regarding the variety of available models (Leamer, 1978). It has been applied in statistics (Raftery, 1995; Raftery et al., 1997; Chipman et al., 2001) and econometrics (Koop and Potter, 2003; Cremers, 2002).

According to Wright (2008), an advantage of Bayesian Model Averaging (BMA) is that 'One does not have to be a subjectivist Bayesian to believe in the usefulness of BMA, or of Bayesian shrinkage techniques more generally. A frequentist econometrician can interpret these methods as pragmatic devices that may be useful for out-of-sample forecasting in the face of model and parameter uncertainty'.

As Wright (2008) explains, the procedure takes under consideration a large number of alternative models' forecasts, assuming one of them is the 'true' data-generating model; however, the researcher is unaware of which one is this. A prior regarding which model is the correct one is set, and then *a posteriori* probabilities of the different models being the true one are obtained to weight the predictions.

Alternative models' weights can be time evolving. For instance, Billio et al. (2011) work with weights that change depending on the predictive densities past performance and learning mechanisms.

Following Wright (2008): let K be the total number of models $M_1, \ldots, M_K$. The i-th model is related to the vector of parameters θ_i. The researcher has *a priori* knowledge of the probability that the i-th model is the true one, $p(M_i)$. Then the data, D, is observed and the probability is updated by calculating the *a posteriori* probability that model i-th is the true one:

$$p(M_i|D) = \frac{p(D|M_i)p(M_i)}{\sum\limits_{i=1}^{K} p(D|M_j)p(M_j)}, \tag{2.4}$$

where $p(D|M_i) = \int p(D|\theta_i, M_i)p(\theta_i|M_i)d\theta_i$ is the marginal likelihood of the i-th model, $p(\theta_i|M_i)$ is the *a priori* density of that model parameters vector, and $p(D|\theta_i, M_i)$ is the likelihood. Inference about a 'future' quantity Δ is based on

$$p(\Delta|D) = \sum_{i=1}^{K} p(\Delta|D, M_i)p(M_i|D). \tag{2.5}$$

In particular, the mean of this posterior distribution can be used as forecast. This procedure minimizes the Mean Squared Forecast Error (MSFE). It is only necessary to specify the set of models, their priors $p(M_i)$, and the parameters' priors $p(\theta_i|M_i)$.

A disadvantage of this approach though, is that the conditional probabilities are, in general, unknown. Therefore, they should be estimated from the data, which could mean that any benefits of Forecast Combination are lost.

Often, all models will have equal *a priori* probabilities, i.e. $p(M_i) = 1/K$. In this case, as Raftery (1995) indicates, the posterior probability $p(D|M_i)$ is proportional to $\exp(-(1/2)BIC_i)$. Therefore, expression (2.4) can be written as follows,

$$p(M_i|D) \approx \frac{\exp(-(1/2)BIC_i)}{\sum_{i=1}^{K} \exp(-(1/2)BIC_i)}. \tag{2.6}$$

Expression (2.6) is easy to calculate and no prior densities need to be set (Raftery, 1995). In this chapter, one of the Forecast Combinations will use weights obtained as indicated in expression (2.6).

Notice that the selection of equal *a priori* probabilities is motivated by the approach of using non-informative *a priori* probabilities. However, other *a priori* probabilities can be considered and, in such cases, expression (2.6) would be:

$$p(M_i|D) \approx \frac{\exp(-(1/2)BIC_i)p(M_i)}{\sum_{i=1}^{K} \exp(-(1/2)BIC_i)p(M_i)}. \tag{2.7}$$

In this chapter the goal is to derive some feasible and reasonable weights, not to estimate conditional probabilities. Of course, it is to be expected that clever *a priori* probabilities would produce better weights in the sense of better forecast performance.

2.3 Methodology

Taking into account the limitations of existent approaches in dimensionality reduction, most importantly the issue of selecting a number of common underlying factors r, as well deciding for a 'best' model for them; and given the advantages of Forecast Combination revisited in the previous sections, our methodological proposal consists of averaging the forecasts of alternative models for each factor.

This allows to capture the factors underlying the behavior of large data-sets, avoiding the risk of committing to a particularly 'bad' specification for them. That is why we consider that this approach improves previously mentioned solutions to open problems described along sections 2.1 and 2.2.

The complete prediction procedure can be summarized in the following steps, repeated for each window of time in the data-set. Notice that each window of time provides a historical data-set as well as out of sample data with which the forecasts will be compared.

For each window of time, the factors underlying the data are estimated by means of SVD, as explained in Section 2.2.1. There are as many common factors as time series in the data-set, N. However, the purpose of applying dimensionality reduction techniques is to be able to describe the data by means of a much smaller number of variables, thus $R << N$. There are many criteria for estimating the value R that would best represent the underlying

Select a window of time (considering a historical length)

↓

Use SVD to calculate the weights and derive the common factors, $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ (we work with models of one and two common factors)

↓

Estimate 36[a] alternative seasonal ARIMA models for each factor

↓

Calculate forecasts employing the estimated seasonal ARIMA models, $f_{1,t+h}$ and $f_{2,t+h}$

↓

Transform to competing forecasts of hourly prices using either one factor ($\hat{y}_{T+h} = \omega_1 f_{1,T+h}$) or two factors ($\hat{y}_{T+h} = \omega_1 f_{1,T+h} + \omega_2 f_{2,T+h}$)

↓

Combine forecasts according to Equation (2.3)

[a]36 = alternatives for $p\times$alternatives for $q\times$alternatives for $P\times$alternatives for Q.

trends in the data. In this regard, a contribution of this work is that, instead of committing to one of them, the possibility of estimating several models is explored. For this reason, at least two settings are estimated: on the one hand $r = 1$, which means that only the most representative underlying factor is used to forecast; and on the other hand $r = 2$, which means that the first and second most important underlying factors are estimated and employed to obtain forecasts. Based on the percentage of the total variability explained by the common factors, having up to $r = 2$ in the case of the Iberian data corresponds to explaining about 80% of the total variability. However, for the Italian data we need $r = 3$ to achieve a similar result in terms of the percent of variability explained. Therefore, we incorporate the option of having up to three common factors in the models, and the steps described should accommodate a third factor $\boldsymbol{f}_3$ for this data-set.

As indicated in the flowchart, the next step consists of estimating models for the factors. The literature review performed in this work reveals that it is difficult, if not impossible, to find a model that by all criteria would outperform all others. Even more, a good fit does not guarantee an accurate forecasting performance. To overcome these difficulties, our proposal consists of fitting 36 ARIMA specifications for each estimated factor, in lieu of selecting a 'best' set of parameters. These specifications result from the following parameters: $p = \{1,2,3\}$, $d = \{0\}$, $q = \{1,2,3\}$, $P = \{0,1\}$, $D = \{1\}$, $Q = \{0,1\}$, $s = 7$. Additional values of the parameters (for example, $p > 3$) are excluded because they increase the computational burden but do not provide a relevant improvement in results.

After forecasts are estimated for all the options of factors (either one, two, or three) and ARIMA models, they are transformed to forecasts for the original variables by means of a multiplication by the matrix of weights following expression (2.1). This will render many forecasts for the data, which will be combined to present with a single forecast for each variable of the original data-set.

2.3.1 Forecast Combinations and Accuracy Metrics

We consider five alternative combinations (2 to 6 below) and compare them to a benchmark (1 in the next enumeration):

1. Forecast resulting from the benchmark model ('BIC-selected model' for future reference). This is the best model according to the BIC (has the lowest BIC). Selecting only one model is equivalent to assigning it a weight $w_i = 1$ in (2.3) and $w_i = 0$ for all other models. The superscript (h) has been eliminated from expression (2.3) because weights will not be adaptive to the forecasting horizons and subscript t has also been omitted to avoid confusion with time-varying weights.

2. Forecast calculated as the median of the forecasts of all the models ('median-based combination'). This is also a case of weights $w_i^{(h)} = 1$ for the model with the median forecast and $w_i^{(h)} = 0$ for all other models.

3. Forecast equal to the mean of all forecasts ('mean-based combination'). In this case, expression (2.3)'s weights are all equal $w_i = 1/K$, where K is the total number of models in the analysis.

4. Forecast obtained using BIC-based weights as in expression (2.6) ('BIC-based combination'). This approach involves equal *a priori* probabilities. Other sensible sets of *a priori* probabilities were considered and similar results were obtained.

5. Forecast obtained with BIC-based weights for the top 50% models ('BIC-50% combination'). In other words, half of the models are included according to their BIC criterion $w_i = p(M_i|D)$ of expression (2.6), and for the half that has the largest BIC values, $w_i = 0$. Let us recall that the BIC evaluates the fit of the model, not how accurate it is when used to forecast.

6. Forecast calculated as the mean of the forecasts of the top 50% models ('mean BIC-based combination'). Only half of the models are included (the 'best' half models depending on their BIC), and the Forecast Combination is simply their average. In other words, the 50% models with the lowest BIC are assigned weights $w_i = 2/K$ and the 50% models with the greatest BIC are assigned weights $w_i = 0$.

In order to evaluate forecasts and assess the most appropriate combination, we need to define a forecasting accuracy metric. We can evaluate the forecasts' accuracy by means of several alternative metrics, see Conejo et al. (2005); Hyndman and Koehler (2006); Weron (2014) for detailed reviews. Some of them are the relative forecast error and the Mean (and Median)

Average Percentage Error (MAPE). However, these measures are not valid when the data have negative and/or positive, but close to zero, values (Hyndman and Koehler, 2006), a frequent occurrence for many electricity markets (Bello et al., 2016; Monteiro et al., 2015, deal with the issue of forecasting extreme prices in the Spanish electricity market).

Therefore, we use the Mean Absolute Error (MAE) and Median Absolute Error (MedAE). They can be obtained as follows,

$$\text{MAE}^i_\tau = \frac{1}{m}\sum_{z=\tau+1}^{\tau+m}\left(\frac{1}{24}\sum_{h=1}^{24}|(y_{h,z} - \hat{y}^i_{h,z})|\right) \tag{2.8}$$

and

$$\text{MedAE}^i_\tau = \frac{1}{m}\sum_{z=\tau+1}^{\tau+m}(median(|y_{h,z} - \hat{y}^i_{h,z})|)), \tag{2.9}$$

where m is the number of days in the out-of-sample period, sub-index h is the hour, and τ is the last observation of the rolling window employed to estimate the model used to compute the forecasts. The error is defined as the difference between $y_{h,z}$, the actual price at hour h for z steps ahead, and the forecast of model or technique i, $\hat{y}^i_{h,z}$.

Additionally, in order to simplify the comparison between the benchmark and Forecast Combinations, the Relative MAE (RelMAE) will be computed. Following Hyndman and Koehler (2006), we calculate it as follows,

$$\text{RelMAE}^i_\tau = \text{MAE}^i_\tau/\text{MAE}^b_\tau, \tag{2.10}$$

where b indicates the benchmark model (BIC-selected model). As indicated in Hyndman and Koehler (2006), whenever $\text{RelMAE}^i_\tau < 1$ the forecast provided by the i-th combination is better than the one provided by the benchmark and the opposite happens when $\text{RelMAE}^i_\tau > 1$.

2.4 Results

2.4.1 Data

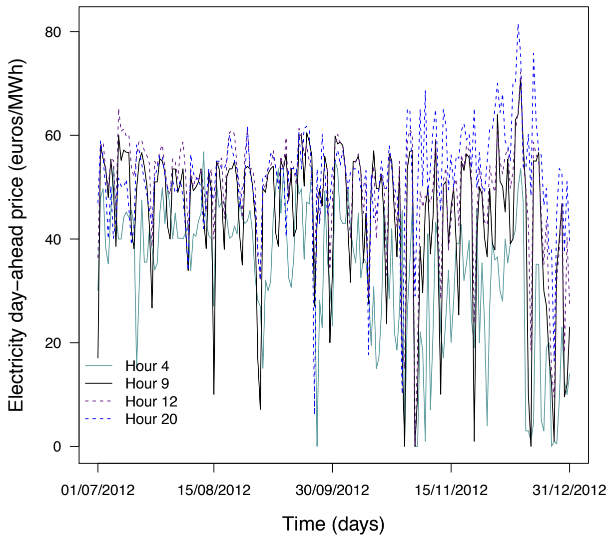
We study two data-sets of electricity spot prices, one for the Iberian market, which includes Spain and Portugal[5] (July 2006 - December 2012) and the other one for the Italian market (January 2008 - December 2012). To illustrate the behavior of these prices, a few representative time series are plotted. Figure 2.1a, corresponds to the Iberian market, while Figure 2.1b, corresponds to the Italian market. Both figures present hours 4, 9, 12, and 20, for the last six months of 2012. In each figure there is a common pattern in the evolution of the hourly series, which is what the common factors attempt to capture.

2.4.2 ANOVA for Comparison of Alternatives for Modeling

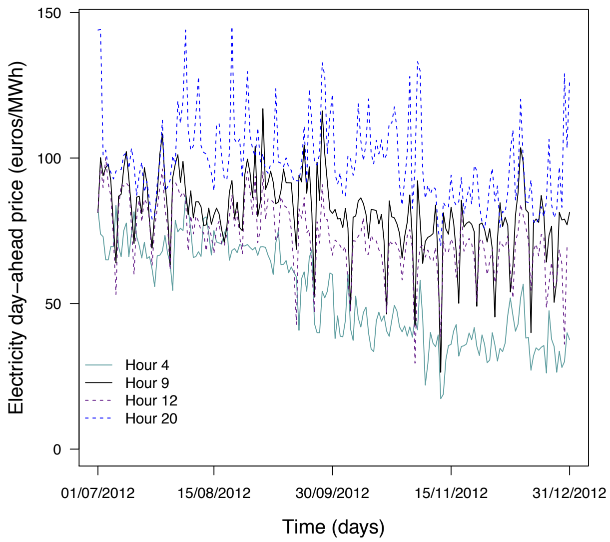
We rely on Design of Experiments (DOE) techniques to assess which is the forecasting methodology that produces the smallest error, measured by MAE, in the forecasts of electricity prices. We consider the following factors:

- *Logarithm*: this factor has two levels, *Logarithm*= {No, Yes}. *Logarithm*= No when we use the prices in the same way they are reported (€ per MWh). *Logarithm*= Yes means that we work with ln(*prices*). Taking logarithm has the effect of producing time series with less volatility.

[5]Nogales et al. (2002) indicates that for the Spanish market there is more uncertainty in hours of high demand than in hours of low demand, which affects the accuracy of forecasts for those hours.



(A) Iberian market.



(B) Italian market.

FIGURE 2.1: Electricity day-ahead prices for four representative hours during the last semester of 2012.

- *Historical Length*: this factor indicates how long is the dataset employed in each rolling window. It has two levels, *Historical Length*= {308 days, 548 days}. *Historical Length*= 308 days indicates that the common factors are extracted from series of prices with an extension of 44 weeks (García-Martos et al., 2012). *Historical Length*= 548 days (time series that extend for 1.5 years), supporting the well known idea that the estimation of common factors benefits from extensive data, but not so extensive as to include sub-periods with large differences in the variance explained by the common factors (an example of sub-periods with different co-movements is in Poncela et al., 2014).

- *Moving Average (MA)*: this factor has two levels, *MA*= {No,Yes}. *MA*= No makes reference to a forecasting methodology in which the common factors are fitted with AutoRegressive-Integrated (ARI) models. *MA*= Yes instead, allows greater complexity since in this case the common factors are modeled as having an AutoRegressive-Integrated-Moving-Average (ARIMA) behavior.

- *Forecast Combinations*: this factor has six levels, *Combinations*= $\{1, 2, 3, 4, 5, 6\}$ which were described in Section 2.3.1.

To compare these features, we have performed a computational experiment in which we computed one-day-ahead to two-month-ahead forecasts for every hour and day. This involves estimations for every day during five years (Iberian data), or three and a half years (Italian data), long periods of time that allow validating the results. There are $2 \times 2 \times 2 \times 6 = 48$ treatments resulting from combining all the levels of the aforementioned factors.

We analyze separately the performance of out-of-sample forecasts for various forecasting horizons: one-day-ahead (forecasting horizon $h = 1$), one-week-ahead ($h = 7$), one-month-ahead ($h = 30$), and two-month-ahead ($h = 60$). The goal is to select the treatment which results in the smallest forecasting error possible (measured by MAE) for each of these forecasting horizons h.

Furthermore, given the fact that forecasts have been computed for a large number of days, the particular *Day* could also explain some significant part of the variability of the response variable, MAE. For instance, if the prices in one day are rather unexpected for being too low or high, the MAE will be large, whatever the values for *Logarithm*, *Historical Length*, *Moving Average (MA)*, and *Forecast Combinations*. Therefore, *Day* is considered as a block in the computational experiment. This helps remove a likely correlation between forecasting errors.

An ANOVA with four factors and one block is conducted to compare the alternative forecasting methodologies (see Montgomery, 1984, for a complete reference on ANOVA and Design of Experiments).

The equation of the model is:

$$\mathrm{MAE}_{ijkld} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \epsilon_d + u_{ijkld}, \tag{2.11}$$

$$u_{ijkld} \sim NIID(0, \sigma_u^2),$$

where μ is the grand mean and α_i, β_j, γ_k, δ_l, and ϵ_d are known as the main effects of the factors *Logarithm*, *Historical Length*, *Moving Average (MA)*, *Forecast Combinations*, and the block *Day*, respectively. For instance, the main effect δ_l measures the increase or decrease of the average response of the *Forecast Combinations* $l = \{1,2,3,4,5,6\}$ with respect to the average level. Similar interpretations apply for the rest of the effects. This is related to the restrictions $\sum_{i=\{\mathrm{No},\mathrm{Yes}\}} \alpha_i = 0$, $\sum_{j=\{308,548\}} \beta_j = 0$, $\sum_{k=\{\mathrm{No},\mathrm{Yes}\}} \gamma_k = 0$, $\sum_{l=1}^{6} \delta_l = 0$, and $\sum_{d=1}^{D} \epsilon_d = 0$ (D represents the total number of days with forecasts, this is $D = 1\,767$ for the Iberian data and $D = 1\,219$ for the Italian data).

The noise term u_{ijkld} includes all that is not explicitly taken into account in the model, but that somehow is able to explain some of the variability of the response variable MAE_{ijkld}.

Since it is assumed that the error term u_{ijkld} is Gaussian, independently and identically distributed, with zero mean and variance σ_u^2, once the model has been estimated, a diagnostic checking must be performed, testing that the $\widehat{u}_{ijkld}$ are homoskedastic, Gaussian and independent, where

$$\widehat{u}_{ijkld} = \mathrm{MAE}_{ijkld} - \widehat{\mu} - \widehat{\alpha}_i - \widehat{\beta}_j - \widehat{\gamma}_k - \widehat{\delta}_l - \widehat{\epsilon}_d.$$

The ANOVA is conducted for each forecasting horizon and the results are summarized in the next Section 2.4.2.1 and fully described in Appendices A.1 and A.2. In all the cases, the response variable was transformed after a first attempt to fit a model to the MAE_{ijkld} because the residuals were heteroskedastic. The results shown hereafter consider the $\ln(\mathrm{MAE}_{ijkld})$ as the response variable. Given that the logarithm is a monotonically increasing function, the results can be interpreted directly, and the best model corresponds to the smallest $\ln(\mathrm{MAE}_{ijkld})$, while the worst model to the largest $\ln(\mathrm{MAE}_{ijkld})$.

Often, in practice the Gaussianity assumption for the ANOVA residuals does not hold. Therefore, the p-values to assess whether the Design of Experiment factor effects are significant or not are recalculated employing bootstrap, following Davison and Hinkley (1997). Likewise, the confidence intervals for the mean of the main effects are obtained employing bootstrap. See Appendix A.3 for further detail on the bootstrap procedures employed.

2.4.2.1 Summarizing the Conclusions from the ANOVA

For the Iberian data-set, for all the forecasting horizons considered, taking *Logarithm* of prices does not make a difference in performance. Regarding the *Historical Length* of the data, the short window of 308 days is preferred for the forecasting horizons of 1 and 7-day-ahead (forecasts for the short term) while the long window of 548 days is preferred for the forecasting horizons of 30 and 60-day-ahead (long term forecasts); this is consistent with the results in Alonso et al. (2011). Furthermore, the *Moving Average* terms for the factor models are statistically significant for all forecasting horizons, which means that modeling the common factors as ARIMA reduces the error in comparison to modeling them as ARI.

Regarding the *Forecast Combinations*, the median-based combination, mean-based combination and mean BIC-based combination result in better forecasts than the benchmark BIC-selected model and the other combinations available for most forecasting horizons ($h = 7$ onward). However, it is not clear that one of these three is best: the confidence intervals for the median-based combination, mean-based combination and mean BIC-based combination usually overlap, indicating no significant difference between them.

In the case of the Italian data-set, employing *Logarithm* = Yes contributes to reduce the forecasting error for all the horizons considered. The *Historical Length* behaves differently than the way it does for the Iberian data-set for $h = 30$, for which case it is convenient to set it to 308 days instead of the 548 days that are suggested for the peninsula. *MA* is also a factor which contributes to reduce the forecasting error, for any forecasting horizon considered. As it occurs with the results for the Iberian electricity prices, *Forecast Combinations* 2, 3 and 6 reveal better results than the benchmark, and it is also not clear that any of the three would be better than the other two in all scenarios. On the contrary, *Combinations* 4 and 5 fail to outperform the benchmark, specially for the long-term forecasting horizons.

For details of the ANOVA results for each forecasting horizon see Appendices A.1 and A.2 for the Iberian and Italian markets, respectively.

2.4.3 Electricity Prices Forecasting

In this section, the results of the *Forecast Combinations* are presented, in comparison with the best model selected by the BIC information criterion[6]. Forecasts are calculated for a long period, for each day and hour. The next paragraphs describe technical details involved in estimation. Subsection 2.4.3.1 sheds light on the results involved in the estimation of each rolling window and Subsection 2.4.3.2 presents the results for all days and hours for up to 60-day-ahead forecasts.

Prices are transformed using logarithms to mitigate the existing heteroskedasticity, present in most commodity prices' time series. Therefore, the series modeled are $\boldsymbol{y}_t = ln(\boldsymbol{P}_t + k)$, where $\boldsymbol{P}_t$ represents the vector of 24 prices for day t and $k = 1\ 000$.

Most of the literature focuses on short-term forecasts (Nogales et al., 2002), but we focus on medium- and long-term forecasting in order to provide with complementary insights. In this case, forecasts of specific components are negligible. Therefore, we do not model these, but only the unobserved common factors, which explain the larger portion of the variability and capture the trend of the series in the long-run. This is in line with the results in Alonso et al. (2011). The prediction horizon will vary from 1 to 60 days, and once the factor(s) are modeled and predicted, the loading matrix is used to obtain the forecasts of the original 24-dimensional vector of prices. Then, the out-of-sample performance of the forecasts is evaluated.

We work with rolling windows of *Historical Length*= 548 days, the best length for medium- and long-term forecasts according to the previous section;

[6]This model may have one or two common factors for the Iberian data and a third factor as well for the Italian case, and each common factor extracted is modeled with a seasonal ARIMA model with parameters selected by BIC. The model is selected anew in each window.

and estimate one and two common factors for the Iberian data-set and also a third factor for the Italian data-set.

In each window, 36 alternative seasonal ARIMA$(p, d, q) \times (P, D, Q)s$ models are estimated for each factor: $p = \{1,2,3\}$, $d = \{0\}$, $q = \{1,2,3\}$, $P = \{0,1\}$, $D = \{1\}$, $Q = \{0,1\}$. Weekly seasonality is included in the model, $s = 7$, but yearly seasonality is not. This follows Alonso et al. (2011), who found no improvement in the prediction error when modeling yearly seasonality in the Iberian market, using a similar length of time for the estimation.

Therefore, in the Iberian case there are 36 models that use only one factor and 1 296 models that use two factors; a total of 1 332 different models, depending on how many factors they include and the parameters of the ARIMA$(p, d, q) \times (P, D, Q)s$. For the Italian data-set, because up to three common factors are estimated, the total number of models for any rolling window is 47 988 (36 models that use only one factor, 1 296 models that use two factors, and 46 656 models that use three factors). These figures make it unfeasible to check the residuals' behavior for each ARIMA model estimated; notwithstanding, TRAMO, the software employed to calculate the ARIMA models, estimates the p-value of the Ljung Box statistic for each model and shows acceptable values for most cases. Additionally, three of the five *Forecast Combinations* under consideration are based on BIC, so "badly" behaved models (poor fit will be associated to a high residuals variance) will be assigned small or negligible weights in the final forecast. Furthermore, the median-based combination is not affected by outliers due to "badly" behaved models. Only the mean combination may be affected by them but, based on Tables 2.3 and 2.4, median and mean combinations reveal similar results. If there were fewer models or the analyses were limited to a shorter period, residual checking could be performed before forecast averaging. In that case, it would be reasonable to obtain slightly better results.

Notwithstanding the large number of models, the estimation for each individual window of time takes only a few minutes; therefore, the procedure could be used in real time. Additionally, even though with such a large number of models some will be superfluous, the combinations that use weights depending on the BIC will assign them nearly null weights.

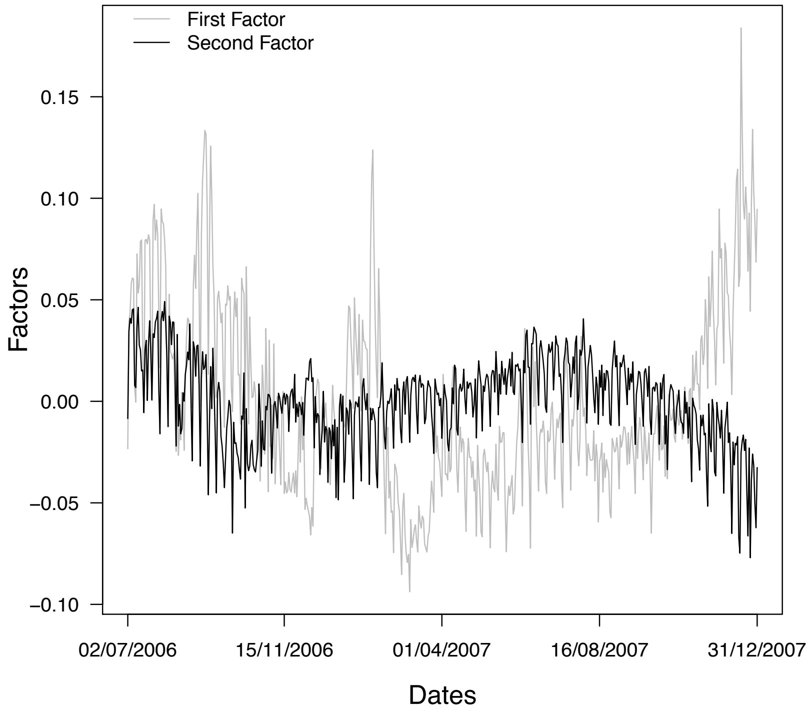
For the Iberian market, the complete data-set comprises the period July, 2006, to December, 2012. The data before 2008 is only used as historical data, therefore the first predicted day is January 1st, 2008 and the last one December 31st, 2012. Thus, there is a total of 1 767 time rolling windows, corresponding to 1 827 days in January 1st, 2008-December 31st, 2012 minus 60 days needed for out-of-sample data (used to compare with up to two-month-ahead forecasts). This data is provided by the market operator (See Conejo et al., 2010a, for details on the role of the market operator), OMIE (Iberian Market Operator). For the Italian market, on the other hand, the complete data-set includes January, 2008 to December, 2012. Therefore, the first predicted day is July 2nd, 2009 and the last one December 31st, 2012. There are 1 219 rolling windows in total corresponding to 1 279 days in July 2nd, 2008-December 31st, 2012 minus 60 days needed for out-of-sample data. The prices for the Italian electricity market are available in the website of the market operator, GME (The Energy Markets Operator).

2.4.3.1 Illustration in a Single Forecasting Window

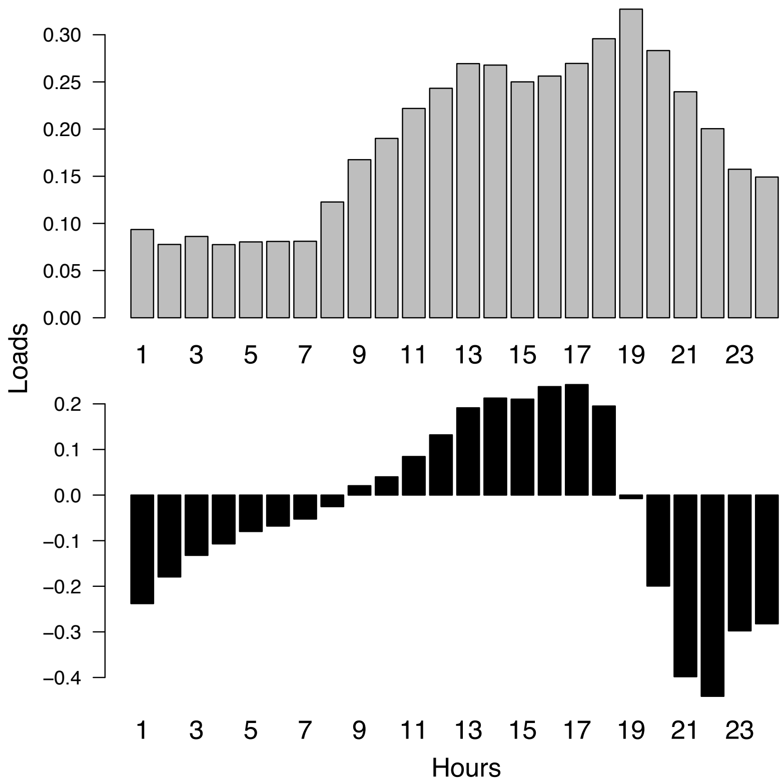
Before proceeding with the presentation of the results, this subsection is used to gain insight into the role of the common factors, as well as the forecasting combinations. With this aim, the estimation for one window of the Iberian data-set is analyzed in further detail.

The role of the underlying factors is hereby clarified. Considering as an example the first rolling window in the estimation for the Iberian data-set, Figure

2.2 presents the first and second common factors, as well as the weights assigned to them for each of the 24 hours. For the first factor, which explains 64.6% of the data variability, weights are heavy from hours 8 to 24, when most people are awake. Then, it is possible to interpret that this factor mainly records the general behavior of prices during hours when people are awake. On the other hand, the weights of the second factor are positive from 9 to 18 and negative otherwise. This coincides with usual working hours or, alternatively, sunlight hours. Therefore, the second factor, which accounts for 13.8% of the data variability, would capture changes in the price relation of working vs. non working hours. In a way, the factors are distinguishing between peak and off-peak hours, which sometimes are modeled as two groups (Conejo et al., 2010b). Notice that some models would include only the first factor, while others will include the first and second common factors. Models with more factors have been excluded from the analysis for the Iberian electricity prices (setting $r \leq 2$) because already around 80% of the data variability is explained by two factors.



(A) First and second estimated common factors.



(B) Estimated loads for first (top) and second (bottom) common factors.

FIGURE 2.2: Common factors and their weighs, first rolling window of the Iberian data-set (*Historical Length*= 548 days).

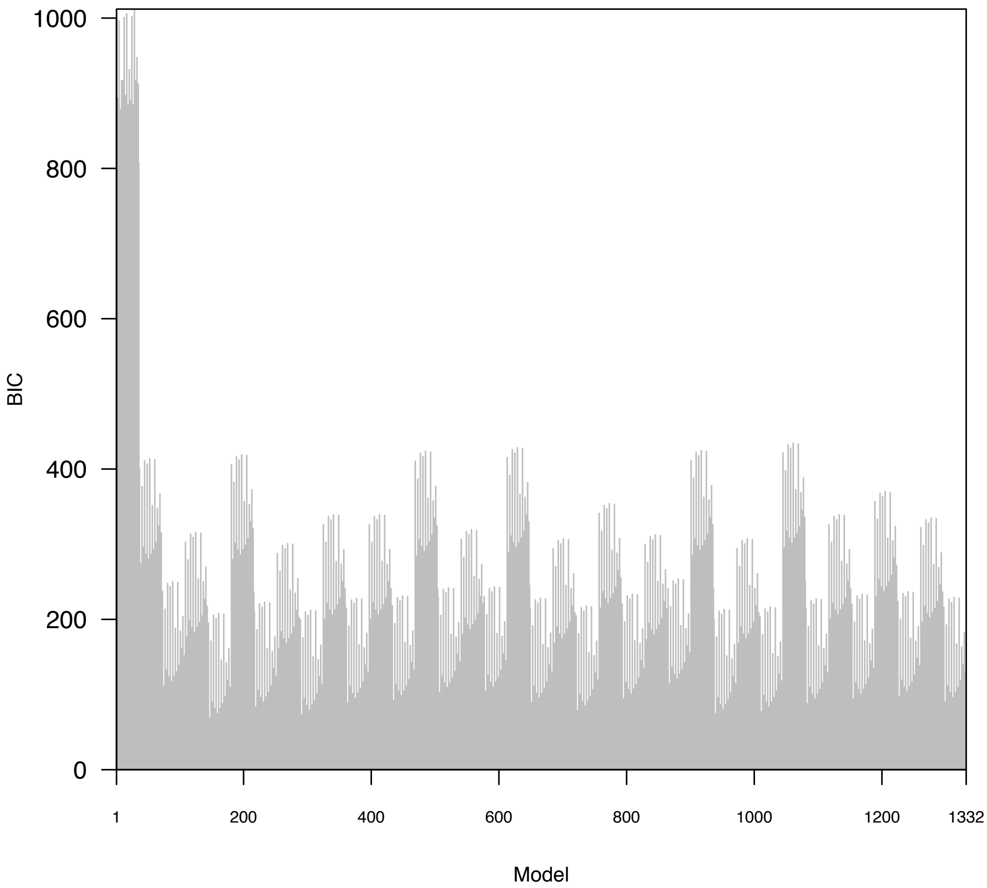


FIGURE 2.3: BIC criteria for models with one and two common factors of the first rolling window of the Iberian data-set (*Historical Length* = 548 days).

The massive estimations performed make it unfeasible to provide with the estimation results for each of the 1 332 models and for each of the 1 767 rolling windows of time. However, as an illustration, for the first rolling window, taking for example the first factor and the ARIMA model of order $p = 1$, $q = 1$, $P = 0$, $Q = 0$, the coefficients are the following: ϕ =0.7145, θ =-0.1319, both significant.

To shed some light on how the alternative models enter the combinations, Figure 2.3 presents the BIC values for the 1 332 previously mentioned models.

For illustrative purposes, also the first rolling window of the Iberian data is employed. The horizontal axis corresponds to the indexes of the models. The first 36 values (X axis from 1 to 36) represent models with only one factor ($r = 1$), starting with parameters $p = 1$, $q = 1$, $P = 0$, $Q = 0$ for X axis= 1, then $p = 1$, $q = 1$, $P = 0$, $Q = 1$ for X axis= 2, until $p = 3$, $q = 3$, $P = 1$, $Q = 1$ for X axis= 36. X axis 37 to 1 336 correspond to models with two factors ($r = 2$), we can see an important reduction of the BIC for these models. In X axis= 1 336 the order of the ARIMA models for the two factors coincide, $p = 3$, $q = 3$, $P = 1$, $Q = 1$. For BIC dependent combinations, the smaller the BIC value, the greater that model's weight. In this way, better performing models are rewarded. It is clear that there are some models with predominant low BIC (i.e. high weights). Of course, if all considered models had poor goodness of fit, then it would be reasonable that the combinations would inherit that bad performance. The claim in this work is that those combinations would be, at least, as good as the best considered model.

2.4.3.2 Forecasting Results

Given that, according to the ANOVA, many of the combinations resulted significantly better than the benchmark, in this section we study more closely those improvements. As it was previously indicated, in this section we work with DOE's factors that have the following characteristics: *Logarithm* = Yes, *Historic Length* = 548 days, and *MA* = Yes. We will consider all the *Combinations*.

Tables 2.1 and 2.2 present some descriptive statistics of interest for the daily average MAE (expression (2.8)). As expected, the error increases with the forecasting horizon h, for all the forecasts available and for both markets.

For $h = 1$, the *Combinations* do not usually do better than the benchmark model, either comparing means or quartiles. Though this may seem contradictory to the ANOVA's findings, it is not: as we explained in subsection 2.4.3.1, we are not employing the suggested values for the DOE factors for short-run forecasts. The ANOVA's outcomes indicate an advantage of using the short *Historic Length*, which we do not do here. The reason for this is to focus on the performance of a particular specification, which in this case reflects an interest in medium- and long-run forecasts rather than short-run forecasts.

On the contrary, we find that, for longer forecasting horizons ($h \geq 7$), the *Combinations* consistently render smaller errors than the benchmark. In particular, *Combinations* 2, 3 and 6 perform well in these horizons, for both data-sets.

For evaluating the performance of the *Combinations* in direct comparison to the benchmark we use the relative MAE (RelMAE). This is presented in Figures 2.4a and 2.4b. For most forecast horizons considered, all combinations of forecasts included hereby outperform the best factor model selected

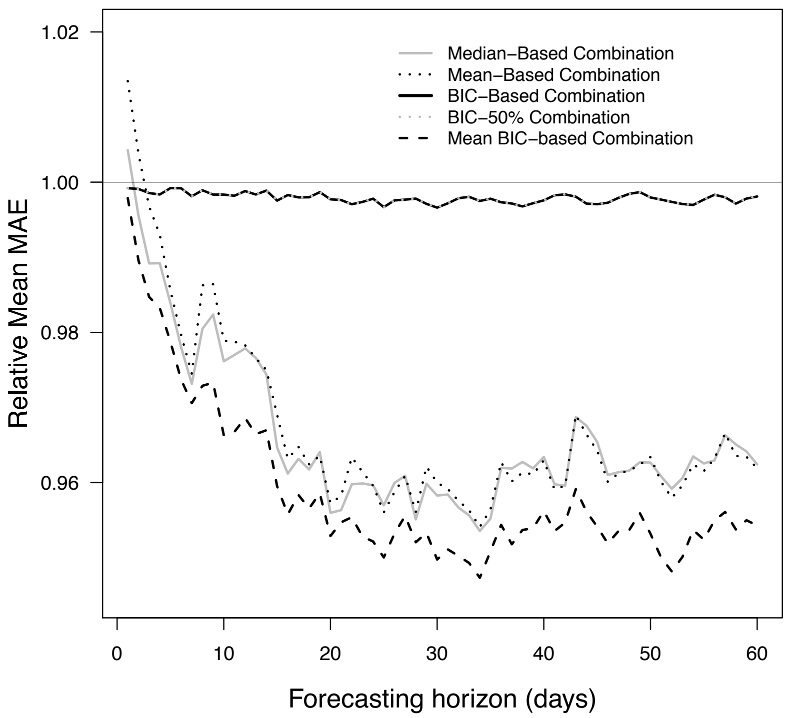
TABLE 2.1: Descriptive statistics for MAE. Iberian Market.

Forecasting horizon		BIC-selected model	Median-based Combination	Mean-based Combination	BIC-based Combination	BIC 50% Combination	Mean-BIC-based Combination
	mean	5.3389	5.3617	5.4107	5.3346	5.3346	5.3276
h=1	Q1	3.2591	3.2709	3.2850	3.2632	3.2632	3.2496
	Q2	4.5014	4.5198	4.5973	4.4970	4.4970	4.4989
	Q3	6.3009	6.3391	6.3991	6.2910	6.2910	6.2734
	mean	6.3261	6.1562	6.1648	6.3142	6.3142	6.1400
h=7	Q1	3.6423	3.5485	3.5326	3.6457	3.6457	3.5344
	Q2	5.1323	4.9770	4.9612	5.1261	5.1261	4.9754
	Q3	7.4496	7.4044	7.3966	7.4500	7.4500	7.3105
	mean	7.8677	7.5395	7.5531	7.8411	7.8411	7.4725
h=30	Q1	4.4447	4.2406	4.2163	4.4416	4.4416	4.1877
	Q2	6.3851	6.1435	6.1057	6.3668	6.3668	6.0195
	Q3	9.8775	9.5081	9.5525	9.8372	9.8372	9.3183
	mean	9.5120	9.1545	9.1496	9.4938	9.4938	9.0772
h=60	Q1	5.2904	4.8470	4.8116	5.2651	5.2651	4.7933
	Q2	7.7493	7.2159	7.2676	7.7290	7.7290	7.2704
	Q3	11.5845	11.4665	11.4378	11.5386	11.5386	11.2074

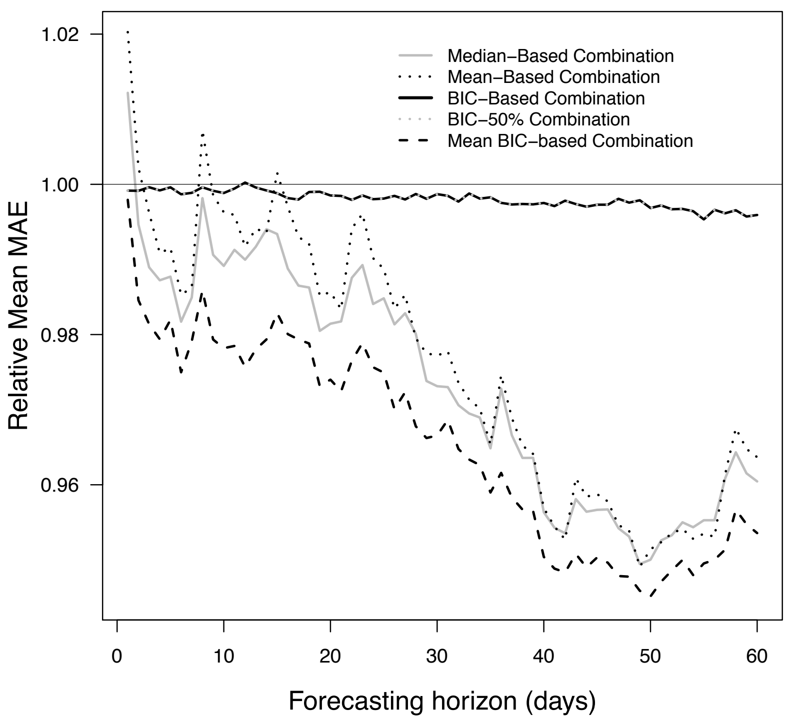
TABLE 2.2: Descriptive statistics for MAE. Italian Market.

Forecasting horizon		BIC-selected model	Median-based Combination	Mean-based Combination	BIC-based Combination	BIC 50% Combination	Mean-BIC-based Combination
	mean	7.7263	7.8204	7.8829	7.7198	7.7198	7.7104
h=1	Q1	4.9388	4.9486	4.9366	4.9253	4.9253	4.8800
	Q2	6.5024	6.5837	6.5857	6.5053	6.5053	6.4864
	Q3	9.0010	9.2146	9.3917	9.0167	9.0167	9.0471
	mean	8.9553	8.8204	8.8323	8.9451	8.9451	8.7682
h=7	Q1	5.3898	5.2714	5.2710	5.3849	5.3849	5.3467
	Q2	7.3988	7.2220	7.2518	7.4005	7.4005	7.2698
	Q3	10.4787	10.3937	10.3449	10.5152	10.5152	10.1510
	mean	10.6954	10.4080	10.4516	10.6815	10.6815	10.3378
h=30	Q1	6.3428	6.1296	6.1021	6.3335	6.3335	6.1352
	Q2	8.9068	8.4948	8.5843	8.9069	8.9069	8.4545
	Q3	13.1087	12.5893	12.7182	13.0952	13.0952	12.4662
	mean	12.2726	11.7872	11.8268	12.2225	12.2225	11.7028
h=60	Q1	7.3961	7.2301	7.3933	7.4088	7.4088	7.2269
	Q2	10.6249	10.0720	10.0532	10.5705	10.5705	10.0262
	Q3	15.1359	14.2388	14.1938	14.9273	14.9273	14.1697

by the BIC (RelMAE$<$ 1). Though for the very short time, the median-based and mean-based *Combinations* are worse than the benchmark, they outperform it in the medium- and long-run, which is the aim of this section's exercise. The mean BIC-based Combination is better than the others for all forecasting horizons and for both data-sets. After a certain h, the performance of the *Combinations* relatively to the benchmark becomes stable as the forecasting horizon increases, an advantage when the focus is obtaining accurate forecasts in the long-run. On the contrary, the vast majority of methods proposed in the literature are only appropriate for short-term forecasting, because their performance dramatically degrades when extending the forecasting horizon.



(A) Iberian market.



(B) Italian market.

FIGURE 2.4: Relative MAE, forecast horizon from 1- to 60-day-ahead. Values smaller than one indicate a result outperforming the benchmark.

TABLE 2.3: Weekly, monthly, and bi-monthly MAE and MedAE for the Iberian Market.

	BIC-selected model	Median-based Combination	Mean-based Combination	BIC-based Combination	BIC 50% Combination	Mean-BIC-based Combination
Weekly						
MAE	5.9455	5.8690	5.8965	5.9384	5.9384	5.8397
MedAE	5.3515	5.2433	5.2677	5.3444	5.3444	5.2275
Monthly						
MAE	6.9069	6.6952	6.7097	6.8934	6.8934	6.6526
MedAE	6.3635	6.1179	6.1367	6.3484	6.3484	6.0882
Bi-Monthly						
MAE	7.8184	7.5456	7.5512	7.8014	7.8014	7.4867
MedAE	7.3047	7.0081	7.0157	7.2844	7.2844	6.9539

TABLE 2.4: Weekly, monthly, and bi-monthly MAE and MedAE for the Italian Market.

	BIC-selected model	Median-based Combination	Mean-based Combination	BIC-based Combination	BIC 50% Combination	Mean-BIC-based Combination
Weekly						
MAE	8.5488	8.4687	8.5109	8.5418	8.5418	8.3989
MedAE	7.2800	7.2110	7.2525	7.2752	7.2752	7.1404
Monthly						
MAE	9.7019	9.5744	9.6197	9.6902	9.6902	9.4794
MedAE	8.3685	8.2531	8.3021	8.3579	8.3579	8.1718
Bi-Monthly						
MAE	10.6000	10.3025	10.3310	10.5779	10.5779	10.2186
MedAE	9.1880	8.9105	8.9379	9.1656	9.1656	8.8332

Last, in Tables 2.3 and 2.4, the MAE and MedAE for the BIC-selected model and for the alternative *Combinations* are presented for weekly, monthly and bi-monthly forecasts[7]. Results are consistent with those of the previous tables. Similar outcomes were obtained with a different accuracy metric, the Root Mean Squared Error (RMSE, see Appendix A.4 for details) (Hyndman and Koehler, 2006).

In conclusion, there is an improvement in prediction when using *Forecast Combinations*, specially median-based, mean-based, and median-BIC-based *Combinations*, in comparison with the best model selected according to the BIC criterion. Even though the decrease in the errors is small, it is steady,

[7]Weekly values are obtained by dividing the average of the week (the MAE of horizons $h = 1$ to $h = 7$) by the forecasting horizon ($h = 7$). Similarly for the monthly and bi-monthly values.

supporting the conclusion obtained in the ANOVA, in which some combinations are statistically significant better than the benchmark.

2.5 Concluding Remarks

In this chapter, Factor Models and Forecast Combination techniques have been jointly employed to obtain predictions of spot market electricity prices in the Iberian and Italian Markets. The main contribution consists, therefore, of combining two streams of literature in order to obtain forecasts that outperform those resulting from the individual models. In this respect, there are three combinations that clearly outperform the benchmark: the median-based combination, the mean-based combination, and the mean BIC-based combination. This conclusion is supported by the ANOVA of the combinations for forecast horizons 1-day-ahead, 7-day-ahead, 30-day-ahead and 60-day-ahead.

In the process of trying to obtain the best possible results, different aspects of the available models were compared. In this regard, the main conclusions are that longer historic data-sets benefit longer forecasting horizons and the error is reduced by the inclusion of MA terms when modeling the unobserved factors (vs. AR models).

This application reflects how the methodology works in empirical applications. The numerical results for electricity prices, which is a difficult to predict series, are good. An effort has been made to obtain the results for many time horizons ($h = 1$ to $h = 60$), for every day and during several years and considering several models for the factors, enhancing the validity of our proposal. In order words, forecasts are obtained for the very short (one day) and short term (a few days ahead), like most of other works, as well as for the medium term, which is an extension not customary in the literature. As previously explained, this approach can be employed to obtain long term forecasts not experiencing a degradation of accuracy, which is a drawback that most applications suffer from.

Numerous lines of research remain open in relation with this topic. For instance, in this work, few techniques for combining forecasts are employed besides the mean, and weights depend on the overall performance of the particular model to be used in the combination in terms of the BIC information criterion. However, there are several other, Bayesian and classical techniques to determine such weights. In particular, it would be interesting to compare the performance of both types of techniques. Moreover, in this article we have worked with fixed weights; however, these could change in a predefined way for different forecasting horizons. Furthermore, weights could be adaptive to the performance of the models (as in Sánchez, 2006).

The use of ARIMA models for the common factors allows to maintain the number of parameters to estimate low, but it may also signify a constraint in the improvement that can be achieved from the combinations of forecasts. A future line of work would be to include other models for the factors, such as the SeaDFA, which assumes that vector F_t follows a VARIMA model, modeling heteroskedasticity in the common factors (García-Ferrer et al., 2012; Pascual et al., 2004, for obtaining prediction intervals), or even including scenarios for pool prices (Pineda et al., 2009).

It is also left for future work to incorporate in the Forecast Combination other forecasting methods, not necessarily involving FM. For example, the predictions obtained by García-Martos et al. (2007) mixed model, which presents extremely accurate short-term predictions for the Iberian market. With weights evolving for different time-horizons, including this model for short-term predictions could improve the results.

A further improvement could consist of employing explanatory variables that drive spot prices as well as other type of models. Some examples could be including data for demand, weather conditions, fuel prices, excess capacity, and prices of the futures markets (when these are developed and liquid enough). Interesting references in these fields are Bello et al. (2016), Monteiro et al. (2015), Cosmo (2015), and Conejo et al. (2010b). It would also be interesting to include models with stochastic producers: Iversen et al. (2014) and Iversen et al. (2016) forecast wind power and solar irradiance and estimates their associated uncertainty. Participation in regulation and adjustment markets is not considered in this work either (Conejo et al., 2010a; Morales et al., 2014) . However, it would be necessary to assess if the uncertainty in the prediction of the explanatory variables does not outweigh the improvement in the forecast of the price. In a similar line of research, regime switching models could be employed to deal with spikes in the price series.

Last, bootstrap procedures could be used to obtain confidence intervals of the predictions and in this way assess the uncertainty involved in the forecasts.

Chapter 3

Bias Correction for Factor Models

3.1 Introduction

Dimensionality reduction techniques have been employed for decades in the context of multiple time series data-sets because “when the series are driven by a set of common factors, (a) a large number of parameters may be needed to obtain an adequate representation of the system and (b) the estimated parameters will be highly correlated. Therefore, a complex and badly defined relationship can appear when, in fact, a simpler and parsimonious model in terms of a few common factors can be operating.” (Peña and Box, 1987).

This idea that large sets of time series can be modelled and forecast by using only a few variables that integrate information of all the data has been successfully applied in diverse research fields. Some examples are:

- Commodities' prices: Peña and Box (1987) extract common factors from wheat prices of different regions, Alonso et al. (2011), García-Martos et al. (2012), and Alonso et al. (2016) use dynamic factors for electricity prices, García-Martos et al. (2013) employ dynamic factors to model the volatility of electricity, fuel, and CO_2, and Poncela et al. (2014) work with one common factor for non-fuel commodities;

- Macroeconomic variables: Stock and Watson (2002) apply principal components to 149 economic indicators, Sargent and Sims (1977) work with index models in the context of business cycles, Forni et al. (1999) combine dynamic principal components and dynamic factor analysis to estimate economic activity indexes;

- Demographic variables: Lee and Carter (1992) use singular value decomposition to estimate indexes that help forecast age specific mortality in the US, Alonso et al. (2008) employ a dynamic factor model for mortality and fertility rates of the Spanish population.

In an iterative process, we can obtain forecasts for common factors, that allow to obtain forecasts for the original data-set or for other dependent variables. To obtain these forecasts, the common factors are modelled to follow, for instance, ARIMA models (Jeong and Bienkiewicz, 1997; García-Ferrer et al., 2011; García-Martos et al., 2012) or VAR and VARIMA models (Stock and Watson, 2002; Alonso et al., 2011; Peña and Poncela, 2006).

In this work we focus on modelling the common factors by means of AR models because we want to study one feature in particular: small sample bias-correction of nearly non-stationary AR coefficients. This aspect has been explored, among others, by Clements and Kim (2007), Roy and Fuller (2001), and Clements and Taylor (2001). However, to the best of our knowledge, it has not been studied in the context of factor models, even though it is not unusual to find highly persistent common factors. Some examples of this can be found in Alonso et al. (2011), for electricity prices of the Spanish market; and Gregory and Head (1999), for macroeconomic relations between investment, productivity and the current account in a multi-country setting. A simulation exercise will show that the improvements of employing the aforementioned corrections do not fade when the factors' forecasts are transformed back (through the estimated relation data series-factors) to become forecasts of the original time series. The proposal is illustrated with an empirical case as well, featuring the Industrial Production Index of several European countries.

The remainder of the chapter is organized as follows. Section 3.2 describes the methodology. Section 3.3 presents the experimental design, and Section 3.4 shows the results of the Monte Carlo simulation. Section 3.5 contains an empirical application of the proposed methodology. Finally, Section 3.6 concludes.

3.2 Methodology

The methodology can be summarized in two steps. Given a vector of variables $\boldsymbol{y}_t$, the first step consists of estimating the common factors. In the second step, an AR model for each factor is estimated. These AR models allow to obtain h-step-ahead (h being the forecasting horizon) forecasts[1] for the common factors, which, using the corresponding weights, are transformed to forecasts for $\boldsymbol{y}_t$.

[1]See Marcellino et al. (2006) for a comparison between iterated multi-period ahead forecasts and direct forecasts for time series.

Though we work with AR models, it would be possible to extend the technique for seasonal ARIMA models. The complete process is described in the following subsections.

3.2.1 The Factor Model

As Geweke and Singleton (1981) explain, given an observable vector of time series, the factor model determines how many common factors there are; these factors can be interpreted as latent variables underlying the covariance structure of the series.

In the factor model, a set of observed variables, $\boldsymbol{y}_t$, is decomposed into unobserved common factors $\boldsymbol{F}_t$ and specific components $\boldsymbol{\varepsilon}_t$. Let $\boldsymbol{y}_t$ be an N-dimensional observed vector of variables at time t, generated by an R-dimensional vector of unobserved common factors, with $R \ll N$. $\boldsymbol{\varepsilon}_t$, the vector of specific components or idiosyncratic errors, is also N-dimensional. The factor model can be expressed as

$$\boldsymbol{y}_t = \boldsymbol{\Omega}\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \tag{3.1}$$

where $\boldsymbol{\Omega}$ is the matrix of loads or weights and has dimension $N \times R$. It indicates the relation of the R unobserved common factors with the observed series in $\boldsymbol{y}_t$. The loadings in $\boldsymbol{\Omega}$ are unknown and we will consider only static weights (therefore, a static factor model). However, in a more general model, the effect of lagged factors may be included as well; in that case we would have a lagged polynomial matrix $\boldsymbol{\Omega}(L)$ instead, where L is the lag operator. That is the so-called dynamic factor model. Bai and Ng (2008) indicate that for empirical applications the two approaches render similar forecasts, but the static approach, for which time domain methods are employed, is easier to estimate and implies fewer decisions regarding auxiliary parameters than the dynamic approach, which is estimated employing frequency domain analysis.

There are several techniques to estimate the unobserved common factors $\boldsymbol{F}_t$. In their survey, Stock and Watson (2010) divide them in three groups: maximum likelihood by means of Kalman filter, non-parametric cross-sectional averaging, and hybrid techniques that combine the former two. See Ruiz and Poncela (2015a) for a comparison of point and interval factor estimates for these procedures.

In this work we will use principal components (adapted to time series), which is included in the second set of methods. One advantage of this methodology is that it is computationally fast. Moreover, Stock and Watson (2002) prove that the factors' estimates obtained by means of principal components are consistent, even if there is serial or cross-sectional correlation in the specific components. Stock and Watson (2010) also indicate that when the number of variables is large, the estimation of the common factors is accurate enough that it can be included as data in regressions. In this line, Ruiz and Poncela (2015a) obtain that the factors extracted by the usual alternative procedures are similar and that the accuracy of point estimates increases only marginally when adding more variables to a system (they start with $N = 11$). We will be operating in a context like this: taking the cross-section dimension of the data to be high, while varying the length of the time dimension.

We obtain the common factors $\boldsymbol{F}_t$ by means of eigen-decomposition. This way we transform a matrix of data $\boldsymbol{Y}$ of size $(T \times N)$, where T represents the time dimension of the data-set and N the cross sectional dimension, to a space with fewer dimensions, keeping those in which the data has the maximum variance. Let us recall the basics of this estimation: given $\boldsymbol{\Sigma_Y} (N \times N)$ (in practice this will be the sample variance-covariance matrix for the data-set), we find real values λ and vectors e such that

$$\mathbf{\Sigma}_Y \boldsymbol{e} = \lambda \boldsymbol{e}, \tag{3.2}$$

where λ are the so called eigenvalues of matrix $\mathbf{\Sigma}_Y$ and $\boldsymbol{e}$ are the corresponding eigenvectors. When we do the multiplication in the left side, $\mathbf{\Sigma}_Y \boldsymbol{e}$, we are transforming the points of matrix $\mathbf{\Sigma}_Y$ into a new coordinates space. The objective is to keep only a few eigenvectors so that the transformation renders a space with fewer dimensions than the original data-set ($R < N$). As a property, assuming that the eigenvalues are different, the first component has greater variance than the second component, the second component has greater variance than the third component, and so on (Mardia et al., 1979, pp. 215).

This procedure is equivalent to minimizing a loss function given by the average squared difference between the data and the commonality, $\boldsymbol{y}_t - \mathbf{\Omega} \boldsymbol{F}_t$, subject to a normalization and orthogonality of the weights (see Stock and Watson, 2010, for further details).

The estimation results in $\hat{\mathbf{\Omega}}$ equal to a matrix of the eigenvectors of $\mathbf{\Sigma}_Y$ associated to the greatest R eigenvalues. Notice that it is infeasible to separately identify the common factors and their weights. Depending on the problem at hand, it is convenient to establish constrains, either for the factors or for the weights, that solve the identification problem. For reasons that will become clear in the simulations, we will constraint the weights to be orthogonal. For other details regarding the theory of factor models see Bai and Ng (2008).

This procedure for obtaining the factors is alike the dynamic extension (incorporates time dimension) of the principal components analysis (PCA) static case described in the appendix of García-Martos et al. (2012), and employed by Stock and Watson (2002). García-Martos et al. (2012) explain that, while

Peña and Box (1987) dealt with stationary data, Lee and Carter (1992) employed non stationary data, suggesting that singular value decomposition (SVD) of the covariance matrix is used to compute the weights.

We are also proceeding similarly to Forni et al. (1999): employ principal components (PC) to separate the dynamics that create correlations in the whole panel, from the noise that characterizes each observed series and that is weakly related to the other observed series; and afterwards we incorporate the components in lieu of the factors in a factor model. However, since Forni et al. (1999) work with dynamic PC, they calculate the eigenvalues and eigenvectors of the spectral density matrix at different frequencies instead of those belonging to the data's covariance matrix.

It will not be an objective of this work to introduce methodology to model the idiosyncratic components. Therefore, we will take the specific factors ε_t as white noise. Furthermore, the specific factors' variances should be small in comparison to the variances of the common factors; otherwise they would be incorporated into the principal components (Mardia et al., 1979, pp.276).

Finally, we recur to criterion IC_3 of Bai and Ng (2002) to consistently estimate the number of factors R to keep in approximate factor models (meaning factor models in which the factors are approximated with PC). These authors define the criterion as $IC_3 = ln(V(r, \hat{\boldsymbol{F}}^r)) + r(\frac{lnC_{NT}^2}{C_{NT}^2})$, where $V(r, \hat{\boldsymbol{F}}^r)$ stands for the mean residual variance of employing r factors and $\frac{lnC_{NT}^2}{C_{NT}^2}$ is the penalty for over-fitting. $C_{NT} = min(N, T)$, where N is number of time series included and T is the series' length. Notice that we have changed their notation to be in line with the one hereby employed. Also, we use capital R to indicate the "true" (unknown) number of factors and small r when referring to the estimated number of factors.

The advantage of this criterion is that it depends on both N and T, while other criteria such as the corrected Akaike Information Criterion (AICc, Hurvich and Tsai, 1989) or the Bayesian Information Criterion (BIC, Schwarz, 1978) only include one dimension (either N or T is taken as fixed). We selected IC_3 from the criteria proposed by Bai and Ng (2002) because it had better or equal performance than the others included in that paper, for values of N, T similar to the ones we employ in the simulation (see Tables I and II of Bai and Ng, 2002). We obtain an excellent performance of this criterion in our simulations, but another option would be to use Ahn and Horenstein (2013) test, which may work better in some circumstances.

3.2.2 AR Factors

The factors $\boldsymbol{F}_t$ described in the previous section can be dynamic, following a time series model. We consider that each unobservable common factor $F_{i,t}$ is generated by an AR processes. Examples of AR common factors are given in Gregory and Head (1999) (study of the interactions of productivity, investment and current account), Peña and Safadi (2008) (five series in a model for air pollution are fitted with two AR(1) common factors), Doz et al. (2012) (the authors contemplate estimating VAR factors, but in their simulation they generate AR factors), García-Ferrer et al. (2012) (though they actually use Independent Component Analysis), García-Martos et al. (2013) (they estimate univariate GARCH models for common factors which represent volatility), and Fiorentini et al. (2016) (the authors estimate a single common factor for sectoral employment data in the United States).

The following transition equation (Gregory and Head, 1999) describes each factor:

$$F_{i,t} = \phi_1 F_{i,t-1} + \phi_2 F_{i,t-2} + ... + \phi_p F_{i,t-p} + \eta_{i,t}. \tag{3.3}$$

We consider that the roots of the AR characteristic equation lie inside the unit circle (i.e. the process is stationary). We will specially pay attention to processes for which the factors are highly persistent, though not integrated, thus our focus is on close to unit roots of the characteristic polynomial of the AR model. For procedures that deal with integrated factors see Peña and Poncela (2006).

Also, $\eta_{i,t}$ will be normally distributed. However, in Appendix B.2 we will see that this is not a restriction, and having other distribution for the errors $\eta_{i,t}$ does not alter the conclusions hereby obtained.

This particular type of model for the factors allows to maintain a low number of parameters to estimate. The AR coefficients are estimated by means of Conditional Sum of Squares (CSS) instead of Ordinary Least Squares (OLS) (Clements and Kim, 2007) in order to facilitate future extensions to include MA terms. In simulations, we assessed the estimates' distributions obtained by OLS and CSS for different values of the AR coefficients and we observed that they overlap almost completely.

In order to select the number of lags, p, in each AR model, we will compare the performance of the AICc and BIC criteria. An alternative option not explored in this work would be to employ an endogenous lag order selection algorithm that re-estimates p in each iteration of the bootstrap (Kilian, 1998b).

3.2.3 Small Sample Bias Correction

CSS estimators for AR process are consistent.[2] However, in small samples some bias and skewness are often present. We employ two approaches in order to improve the estimation for highly persistent factors. On the one hand, the bootstrap bias-corrected estimator of Clements and Kim (2007) and on the other hand, the Roy-Fuller estimator (Roy and Fuller, 2001).

Clements and Kim (2007) bootstrap bias correction can be interpreted as a constant bias correction (MacKinnon and Smith, 1998); this means that the correction depends linearly on the value of the population parameter. This is a different approach from Roy and Fuller (2001) in that Roy-Fuller's estimate is mainly a function of the unit root test statistic.

To verify the accuracy of prediction intervals obtained based on these corrections, we will perform an extensive Monte Carlo experiment in Section 3.4.

3.2.3.1 Bootstrap Bias Correction

The procedure for Clements and Kim bootstrap bias correction may be summarized in the following steps. This description follows Clements and Kim (2007) and we adapt their notation to indicate that we are doing the correction in the models for the common factors, an extension of their procedure for a single series.

This process takes place after a first estimation of the AR model for each factor by means of CSS; we identify these coefficients as $\hat{\phi}_1, \hat{\phi}_2, ..., \hat{\phi}_p$. Notice that there is a small difference of our approach with the one in Clements and

[2]Robinson (2006) shows that the CSS estimation converges a.s. in the context of long memory models. Also for long memory models, SARFIMA, Egrioglu et al. (2011) use simulation to show that CSS does better than a two-stage methodology.

Kim (2007); these authors use OLS in their estimations while we employ CSS in order to be able to incorporate MA terms in a future extension of this work. We take a shortcut and drop the sub-indexes for the factors since the procedure is the same for all of them.

1. Generate a bootstrap replica of the common factor f_t^* employing the estimated AR coefficients $\hat{\phi}_1, \hat{\phi}_2, ..., \hat{\phi}_p$, randomly selected residuals (η_t^*) and the first p estimates of the factor as starting values, $f_1, f_2, ..., f_p$.[3] We will repeat this step a number of times B.

$$f_t^* = \hat{\phi}_1 f_{t-1}^* + \hat{\phi}_2 f_{t-2}^* + ... + \hat{\phi}_p f_{t-p}^* + \eta_t^*. \qquad (3.4)$$

2. Obtain the so called bootstrap estimates by re-estimating the AR coefficients for each pseudo-data-set $f_1^*, f_2^*, ..., f_T^*$ generated in the previous step. This means we will have B values $\hat{\phi}_1^*, \hat{\phi}_2^*, ..., \hat{\phi}_p^*$.

3. Clements and Kim (2007) explain that the bias can be estimated with the formula

$$bias = mean(\hat{\phi}^*) - \hat{\phi}. \qquad (3.5)$$

   They obtain the bias-corrected estimator, $\hat{\phi}^{BC}$, by subtracting the bias from the OLS estimate (CSS for us instead here) and get

$$\hat{\phi}^{BC} = 2 \times \hat{\phi} - mean(\hat{\phi}^*). \qquad (3.6)$$

4. Last, if needed, Kilian (1998a)'s algorithm is employed in order to adjust bootstrap estimates when they fall outside the stationary region. Any of the next three situations may arise:

[3]The notation for the common factors is in lowercase to emphasize that at this point we are working with factors that are estimates themselves; in other words, for each factor, $\hat{F} = f$.

- When the original CSS estimates $\hat{\phi}$ are not stationary, we do not perform a bootstrap bias correction, thus $\hat{\phi}^{BC} = \hat{\phi}$.
- The corrected $\hat{\phi}^{BC}$ estimates should be used directly if they are stationary and the CSS $\hat{\phi}$ estimates are stationary as well.
- When the estimates of $\hat{\phi}$ are stationary but $\hat{\phi}^{BC}$ is not, then iterate i times until $\hat{\phi}_i^{BC}$ becomes stationary in the following way. We start with the values $\delta_1 = 1$, $\Delta_1 = bias$ (calculated in (3.5)) and calculate $\hat{\phi}_1^{BC} = \hat{\phi} - \Delta_1$. We will iterate i times, each time calculating $\Delta_{i+1} = \delta_i \Delta_i$, $\delta_{i+1} = \delta_i - 0.01$, $\hat{\phi}_i^{BC} = \hat{\phi} - \Delta_i$, until the estimates imply stationarity.

Kilian (1998a) shows that, because of small sample bias and skewness, bias-corrected bootstrap intervals are usually more accurate than the intervals obtained with other techniques, such as delta method, standard bootstrap, and Monte Carlo integration. This author works with bivariate models including VAR models, random walk models, and cointegrated processes, though not particularly with AR models like we do. Interestingly, Kilian (1998a) indicates that the procedure in step 4 does not have an effect asymptotically and it is not constraining the OLS estimator because it affects the estimation of the bias and does not directly affect the OLS estimate.

3.2.3.2 Roy Fuller Estimator

As an alternative, we consider the estimator developed by Roy and Fuller (2001). The explanation in this section summarizes the relevant parts of that reference for this work. These authors' purpose is to obtain an estimator which provides with considerable gains in terms of mean square error for models that are close to the unit root, while maintaining a small loss in mean square error efficiency for the remainder parameter space. According to their simulations, the bias is reduced even if the process is not highly persistent.

Roy and Fuller (2001) start with a regression that works as an ARX (Auto Regressive with Exogenous Variables) with exogenous variables given by the lagged differences of the process, as it is done to test for a unit root in an AR(p) process. Since at this point we would be working with the estimated common factors, f, for our problem the regression would be

$$f_t = \hat{\theta}_1 f_{t-1} + \hat{\theta}_2 \Delta f_{t-1} + ... + \hat{\theta}_p \Delta f_{t-p+1} + u_t, \tag{3.7}$$

where $\hat{\theta}_1 = -\sum_{i=1}^{p} \hat{\phi}_i$, $\hat{\theta}_i = -\sum_{j=i}^{p} \hat{\phi}_j$, and $\Delta f_t = f_t - f_{t-1}$. Roy and Fuller (2001)'s correction depends on the LS estimator $\hat{\theta}_1$ (in our case estimated by CSS by adding up the auto-regressive coefficients $\hat{\phi}_i$), its standard error $\hat{\sigma}_1$, the unit root test statistic $\hat{\tau} = \frac{(\hat{\theta}_1 - 1)}{\hat{\sigma}_1}$, and a function C_p that corrects the bias and adapts depending on how close to the unit root the process is. Based on their paper, for us Roy-Fuller's corrected estimate would be,

$$\hat{\theta}_1^{RF} = \hat{\theta}_1^{CSS} + [C_p(\hat{\tau}) + C_{-p}(\hat{\tau})]\hat{\sigma}_1^{1/2}, \tag{3.8}$$

where the authors have established

$$C_p(\hat{\tau}) = \begin{cases} 0 & \text{for } \hat{\tau} \leq -(k_1)^{1/2} \\ \lfloor \frac{p+1}{2} \rfloor n^{-1}\hat{\tau} - (s+1)\hat{\tau}^{-1} & \text{for } (k_1)^{1/2} < \hat{\tau} \leq K \\ \lfloor \frac{p+1}{2} \rfloor n^{-1}\hat{\tau} - (s+1)(\hat{\tau} + k_2(\hat{\tau} - K))^{-1} & \text{for } K < \hat{\tau} \leq \tau_{0.5} \\ -\tau_{0.5} + d_n(\hat{\tau} - \tau_{0.5}) & \text{for } \tau_{0.5} \leq \hat{\tau}, \end{cases} \tag{3.9}$$

$k_1 = \lfloor (p+1)/2 \rfloor^{-1}(s+1)n$, $k_2 = [(1+\lfloor (p+1)/2 \rfloor n^{-1})\tau_{0.5}(\tau_{0.5} - K)]^{-1}[(s+1) - \lfloor (p+1)/2 \rfloor n^{-1}\tau_{0.5}^2]$; $\tau_{0.5}$ is the median of τ when there is unit root; and d_n is a slope set to 0.1111 in Roy and Fuller (2001)'s simulations. Also $K = 5$ and s is the rank of exogenous explanatory variables (if any). Functions $C_p(\hat{\tau})$

and $C_{-p}(\hat{\tau})$ are defined similarly. Clements and Kim (2007) indicate that $C_{-p}(\hat{\tau})$ is close to null for most time series employed in economics because these tend to have a unit root, or be close to unit root processes. See Roy and Fuller (2001) for details on this function.

3.2.4 Complete Process: Obtaining Forecasting Intervals

For calculating forecast intervals, we employ a bootstrap procedure based on Alonso et al. (2008). We follow the same steps, but we exclude estimation and forecasts of specific factors in the factor model and we include bias corrections in the estimation of the AR coefficients.

We are using a parametric bootstrap, since we are estimating the model from the data only once and then using this model as if it were the true model.[4] Ignoring model uncertainty will not be a problem when we specify in advance the value of p, known in simulations, but can definitely affect the estimation when p is unknown.

The process can be summarized in the next steps:

1. Using multiple series in a matrix, Y, we extract common factors by eigen-decomposition of the variance-covariance matrix.

Then, we conduct steps 2 to 4 separately for each extracted common factor.

2. Estimate an AR model for each factor. This involves two steps: first, selecting $\hat{p}$, and then estimating $\hat{\phi}_0, \hat{\phi}_1, ..., \hat{\phi}_{\hat{p}}, \hat{\sigma}^2_\eta$ and $\hat{\eta}_t$. To estimate

[4]See Alonso et al. (2004) for a discussion on how to introduce model selection in the bootstrap algorithm and an assessment of results of alternative methods for the estimation of prediction intervals.

the coefficients $\hat{\phi}$ we can decide to use the small sample bias correction methods outlined in the previous section.[5]

3. Residuals re-sampling: in this step the residuals are centred. We name their distribution $\Phi_{\hat{\eta}}$. To avoid excessive notation, we do not use superscripts, but it should be clear that if we use bias correction ($\hat{\phi}^{BC}$) then the residuals will correspond to these coefficients, with a distribution $\Phi_{\hat{\eta}^{BC}}$, and analogously for Roy-Fuller's correction. We draw a random sample from the residuals' distribution function $\Phi_{\hat{\eta}}$ for $t = T+1, ..., T+H$, H being the maximum forecasting horizon considered.

4. Recursively generate factor's forecasts by using the AR estimated coefficients (with or without correction of the bias), the re-sampled residuals $\hat{\eta}_{T+h}$, and the last values for the common factor $f_{T-p+1}, ..., f_T$ (See Pascual et al., 2004, for bootstrap estimates not conditional on the last p observations of the process).

$$\hat{f}_{T+h} = \hat{\phi}_1 \hat{f}_{T+h-1} + \hat{\phi}_2 \hat{f}_{T+h-2} + ... + \hat{\phi}_p \hat{f}_{T+h-p} + \hat{\eta}_{T+h}. \tag{3.10}$$

Notice that by using the last values of f we are conditioning on the "observed"[6] sample realization (following Pascual et al., 2001).

Steps 3 and 4 are carried out B times ($B = 500$ in our simulation study) and they render an empirical forecast distribution for each factor, Φ_f. We employ Efron percentiles to obtain prediction intervals for f_{T+h}, $h = 1, ..., H$. Therefore, for a nominal coverage of $(1-\alpha)$ and forecasting horizon h, the interval for factor f is given by $[\Phi^{-1}_{\hat{f}_{T+h}}(\alpha/2), \Phi^{-1}_{\hat{f}_{T+h}}(1-\alpha/2)]$. As an advantage, this bootstrap approach does not assume normality in the errors of the

[5] If so, we adapted package BootPR in the software R to use CSS to get the bias corrected $\hat{\phi}$.

[6] "observed" is between quotes because the factors are not actually observed, but they themselves are estimates.

models for the factors (Fresoli et al., 2015, for an assessment of the effect of this assumption in forecasting densities of VAR(2) models with χ^2 errors).

5. Calculate forecasts for the series using the forecasts for each factor and the estimated weights (equation 3.11, in vector notation). We also obtain prediction intervals for the series employing Efron percentiles.

$$\hat{\boldsymbol{y}}_{T+h} = \hat{\boldsymbol{\Omega}}\hat{\boldsymbol{f}}_{T+h} + \boldsymbol{\varepsilon}_{T+h}, \tag{3.11}$$

where $\hat{\boldsymbol{y}}_{T+h}$ is a vector that contains the forecasts for the N series, $\hat{\boldsymbol{\Omega}}$ is the $(N \times r)$ estimated matrix of loadings, and $\hat{\boldsymbol{f}}_{T+h}$ the forecasted factors $(r \times 1)$.

3.3 Experimental Design

We employ simulated data-sets to illustrate the performance of the methodology. The data matrix $\boldsymbol{Y}$ has dimension $N \times T$, $N = 25$ being the number of time series included and $T = 50,\ 100,\ 200$, the time dimension. Notice that T assumes three alternative values in order to allow comparison of the estimation's performance for different sample sizes. As Clements and Kim (2007) indicate, we expect the bias in the estimates of the AR coefficients to be worse the smaller the series' length. The data matrix $\boldsymbol{Y}$ results from pre-specified AR common factors, orthonormal weight vectors (following Stock and Watson, 2002), and normally distributed idiosyncratic errors $\varepsilon_t \sim N(0, 0.1)$ with mean and standard deviation values like Alonso et al. (2011). In particular, we start with weights following model 2 of Alonso et al. (2011)[7] and then transform them into orthonormal vectors to obtain

[7] Alonso et al. (2011) loading matrix $\boldsymbol{\Omega}$ is made up of vectors $\boldsymbol{\omega}_1 = [1,1,1,1,2,1,1,1,1,-0.5,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1]^T$ and $\boldsymbol{\omega}_2 = 0.3 \times [0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,3]^T$.

the following weights:

$$\begin{pmatrix} \boldsymbol{\omega_1} \\ \boldsymbol{\omega_2} \end{pmatrix}^T = \begin{pmatrix} 0.19 & 0 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.38 & -0.31 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ -0.1 & 0.51 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.02 \\ 0.19 & 0.72 \end{pmatrix}$$

Even if Principal Components allows to identify the factor space by assuming orthonormal common factors, we need to impose the condition that the first load of the second factor equals 0 in order to identify the individual common factors in a two-factor model and model each of them as independent AR(p) processes. We experimented with alternative, in particular more diverse weights, and obtained similar results.

Additionally, the factors are created with no cross correlation between them and the variance pertaining to the first factor is strictly greater than the one of the second factor $\sigma^2_{F1} > \sigma^2_{F2}$. The persistence in the factors will be such that their variances will maintain this relation. This is in line with assumption F1.b of Stock and Watson (2002) : "$E(\boldsymbol{F}_t\boldsymbol{F}'_t) = \boldsymbol{\Sigma}_{FF}$, where $\boldsymbol{\Sigma}_{FF}$ is a diagonal matrix with elements $\sigma_{ii} > \sigma_{jj} > 0$ for $i < j$".

We consider AR models of orders one and two for the common factors. The data generating process is deliberately simple in order to easily compare the effects of sample size, as well as the proximity to unit root.

Also, $\eta_{1,t} \sim N(0,1)$ while $\eta_{2,t} \sim N(0,0.5)$, according to the principle that the variance of the second factor should be smaller than the variance of the first factor.

We perform a Monte Carlo study (10 000 trials) to evaluate the benefits of small sample bias correction. To separate the sources of uncertainty (similarly to Ruiz and Poncela, 2015b), the estimations will be evaluated in these two situations:

- the number of factors and their AR order are known,
- the number of factors and their AR order are unknown.

Even though known number of factors and AR order is an infeasible scenario, it will allow us to isolate any influence the estimation of these parameters may cast. The forecasting horizon ranges from 1 to 10. The bootstrap for the AR coefficients (in the bootstrap bias-corrected approach) and for the prediction intervals (all approaches) are based on 500 replications.

We compare the performance of not using a bias correction (denoted as *none* in the tables), the bootstrap bias-corrected estimator of Clements and Kim (2007) (denoted as BC), and Roy-Fuller estimates of Roy and Fuller (2001) (denoted as RF). We evaluate their performance for a 95% nominal coverage.

In order to assess performance, for the prediction intervals we obtain average coverage rates (C_m), average length (L_m), and CQ_m (a measure combining C_m and L_m) introduced in Alonso et al. (2002). The coverage rates are estimated as the average of the Monte Carlo trials coverage rates for the prediction intervals. The effective coverage rate in each Monte Carlo trial is the relative frequency indicating the proportion of "true observations" included in the bootstrap interval. These "true processes" or continuations are created following Alonso et al. (2002). Furthermore, like these authors, we calculate a "theoretical" interval length (L_t) that can be used for comparison. Last, CQ_m is calculated as $CQ_m = |1 - C_m/C_t| + |1 - L_m/L_t|$, where C_t is the nominal coverage and L_t the estimated theoretical mean interval length (Alonso et al., 2002).

3.4 Results for the Simulation

In this section we present the results for the Monte Carlo simulation. To make a clear presentation, we divide them in two parts. In the first part we present the results when the number of factors R and the factors' AR order p are known. In the second part (Section 3.4.2) we present the results when R and p are unknown and selected using IC_3 and BIC, respectively.

3.4.1 Number of Factors and AR orders Known

Firstly, we present the results for factors that follow AR(1) models. In order to ensure a higher variance of the first factor, its AR coefficient $\phi_{F1} = 0.975$ is greater than the corresponding one to the second factor, $\phi_{F2} = 0.90$, and the same for the variance of the noise ($\eta_{t,F1} = 1$ while $\eta_{t,F2} = 0.50$).

Results are obtained for sample sizes $T = 50, 100, 200$. Tables 3.1, 3.2, and 3.3 present the outcomes that correspond to five representative series ($Y1$, $Y2$, $Y5$, $Y10$, and $Y25$) out of the $N = 25$ observed series generated. The tables in appendix B.1 present results and explanations in detail for the factors.

In Table 3.1, for $T = 50$, we obtain that the coverage of the intervals, though usually well below the 95% theoretical value, is improved when using *BC* and *RF* (in comparison to *none*). Furthermore, the improvement is more noticeable the longer the forecasting horizon, i.e. $h = 10$ presents a greater improvement than $h = 1$. In this line, Clements and Taylor (2001) explain that the bias can increase with the forecasting horizon h because we power up the biased estimates to produce forecasts.

Be aware that there are very small differences in the standard errors (referred as "se"), presented between parenthesis. The average length of all the intervals, L_m, is larger when a correction is performed. Most often, the interval length for *none* underestimates the theoretical length, while bias correction renders intervals with length closer to the theoretical length reported. Last, CQ_m is never worse for the estimations with correction, with the exception of $Y25$ for $h = 1$. Recall that a value of $CQ_m = 0$ would mean a perfect estimation in the sense that both coverage and length coincide with the theoretical values.

Table 3.2 corresponds to a sample size $T = 100$. C_m of *BC* and *RF* are always better than that for *none*. And again, the improvement of using corrections is more noticeable in long ($h = 10$) than short horizons ($h = 1$). The gains in C_m of using *BC* or *RF* are smaller than those for the smaller sample size of $T = 50$, which is consistent with the idea that, the smaller the sample, the greater the bias and the more useful the role of bias correction in the AR estimates. L_m continues to be greater when a correction is performed and in most cases closer to L_t. Furthermore, in some cases, for the shortest horizons ($h = 1$) the value of CQ_m for *none* results equal than that for BC or RF.

Table 3.3 for $T = 200$ obtains that coverage C_m is always better for BC and RF than for *none*, and like in the previous cases, the improvement is more noticeable for long than for short horizons. As expected, the improvement in terms of coverage tends to be smaller (across forecasting horizons and estimation techniques) than for the smaller sample sizes. Like in the previous cases, L_m tends to be greater (and closer to L_t) when some type of correction is performed. CQ_m tends to be better (equal for $h = 1$) when performing a correction, though the improvements are usually not as good as those for smaller sample sizes. Finally, as expected, the behavior (in terms of C_m, L_m, and CQ_m) of the three procedures improves with the series' length.

Tables 3.4 to 3.6 provide with the results when the factors follow AR(2) models instead. The two roots for the characteristic equation of the factors are: $a^1_{F1} = 0.50, a^2_{F1} = 0.975, a^1_{F2} = 0.50, a^2_{F2} = 0.90$. The findings are similar to those obtained for AR(1) models. As before, the improvements from using small sample bias corrections deteriorate as the sample size increases from $T = 50$ to $T = 200$. Again, the improvements from the corrections are more noticeable as the prediction horizon increases. And even though coverage is always better for the estimations with bias correction, the interval length L_m and CQ_m are sometimes similar for corrected and *none*, specially for $h = 1$.

TABLE 3.1: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created using two common factors, both following AR(1) models with normal errors and coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. $T = 50$. 95% nominal coverage.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	90.68 (0.052)	0.98 (0.001)	1.05 (0.000)	0.11
		BC	91.11 (0.049)	0.99 (0.001)	1.05 (0.000)	0.10
		RF	91.30 (0.048)	0.99 (0.001)	1.05 (0.000)	0.09
	h=5	none	85.29 (0.088)	1.67 (0.003)	2.01 (0.001)	0.27
		BC	88.91 (0.077)	1.87 (0.004)	2.01 (0.001)	0.13
		RF	90.25 (0.072)	1.92 (0.004)	2.01 (0.001)	0.10
	h=10	none	80.36 (0.112)	1.92 (0.005)	2.55 (0.001)	0.40
		BC	86.33 (0.105)	2.34 (0.006)	2.55 (0.001)	0.17
		RF	89.11 (0.094)	2.47 (0.007)	2.55 (0.001)	0.09
Y2	h=1	none	87.36 (0.059)	0.71 (0.001)	0.84 (0.000)	0.24
		BC	87.92 (0.057)	0.72 (0.001)	0.84 (0.000)	0.22
		RF	88.17 (0.055)	0.72 (0.001)	0.84 (0.000)	0.22
	h=5	none	83.73 (0.090)	1.30 (0.002)	1.63 (0.001)	0.33
		BC	87.78 (0.080)	1.45 (0.003)	1.63 (0.001)	0.19
		RF	89.54 (0.074)	1.48 (0.003)	1.63 (0.001)	0.15
	h=10	none	77.99 (0.121)	1.54 (0.004)	2.15 (0.001)	0.46
		BC	84.86 (0.114)	1.89 (0.005)	2.15 (0.001)	0.23
		RF	88.19 (0.105)	1.98 (0.005)	2.15 (0.001)	0.15
Y5	h=1	none	91.15 (0.048)	1.55 (0.002)	1.65 (0.001)	0.10
		BC	91.58 (0.045)	1.57 (0.002)	1.65 (0.001)	0.09
		RF	91.84 (0.044)	1.57 (0.002)	1.65 (0.001)	0.08
	h=5	none	84.93 (0.088)	2.77 (0.005)	3.38 (0.001)	0.29
		BC	88.79 (0.076)	3.10 (0.006)	3.38 (0.001)	0.15
		RF	90.41 (0.070)	3.18 (0.006)	3.38 (0.001)	0.11
	h=10	none	79.16 (0.116)	3.26 (0.008)	4.43 (0.002)	0.43
		BC	85.71 (0.109)	4.00 (0.010)	4.43 (0.002)	0.19
		RF	88.90 (0.097)	4.19 (0.011)	4.43 (0.002)	0.12
Y10	h=1	none	92.13 (0.051)	1.13 (0.002)	1.13 (0.000)	0.03
		BC	92.48 (0.048)	1.14 (0.002)	1.13 (0.000)	0.03
		RF	92.56 (0.048)	1.14 (0.002)	1.13 (0.000)	0.03
	h=5	none	86.74 (0.085)	1.78 (0.004)	2.05 (0.001)	0.22
		BC	89.83 (0.077)	1.99 (0.005)	2.05 (0.001)	0.08
		RF	90.67 (0.075)	2.05 (0.005)	2.05 (0.001)	0.05
	h=10	none	83.36 (0.101)	1.95 (0.005)	2.43 (0.001)	0.32
		BC	87.87 (0.099)	2.36 (0.007)	2.43 (0.001)	0.10
		RF	89.64 (0.092)	2.50 (0.008)	2.43 (0.001)	0.09
Y25	h=1	none	92.78 (0.048)	1.66 (0.003)	1.64 (0.001)	0.04
		BC	93.14 (0.045)	1.68 (0.003)	1.64 (0.001)	0.04
		RF	93.25 (0.044)	1.68 (0.003)	1.64 (0.001)	0.05
	h=5	none	86.66 (0.085)	2.66 (0.005)	3.07 (0.001)	0.22
		BC	89.83 (0.077)	2.97 (0.007)	3.07 (0.001)	0.09
		RF	90.82 (0.073)	3.06 (0.007)	3.07 (0.001)	0.05
	h=10	none	82.85 (0.104)	2.94 (0.007)	3.71 (0.002)	0.33
		BC	87.62 (0.101)	3.56 (0.010)	3.71 (0.002)	0.12
		RF	89.66 (0.090)	3.77 (0.011)	3.71 (0.002)	0.07

TABLE 3.2: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created using two common factors, both following AR(1) models with normal errors and coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. $T = 100$. 95% nominal coverage.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	91.93 (0.037)	0.99 (0.001)	1.05 (0.000)	0.09
		BC	92.06 (0.036)	0.99 (0.001)	1.05 (0.000)	0.09
		RF	92.13 (0.036)	1.00 (0.001)	1.05 (0.000)	0.08
	h=5	none	89.84 (0.057)	1.82 (0.003)	2.01 (0.001)	0.15
		BC	91.77 (0.049)	1.94 (0.003)	2.01 (0.001)	0.07
		RF	92.26 (0.047)	1.97 (0.003)	2.01 (0.001)	0.05
	h=10	none	87.01 (0.076)	2.16 (0.004)	2.55 (0.001)	0.23
		BC	90.53 (0.067)	2.45 (0.005)	2.55 (0.001)	0.08
		RF	91.55 (0.061)	2.52 (0.005)	2.55 (0.001)	0.05
Y2	h=1	none	89.50 (0.040)	0.73 (0.001)	0.84 (0.000)	0.19
		BC	89.64 (0.039)	0.73 (0.001)	0.84 (0.000)	0.18
		RF	89.82 (0.039)	0.74 (0.001)	0.84 (0.000)	0.18
	h=5	none	89.08 (0.057)	1.43 (0.002)	1.63 (0.001)	0.19
		BC	91.21 (0.049)	1.53 (0.002)	1.63 (0.001)	0.10
		RF	91.95 (0.045)	1.56 (0.002)	1.63 (0.001)	0.08
	h=10	none	85.86 (0.082)	1.77 (0.003)	2.15 (0.001)	0.27
		BC	89.92 (0.072)	2.03 (0.004)	2.15 (0.001)	0.11
		RF	91.25 (0.066)	2.09 (0.004)	2.15 (0.001)	0.07
Y5	h=1	none	92.66 (0.033)	1.59 (0.002)	1.65 (0.001)	0.06
		BC	92.79 (0.032)	1.59 (0.002)	1.65 (0.001)	0.06
		RF	92.93 (0.031)	1.60 (0.002)	1.65 (0.001)	0.05
	h=5	none	89.89 (0.056)	3.04 (0.004)	3.38 (0.001)	0.15
		BC	91.89 (0.048)	3.26 (0.004)	3.38 (0.001)	0.07
		RF	92.54 (0.045)	3.30 (0.004)	3.38 (0.001)	0.05
	h=10	none	86.55 (0.080)	3.72 (0.007)	4.43 (0.002)	0.25
		BC	90.41 (0.069)	4.25 (0.008)	4.43 (0.002)	0.09
		RF	91.64 (0.064)	4.37 (0.008)	4.43 (0.002)	0.05
Y10	h=1	none	93.00 (0.034)	1.12 (0.001)	1.14 (0.000)	0.03
		BC	93.11 (0.034)	1.12 (0.001)	1.14 (0.000)	0.03
		RF	93.11 (0.034)	1.12 (0.001)	1.14 (0.000)	0.03
	h=5	none	90.74 (0.054)	1.91 (0.003)	2.05 (0.001)	0.12
		BC	92.40 (0.049)	2.03 (0.003)	2.05 (0.001)	0.04
		RF	92.54 (0.049)	2.05 (0.003)	2.05 (0.001)	0.03
	h=10	none	88.67 (0.067)	2.14 (0.004)	2.42 (0.001)	0.18
		BC	91.33 (0.063)	2.40 (0.005)	2.42 (0.001)	0.05
		RF	91.76 (0.062)	2.45 (0.005)	2.42 (0.001)	0.05
Y25	h=1	none	93.73 (0.032)	1.66 (0.002)	1.64 (0.001)	0.03
		BC	93.85 (0.031)	1.66 (0.002)	1.64 (0.001)	0.03
		RF	93.87 (0.031)	1.67 (0.002)	1.64 (0.001)	0.03
	h=5	none	90.84 (0.054)	2.86 (0.004)	3.07 (0.001)	0.11
		BC	92.54 (0.048)	3.05 (0.005)	3.07 (0.001)	0.03
		RF	92.76 (0.047)	3.09 (0.005)	3.07 (0.001)	0.03
	h=10	none	88.49 (0.069)	3.26 (0.006)	3.71 (0.002)	0.19
		BC	91.35 (0.063)	3.67 (0.007)	3.71 (0.002)	0.05
		RF	91.94 (0.060)	3.76 (0.008)	3.71 (0.002)	0.05

TABLE 3.3: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created using two common factors, both following AR(1) models with normal errors and coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. $T = 200$. 95% nominal coverage.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	92.55 (0.027)	1.00 (0.001)	1.05 (0.000)	0.08
		BC	92.59 (0.027)	1.00 (0.001)	1.05 (0.000)	0.08
		RF	92.59 (0.027)	1.00 (0.001)	1.05 (0.000)	0.08
	h=5	none	92.36 (0.035)	1.91 (0.002)	2.01 (0.001)	0.08
		BC	93.29 (0.032)	1.98 (0.002)	2.01 (0.001)	0.03
		RF	93.42 (0.031)	1.99 (0.002)	2.01 (0.001)	0.03
	h=10	none	90.96 (0.047)	2.33 (0.003)	2.55 (0.001)	0.13
		BC	92.81 (0.041)	2.51 (0.003)	2.55 (0.001)	0.04
		RF	93.11 (0.040)	2.55 (0.003)	2.55 (0.001)	0.02
Y2	h=1	none	90.66 (0.029)	0.74 (0.001)	0.84 (0.000)	0.16
		BC	90.69 (0.029)	0.74 (0.001)	0.84 (0.000)	0.16
		RF	90.71 (0.029)	0.75 (0.001)	0.84 (0.000)	0.16
	h=5	none	91.86 (0.034)	1.51 (0.001)	1.63 (0.001)	0.10
		BC	92.93 (0.030)	1.58 (0.001)	1.63 (0.001)	0.06
		RF	93.10 (0.029)	1.59 (0.001)	1.63 (0.001)	0.05
	h=10	none	90.44 (0.049)	1.94 (0.002)	2.15 (0.001)	0.15
		BC	92.53 (0.043)	2.10 (0.003)	2.15 (0.001)	0.05
		RF	92.91 (0.041)	2.14 (0.003)	2.15 (0.001)	0.02
Y5	h=1	none	93.47 (0.024)	1.61 (0.001)	1.65 (0.001)	0.04
		BC	93.52 (0.024)	1.61 (0.001)	1.65 (0.001)	0.04
		RF	93.51 (0.024)	1.61 (0.001)	1.65 (0.001)	0.04
	h=5	none	92.52 (0.034)	3.21 (0.003)	3.37 (0.001)	0.07
		BC	93.50 (0.030)	3.34 (0.003)	3.37 (0.001)	0.03
		RF	93.68 (0.029)	3.37 (0.003)	3.37 (0.001)	0.02
	h=10	none	90.93 (0.048)	4.05 (0.005)	4.42 (0.002)	0.13
		BC	92.89 (0.042)	4.39 (0.006)	4.42 (0.002)	0.03
		RF	93.24 (0.040)	4.47 (0.006)	4.42 (0.002)	0.03
Y10	h=1	none	93.33 (0.026)	1.11 (0.001)	1.14 (0.000)	0.04
		BC	93.36 (0.026)	1.11 (0.001)	1.14 (0.000)	0.04
		RF	93.35 (0.026)	1.11 (0.001)	1.14 (0.000)	0.04
	h=5	none	92.84 (0.034)	1.98 (0.002)	2.05 (0.001)	0.06
		BC	93.66 (0.031)	2.04 (0.002)	2.05 (0.001)	0.02
		RF	93.68 (0.031)	2.05 (0.002)	2.05 (0.001)	0.01
	h=10	none	91.68 (0.044)	2.27 (0.003)	2.42 (0.001)	0.10
		BC	93.11 (0.040)	2.41 (0.003)	2.42 (0.001)	0.03
		RF	93.21 (0.040)	2.43 (0.003)	2.42 (0.001)	0.02
Y25	h=1	none	94.09 (0.024)	1.65 (0.001)	1.64 (0.001)	0.02
		BC	94.11 (0.024)	1.65 (0.001)	1.64 (0.001)	0.02
		RF	94.11 (0.024)	1.65 (0.001)	1.64 (0.001)	0.02
	h=5	none	93.00 (0.034)	2.98 (0.003)	3.07 (0.001)	0.05
		BC	93.83 (0.031)	3.09 (0.003)	3.07 (0.001)	0.02
		RF	93.84 (0.031)	3.10 (0.003)	3.07 (0.001)	0.02
	h=10	none	91.72 (0.044)	3.48 (0.005)	3.70 (0.002)	0.10
		BC	93.26 (0.039)	3.71 (0.005)	3.70 (0.002)	0.02
		RF	93.37 (0.039)	3.74 (0.005)	3.70 (0.002)	0.03

TABLE 3.4: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created using two common factors, both following AR(2) models with normal errors. Model with coefficients ϕ_1^{F1} =1.475, ϕ_2^{F1} =-0.4875, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. $T = 50$. 95% nominal coverage.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	89.88 (0.059)	0.98 (0.001)	1.05 (0.000)	0.12
		BC	90.46 (0.056)	1.00 (0.001)	1.05 (0.000)	0.10
		RF	90.72 (0.054)	1.00 (0.002)	1.05 (0.000)	0.10
	h=5	none	83.06 (0.116)	2.75 (0.006)	3.29 (0.001)	0.29
		BC	87.28 (0.101)	3.06 (0.007)	3.29 (0.001)	0.15
		RF	88.46 (0.089)	3.06 (0.007)	3.29 (0.001)	0.14
	h=10	none	76.52 (0.144)	3.39 (0.010)	4.61 (0.002)	0.46
		BC	83.76 (0.137)	4.22 (0.013)	4.61 (0.002)	0.20
		RF	86.67 (0.113)	4.29 (0.013)	4.61 (0.002)	0.16
Y2	h=1	none	86.06 (0.065)	0.70 (0.001)	0.84 (0.000)	0.26
		BC	86.71 (0.062)	0.71 (0.001)	0.84 (0.000)	0.25
		RF	87.10 (0.061)	0.71 (0.001)	0.84 (0.000)	0.24
	h=5	none	81.90 (0.117)	2.13 (0.005)	2.65 (0.001)	0.33
		BC	86.22 (0.105)	2.37 (0.006)	2.65 (0.001)	0.20
		RF	87.96 (0.090)	2.37 (0.005)	2.65 (0.001)	0.18
	h=10	none	74.35 (0.152)	2.75 (0.008)	3.89 (0.002)	0.51
		BC	82.22 (0.146)	3.41 (0.011)	3.89 (0.002)	0.26
		RF	85.96 (0.120)	3.47 (0.010)	3.89 (0.002)	0.20
Y5	h=1	none	90.06 (0.055)	1.53 (0.002)	1.65 (0.001)	0.12
		BC	90.66 (0.052)	1.55 (0.002)	1.65 (0.001)	0.10
		RF	91.00 (0.050)	1.56 (0.002)	1.65 (0.001)	0.10
	h=5	none	82.51 (0.116)	4.55 (0.010)	5.55 (0.002)	0.31
		BC	86.83 (0.101)	5.07 (0.012)	5.55 (0.002)	0.17
		RF	88.41 (0.087)	5.06 (0.011)	5.55 (0.002)	0.16
	h=10	none	75.10 (0.148)	5.79 (0.018)	8.08 (0.003)	0.49
		BC	82.87 (0.141)	7.20 (0.023)	8.08 (0.003)	0.24
		RF	86.36 (0.115)	7.32 (0.022)	8.08 (0.003)	0.18
Y10	h=1	none	91.51 (0.061)	1.15 (0.002)	1.13 (0.000)	0.05
		BC	92.00 (0.056)	1.16 (0.002)	1.13 (0.000)	0.06
		RF	92.20 (0.055)	1.17 (0.002)	1.13 (0.000)	0.06
	h=5	none	84.21 (0.117)	2.94 (0.007)	3.36 (0.001)	0.24
		BC	88.10 (0.103)	3.28 (0.008)	3.36 (0.001)	0.10
		RF	88.78 (0.093)	3.28 (0.008)	3.36 (0.001)	0.09
	h=10	none	79.46 (0.136)	3.42 (0.010)	4.37 (0.002)	0.38
		BC	85.39 (0.130)	4.24 (0.014)	4.37 (0.002)	0.13
		RF	87.24 (0.110)	4.33 (0.014)	4.37 (0.002)	0.09
Y25	h=1	none	92.14 (0.057)	1.69 (0.003)	1.64 (0.001)	0.07
		BC	92.66 (0.053)	1.71 (0.003)	1.64 (0.001)	0.07
		RF	92.85 (0.051)	1.72 (0.003)	1.64 (0.001)	0.07
	h=5	none	84.19 (0.115)	4.41 (0.011)	5.06 (0.002)	0.24
		BC	88.16 (0.100)	4.92 (0.012)	5.06 (0.002)	0.10
		RF	88.94 (0.091)	4.92 (0.012)	5.06 (0.002)	0.09
	h=10	none	78.91 (0.136)	5.21 (0.015)	6.73 (0.003)	0.40
		BC	85.20 (0.128)	6.46 (0.021)	6.73 (0.003)	0.14
		RF	87.36 (0.109)	6.59 (0.021)	6.73 (0.003)	0.10

TABLE 3.5: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created using two common factors, both following AR(2) models with normal errors. Model with coefficients ϕ_1^{F1} =1.475, ϕ_2^{F1} =-0.4875, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. $T = 100$. 95% nominal coverage.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	91.70 (0.041)	1.00 (0.001)	1.05 (0.000)	0.08
		BC	91.91 (0.039)	1.01 (0.001)	1.05 (0.000)	0.08
		RF	91.97 (0.039)	1.01 (0.001)	1.05 (0.000)	0.07
	h=5	none	89.55 (0.067)	3.02 (0.005)	3.29 (0.001)	0.14
		BC	91.39 (0.057)	3.20 (0.005)	3.29 (0.001)	0.07
		RF	91.52 (0.054)	3.19 (0.005)	3.29 (0.001)	0.07
	h=10	none	85.91 (0.092)	3.92 (0.008)	4.61 (0.002)	0.24
		BC	89.71 (0.079)	4.45 (0.010)	4.61 (0.002)	0.09
		RF	90.43 (0.071)	4.49 (0.010)	4.61 (0.002)	0.07
Y2	h=1	none	89.11 (0.043)	0.73 (0.001)	0.84 (0.000)	0.19
		BC	89.39 (0.042)	0.74 (0.001)	0.84 (0.000)	0.18
		RF	89.50 (0.041)	0.74 (0.001)	0.84 (0.000)	0.18
	h=5	none	89.12 (0.066)	2.38 (0.004)	2.65 (0.001)	0.16
		BC	91.00 (0.057)	2.52 (0.004)	2.65 (0.001)	0.09
		RF	91.42 (0.051)	2.51 (0.004)	2.65 (0.001)	0.09
	h=10	none	85.03 (0.096)	3.25 (0.007)	3.89 (0.002)	0.27
		BC	89.14 (0.084)	3.69 (0.008)	3.89 (0.002)	0.11
		RF	90.25 (0.073)	3.71 (0.008)	3.89 (0.002)	0.09
Y5	h=1	none	92.31 (0.037)	1.59 (0.002)	1.65 (0.001)	0.06
		BC	92.52 (0.035)	1.60 (0.002)	1.65 (0.001)	0.06
		RF	92.64 (0.035)	1.60 (0.002)	1.65 (0.001)	0.05
	h=5	none	89.42 (0.067)	5.06 (0.008)	5.55 (0.002)	0.15
		BC	91.29 (0.057)	5.36 (0.008)	5.55 (0.002)	0.07
		RF	91.62 (0.052)	5.34 (0.008)	5.55 (0.002)	0.07
	h=10	none	85.39 (0.095)	6.80 (0.014)	8.08 (0.003)	0.26
		BC	89.40 (0.082)	7.73 (0.016)	8.08 (0.003)	0.10
		RF	90.39 (0.072)	7.78 (0.016)	8.08 (0.003)	0.09
Y10	h=1	none	92.84 (0.040)	1.14 (0.001)	1.13 (0.000)	0.03
		BC	93.01 (0.039)	1.15 (0.001)	1.13 (0.000)	0.03
		RF	93.00 (0.039)	1.15 (0.001)	1.13 (0.000)	0.03
	h=5	none	90.18 (0.066)	3.16 (0.005)	3.35 (0.001)	0.11
		BC	91.84 (0.058)	3.34 (0.006)	3.35 (0.001)	0.04
		RF	91.69 (0.057)	3.33 (0.006)	3.35 (0.001)	0.04
	h=10	none	87.20 (0.085)	3.82 (0.008)	4.37 (0.002)	0.21
		BC	90.37 (0.077)	4.33 (0.010)	4.37 (0.002)	0.06
		RF	90.48 (0.073)	4.37 (0.010)	4.37 (0.002)	0.05
Y25	h=1	none	93.53 (0.037)	1.69 (0.002)	1.64 (0.001)	0.05
		BC	93.66 (0.036)	1.70 (0.002)	1.64 (0.001)	0.05
		RF	93.67 (0.036)	1.70 (0.002)	1.64 (0.001)	0.05
	h=5	none	90.21 (0.065)	4.75 (0.008)	5.05 (0.002)	0.11
		BC	91.89 (0.057)	5.04 (0.008)	5.05 (0.002)	0.04
		RF	91.82 (0.055)	5.02 (0.008)	5.05 (0.002)	0.04
	h=10	none	87.03 (0.085)	5.86 (0.012)	6.72 (0.003)	0.21
		BC	90.29 (0.076)	6.63 (0.014)	6.72 (0.003)	0.06
		RF	90.59 (0.071)	6.70 (0.014)	6.72 (0.003)	0.05

TABLE 3.6: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created using two common factors, both following AR(2) models with normal errors. Model with coefficients ϕ_1^{F1} =1.475, ϕ_2^{F1} =-0.4875, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. T = 200. 95% nominal coverage.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	92.51 (0.029)	1.01 (0.001)	1.05 (0.000)	0.07
		BC	92.55 (0.029)	1.01 (0.001)	1.05 (0.000)	0.07
		RF	92.58 (0.029)	1.01 (0.001)	1.05 (0.000)	0.07
	h=5	none	92.36 (0.039)	3.15 (0.004)	3.29 (0.001)	0.07
		BC	93.14 (0.035)	3.25 (0.004)	3.29 (0.001)	0.03
		RF	93.05 (0.035)	3.24 (0.004)	3.29 (0.001)	0.03
	h=10	none	90.67 (0.053)	4.25 (0.006)	4.61 (0.002)	0.12
		BC	92.47 (0.046)	4.57 (0.007)	4.61 (0.002)	0.04
		RF	92.48 (0.045)	4.58 (0.007)	4.61 (0.002)	0.03
Y2	h=1	none	90.50 (0.030)	0.75 (0.001)	0.84 (0.000)	0.16
		BC	90.57 (0.030)	0.75 (0.001)	0.84 (0.000)	0.16
		RF	90.65 (0.030)	0.75 (0.001)	0.84 (0.000)	0.15
	h=5	none	92.16 (0.038)	2.51 (0.003)	2.65 (0.001)	0.08
		BC	92.95 (0.035)	2.59 (0.003)	2.65 (0.001)	0.04
		RF	92.94 (0.033)	2.59 (0.003)	2.65 (0.001)	0.04
	h=10	none	90.38 (0.055)	3.55 (0.005)	3.89 (0.002)	0.13
		BC	92.23 (0.049)	3.83 (0.005)	3.89 (0.002)	0.04
		RF	92.37 (0.046)	3.84 (0.005)	3.89 (0.002)	0.04
Y5	h=1	none	93.35 (0.026)	1.62 (0.001)	1.65 (0.001)	0.04
		BC	93.41 (0.026)	1.62 (0.001)	1.65 (0.001)	0.03
		RF	93.47 (0.025)	1.62 (0.001)	1.65 (0.001)	0.03
	h=5	none	92.42 (0.038)	5.33 (0.006)	5.56 (0.002)	0.07
		BC	93.19 (0.034)	5.49 (0.006)	5.56 (0.002)	0.03
		RF	93.14 (0.033)	5.48 (0.006)	5.56 (0.002)	0.03
	h=10	none	90.58 (0.054)	7.43 (0.010)	8.08 (0.003)	0.13
		BC	92.41 (0.048)	8.00 (0.011)	8.08 (0.003)	0.04
		RF	92.53 (0.045)	8.03 (0.011)	8.08 (0.003)	0.03
Y10	h=1	none	93.24 (0.029)	1.13 (0.001)	1.13 (0.000)	0.02
		BC	93.28 (0.029)	1.13 (0.001)	1.13 (0.000)	0.02
		RF	93.27 (0.029)	1.13 (0.001)	1.13 (0.000)	0.02
	h=5	none	92.56 (0.038)	3.24 (0.004)	3.35 (0.001)	0.06
		BC	93.29 (0.035)	3.34 (0.004)	3.35 (0.001)	0.02
		RF	93.11 (0.036)	3.33 (0.004)	3.35 (0.001)	0.03
	h=10	none	91.04 (0.050)	4.06 (0.006)	4.37 (0.002)	0.11
		BC	92.65 (0.045)	4.34 (0.006)	4.37 (0.002)	0.03
		RF	92.49 (0.047)	4.34 (0.006)	4.37 (0.002)	0.03
Y25	h=1	none	93.91 (0.027)	1.67 (0.001)	1.64 (0.001)	0.03
		BC	93.94 (0.027)	1.67 (0.001)	1.64 (0.001)	0.03
		RF	93.94 (0.028)	1.68 (0.001)	1.64 (0.001)	0.04
	h=5	none	92.57 (0.037)	4.89 (0.006)	5.06 (0.002)	0.06
		BC	93.29 (0.034)	5.03 (0.006)	5.06 (0.002)	0.02
		RF	93.17 (0.035)	5.02 (0.006)	5.06 (0.002)	0.03
	h=10	none	90.96 (0.050)	6.23 (0.009)	6.73 (0.003)	0.12
		BC	92.54 (0.045)	6.66 (0.010)	6.73 (0.003)	0.04
		RF	92.46 (0.046)	6.67 (0.010)	6.73 (0.003)	0.04

3.4.2 Number of Factors and AR orders Unknown

We consider a model of two factors in this experiment and use a sample size of $T = 100$ (see Appendix B.3 for other values of T). The number of factors is estimated by IC_3 of Bai and Ng (2002), as explained in Subsection 2.1. This criterion correctly estimated the number of factors R in more than 99.9% of cases.

We consider factors that are AR(2), as Clements and Kim (2007) explain, in order to allow under and over specification of p. The lag order estimated is restricted to at most six and for selection criteria we compare AICc and BIC. We did not endogenise the selection of p in the bootstrap algorithm because of the small improvements obtained by doing so in Clements and Kim (2007).

Table 3.7 presents the results when we use BIC as the criteria for selecting p and Table 3.8 presents the results when AICc is the criteria for selecting p.

In both cases, for the selected series coverage C_m, length L_m, and CQ tend to be better for the models that use bias-corrected estimators than for *none* (the correction never results in worse off results than *none*). Furthermore, we can verify the same pattern than in the previous section: improvements become more noticeable the longer the forecasting horizon.

Last, comparing the two selection criteria we can see that BIC does a much better job selecting p than AICc (see Table 3.9 for a comparison of the distribution of $\hat{p}$) and it translates in better values of C_m as well as a slight general improvement in CQ_m.

TABLE 3.7: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created with both common factors following AR(2) processes with normal errors. Model with coefficients ϕ_1^{F1} =1.475, ϕ_2^{F1} =-0.4875, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. T = 100. 95% nominal coverage. IC_3 used in the estimation of R, BIC used in the selection of $\hat{p}$.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	91.39 (0.042)	1.00 (0.001)	1.05 (0.000)	0.09
		BC	91.55 (0.040)	1.00 (0.001)	1.05 (0.000)	0.09
		RF	91.66 (0.040)	1.00 (0.001)	1.05 (0.000)	0.08
	h=5	none	89.27 (0.068)	3.01 (0.005)	3.29 (0.001)	0.15
		BC	91.07 (0.059)	3.18 (0.005)	3.29 (0.001)	0.07
		RF	91.24 (0.056)	3.17 (0.005)	3.29 (0.001)	0.07
	h=10	none	85.66 (0.094)	3.92 (0.008)	4.61 (0.002)	0.25
		BC	89.43 (0.082)	4.45 (0.010)	4.61 (0.002)	0.09
		RF	90.16 (0.073)	4.48 (0.010)	4.61 (0.002)	0.08
Y2	h=1	none	88.72 (0.045)	0.73 (0.001)	0.84 (0.000)	0.20
		BC	88.95 (0.045)	0.73 (0.001)	0.84 (0.000)	0.20
		RF	89.10 (0.043)	0.73 (0.001)	0.84 (0.000)	0.19
	h=5	none	88.71 (0.071)	2.37 (0.004)	2.65 (0.001)	0.17
		BC	90.56 (0.063)	2.51 (0.004)	2.65 (0.001)	0.10
		RF	91.02 (0.056)	2.50 (0.004)	2.65 (0.001)	0.10
	h=10	none	84.70 (0.102)	3.25 (0.007)	3.88 (0.002)	0.27
		BC	88.75 (0.091)	3.69 (0.008)	3.88 (0.002)	0.12
		RF	89.92 (0.078)	3.71 (0.008)	3.88 (0.002)	0.10
Y5	h=1	none	91.98 (0.039)	1.58 (0.002)	1.65 (0.001)	0.07
		BC	92.18 (0.038)	1.59 (0.002)	1.65 (0.001)	0.07
		RF	92.32 (0.037)	1.59 (0.002)	1.65 (0.001)	0.06
	h=5	none	89.07 (0.069)	5.04 (0.008)	5.55 (0.002)	0.15
		BC	90.91 (0.060)	5.33 (0.009)	5.55 (0.002)	0.08
		RF	91.25 (0.055)	5.31 (0.008)	5.55 (0.002)	0.08
	h=10	none	85.10 (0.099)	6.80 (0.014)	8.07 (0.003)	0.26
		BC	89.05 (0.087)	7.72 (0.017)	8.07 (0.003)	0.11
		RF	90.09 (0.075)	7.76 (0.017)	8.07 (0.003)	0.09
Y10	h=1	none	92.56 (0.041)	1.14 (0.001)	1.14 (0.000)	0.03
		BC	92.70 (0.040)	1.14 (0.001)	1.14 (0.000)	0.03
		RF	92.73 (0.040)	1.14 (0.001)	1.14 (0.000)	0.03
	h=5	none	89.89 (0.068)	3.15 (0.005)	3.35 (0.001)	0.12
		BC	91.54 (0.061)	3.33 (0.006)	3.35 (0.001)	0.04
		RF	91.46 (0.060)	3.32 (0.006)	3.35 (0.001)	0.05
	h=10	none	86.98 (0.087)	3.83 (0.008)	4.37 (0.002)	0.21
		BC	90.09 (0.080)	4.33 (0.010)	4.37 (0.002)	0.06
		RF	90.24 (0.076)	4.36 (0.010)	4.37 (0.002)	0.05
Y25	h=1	none	93.26 (0.039)	1.68 (0.002)	1.64 (0.001)	0.04
		BC	93.41 (0.038)	1.69 (0.002)	1.64 (0.001)	0.05
		RF	93.45 (0.037)	1.69 (0.002)	1.64 (0.001)	0.05
	h=5	none	89.85 (0.068)	4.75 (0.008)	5.06 (0.002)	0.12
		BC	91.55 (0.060)	5.02 (0.009)	5.06 (0.002)	0.04
		RF	91.51 (0.058)	5.01 (0.008)	5.06 (0.002)	0.05
	h=10	none	86.69 (0.090)	5.88 (0.012)	6.72 (0.003)	0.21
		BC	90.04 (0.081)	6.65 (0.015)	6.72 (0.003)	0.06
		RF	90.35 (0.074)	6.68 (0.015)	6.72 (0.003)	0.05

TABLE 3.8: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created using two common factors, both following AR(2) models with normal errors. Model with coefficients ϕ_1^{F1} =1.475, ϕ_2^{F1} =-0.4875, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. T = 100. 95% nominal coverage. IC_3 used in the estimation of R, AICc used in the selection of $\hat{p}$.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	90.73 (0.045)	0.98 (0.001)	1.05 (0.000)	0.11
		BC	90.88 (0.044)	0.99 (0.001)	1.05 (0.000)	0.11
		RF	91.00 (0.044)	0.99 (0.001)	1.05 (0.000)	0.10
	h=5	none	88.35 (0.073)	2.97 (0.005)	3.29 (0.001)	0.17
		BC	90.27 (0.065)	3.14 (0.005)	3.29 (0.001)	0.09
		RF	90.42 (0.061)	3.12 (0.005)	3.29 (0.001)	0.10
	h=10	none	84.77 (0.100)	3.90 (0.008)	4.61 (0.002)	0.26
		BC	88.59 (0.089)	4.41 (0.010)	4.61 (0.002)	0.11
		RF	89.25 (0.079)	4.39 (0.010)	4.61 (0.002)	0.11
Y2	h=1	none	87.98 (0.048)	0.72 (0.001)	0.84 (0.000)	0.22
		BC	88.27 (0.048)	0.72 (0.001)	0.84 (0.000)	0.21
		RF	88.41 (0.047)	0.72 (0.001)	0.84 (0.000)	0.21
	h=5	none	87.75 (0.076)	2.34 (0.004)	2.65 (0.001)	0.19
		BC	89.69 (0.069)	2.47 (0.004)	2.65 (0.001)	0.12
		RF	90.15 (0.062)	2.46 (0.004)	2.65 (0.001)	0.12
	h=10	none	83.76 (0.109)	3.22 (0.007)	3.88 (0.002)	0.29
		BC	87.84 (0.099)	3.65 (0.008)	3.88 (0.002)	0.14
		RF	88.95 (0.084)	3.63 (0.008)	3.88 (0.002)	0.13
Y5	h=1	none	91.35 (0.042)	1.56 (0.002)	1.65 (0.001)	0.09
		BC	91.53 (0.041)	1.57 (0.002)	1.65 (0.001)	0.08
		RF	91.68 (0.040)	1.57 (0.002)	1.65 (0.001)	0.08
	h=5	none	88.13 (0.074)	4.96 (0.008)	5.55 (0.002)	0.18
		BC	90.09 (0.066)	5.26 (0.008)	5.55 (0.002)	0.10
		RF	90.46 (0.060)	5.23 (0.008)	5.55 (0.002)	0.11
	h=10	none	84.19 (0.106)	6.75 (0.015)	8.07 (0.003)	0.28
		BC	88.20 (0.095)	7.64 (0.017)	8.07 (0.003)	0.12
		RF	89.16 (0.081)	7.61 (0.017)	8.07 (0.003)	0.12
Y10	h=1	none	91.91 (0.047)	1.12 (0.001)	1.14 (0.000)	0.04
		BC	92.00 (0.046)	1.13 (0.001)	1.14 (0.000)	0.04
		RF	92.07 (0.046)	1.13 (0.001)	1.14 (0.000)	0.04
	h=5	none	89.02 (0.076)	3.12 (0.005)	3.35 (0.001)	0.13
		BC	90.76 (0.068)	3.29 (0.006)	3.35 (0.001)	0.06
		RF	90.60 (0.066)	3.27 (0.006)	3.35 (0.001)	0.07
	h=10	none	86.09 (0.094)	3.81 (0.008)	4.37 (0.002)	0.22
		BC	89.26 (0.088)	4.28 (0.010)	4.37 (0.002)	0.08
		RF	89.27 (0.082)	4.27 (0.010)	4.37 (0.002)	0.08
Y25	h=1	none	92.55 (0.044)	1.66 (0.002)	1.64 (0.001)	0.04
		BC	92.75 (0.043)	1.66 (0.002)	1.64 (0.001)	0.04
		RF	92.80 (0.043)	1.67 (0.002)	1.64 (0.001)	0.04
	h=5	none	88.98 (0.075)	4.69 (0.008)	5.06 (0.002)	0.14
		BC	90.78 (0.066)	4.96 (0.009)	5.06 (0.002)	0.06
		RF	90.70 (0.064)	4.93 (0.008)	5.06 (0.002)	0.07
	h=10	none	85.80 (0.097)	5.84 (0.013)	6.72 (0.003)	0.23
		BC	89.17 (0.090)	6.57 (0.015)	6.72 (0.003)	0.08
		RF	89.38 (0.082)	6.55 (0.015)	6.72 (0.003)	0.08

TABLE 3.9: Comparison of relative frequencies in the estimation of $\hat{p}$ by BIC and AICc. The values correspond to a Monte Carlo simulation with 10 000 replications. Two common factors, both following AR(2) models with normal errors. Model with coefficients ϕ_1^{F1} =1.475, ϕ_2^{F1} =-0.4875, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. $T = 100$. 95% nominal coverage.

Factor	$\hat{p}=1$	$\hat{p}=2$	$\hat{p}=3$	$\hat{p}=4$	$\hat{p}=5$	$\hat{p}=6$
BIC						
F1	0.23	82.47	10.15	3.94	1.85	1.36
F2	1.84	78.47	12.32	4.01	1.95	1.41
AICc						
F1	0.02	36.66	16.17	13.50	14.13	19.52
F2	0.22	33.94	17.52	14.28	13.96	20.08

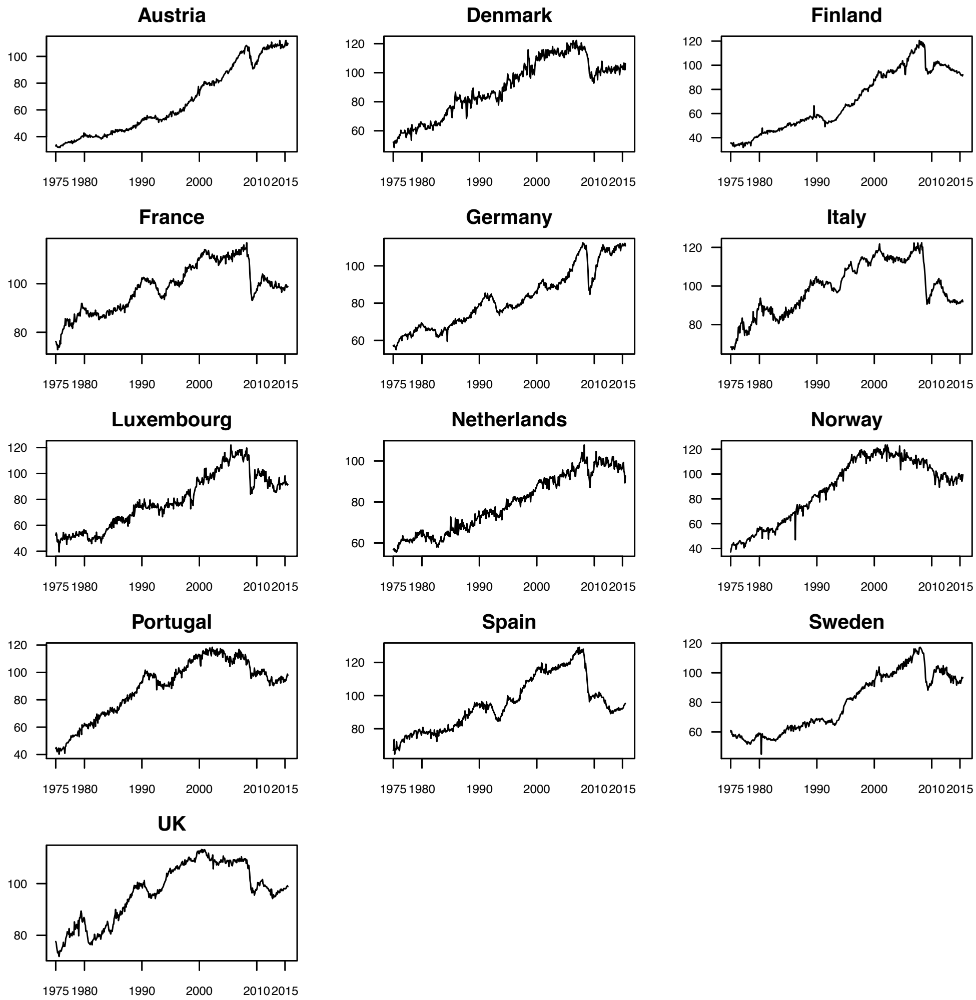
3.5 Empirical Example

As an application, we employ data of industrial production (486 seasonally adjusted monthly observations of the Industrial Production Index, IPI, from January, 1975, to June, 2015) in 13 European countries. These include Austria, Denmark, Finland, France, Germany, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, and the United Kingdom. Other European countries with available data have been excluded for having small cross correlations with the former. The data was obtained from OECD Statistics. See Figure 3.1 for a graph of the series included in the analysis.

In order to compare the results of the corrections we start with a rolling window of length $T = 50$ and forecast from $h = 1$ to $h = 12$ steps ahead. This means that, for the vector of 13 countries, the first window starts from the first observation in the data-set (January, 1975), until $T = 50$ (February, 1979). We work with this window to extract common factors[8], specify an AR model for each factor, and generate forecasts for the next 12 observations (March, 1979, to February, 1980). Repeating this process to the last window

[8]For an assessment of the factors' loads see Figure 3.3a in Subsection 3.5.1.

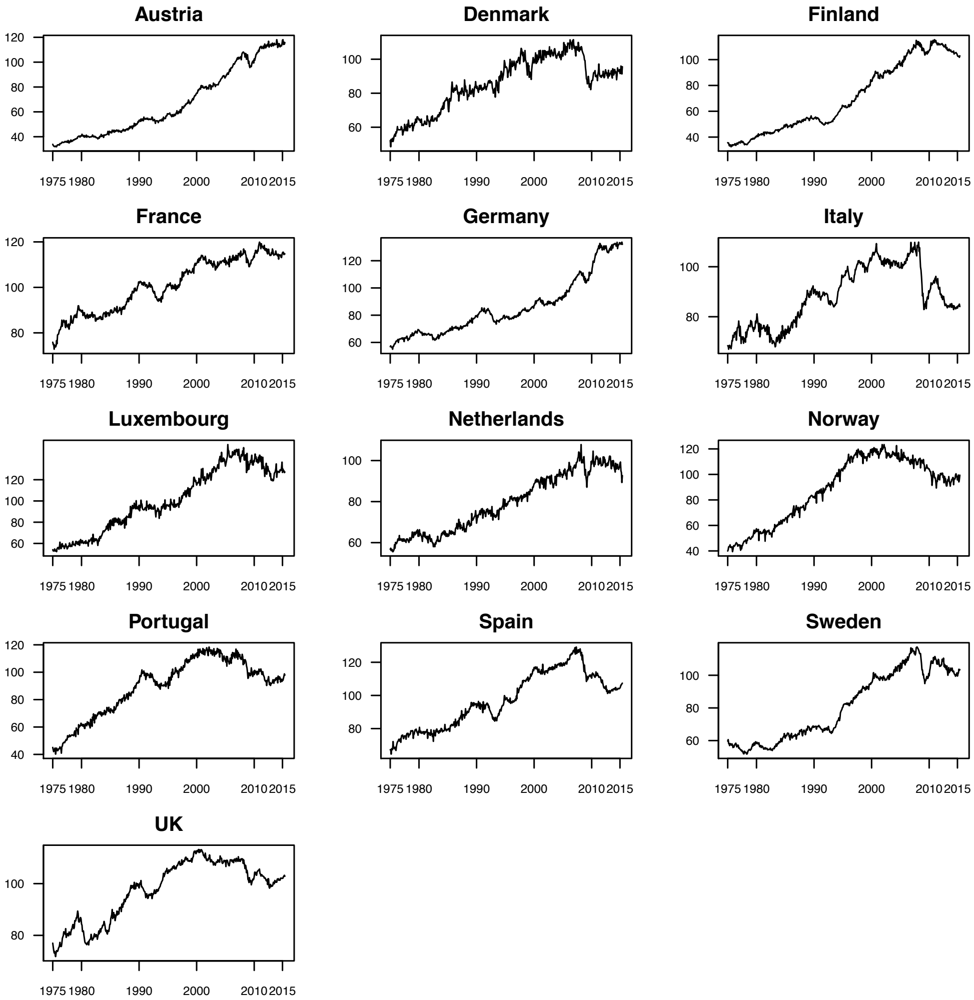
FIGURE 3.1: Industrial Production Index. January, 1975 to June, 2015.



(from May, 2010, to June, 2014), we obtain 424 one- to twelve-step-ahead forecasts. The AR model for each factor is selected in each window, employing BIC. The prediction intervals will have 95% nominal coverage rates. We also performed these estimations employing longer windows of time $T = 100$ and $T = 200$, particularly to show how coverage rates C_m are linked to T in this data-set.

Some additional features outside the scope of the simulations of the previous sections help improve forecasts (equally for *none*, *BC*, and *RF*) in this application. Outliers are intervened beforehand using the statistical software

FIGURE 3.2: Industrial Production Index with intervention of outliers performed using TRAMO. January, 1975 to June, 2015.



TRAMO, through its Matlab interface. Figure 3.2 shows the series after intervening outliers. Furthermore, we obtained an important improvement of using three rather than two common factor to reduce the dimension of this data-set. Last, while in the simulations we know that the specific factors are white noise, in this practical application these are modeled as AR when necessary.

The analysis is performed for the logarithm of the series, but this transformation does not affect the conclusions.

To compare the results of using *none*, *BC*, and *RF* as small sample bias correction methods in the AR models for the common factors, we present actual coverage rates C_m, mean interval lengths L_m, and Mean Absolute Error (MAE). The MAE is calculated in the following way,

$$\mathrm{MAE}^j = \frac{1}{W}\sum_{w=1}^{W}\left(\frac{1}{13}\sum_{i=1}^{13}|(y_{i,z} - \hat{y}^j_{i,z})|\right), \tag{3.12}$$

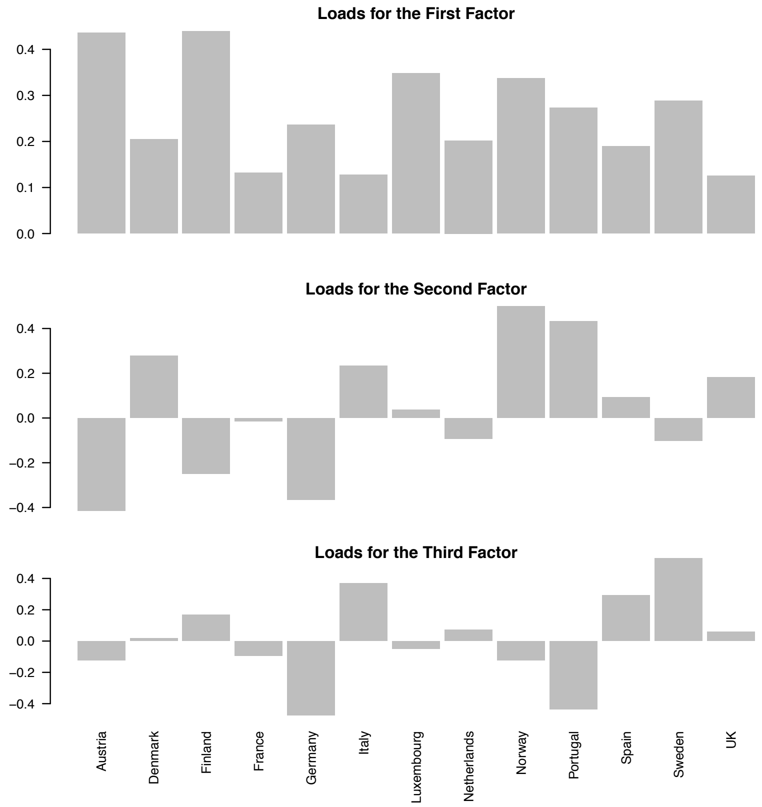
where W is the number of months in the out of sample period (the total number of rolling windows), $i = 1, ..., n$ the series included (in this case we have $n = 13$), and $j = \{none, BC, RF\}$. It is calculated for each forecasting horizon h.

3.5.1 Description of Loads

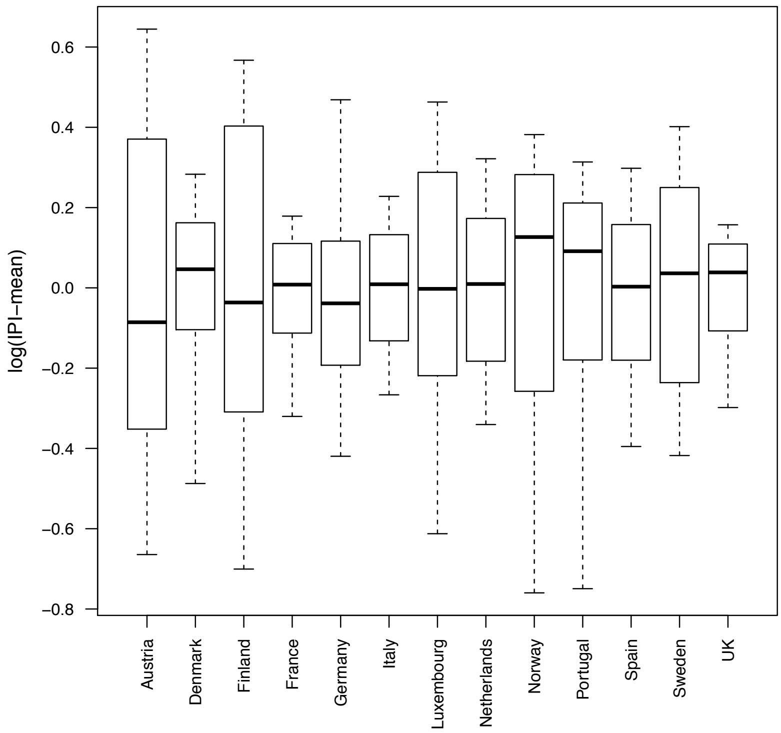
A feature of interest in the empirical estimation are the factors' loads. Because we are working with rolling windows (of diverse length T), loads are estimated together with the unobserved common factors in each window and may change from on window to the next. For this reason, in Figure 3.3a we present the loads we would obtain for the whole data-set instead of any particular window of time; we do this to get an approximate representation of the matrix of weights. We also include box plots of the logarithm of centred IPI for the countries in this study, to identify similarities and differences in the distributions by country.

Oftentimes it is possible to visually find associations between loads and patterns or groupings in the data. The estimated weights for the first factor are highly associated to the variance of IPI in each country (see Table 3.10). The weights for the second factor distinguish two groups of countries: Denmark, Italy, Norway, Portugal, and the United Kingdom on the one hand, and Austria, Germany and, to a lesser degree, Finland, on the other hand.

Last, the weights for the third factor (which, of course, contributes less to the total variability of the data than the other two) separate Germany and Portugal from Italy, Spain and Sweden.



(A) Loads corresponding to three unobserved common factors.



(B) Boxplots of the logarithm of the centred IPI, by country.

FIGURE 3.3: Loads and boxplots for the IPI with outliers intervened ($T = 485$).

TABLE 3.10: IPI complete data-set January, 1975, to June, 2015 ($T = 485$). Loads of the first common factor and variances for 13 European countries.

	Loads first factor	Variances IPI
Austria	0.44	0.16
Denmark	0.20	0.04
Finland	0.44	0.16
France	0.13	0.01
Germany	0.24	0.05
Italy	0.13	0.02
Luxembourg	0.35	0.10
Netherlands	0.20	0.03
Norway	0.34	0.10
Portugal	0.27	0.07
Spain	0.19	0.03
Sweden	0.29	0.07
UK	0.13	0.01

3.5.2 Results

In Tables 3.11 to 3.13 we present the average of the results for the 13 series and in Tables 3.14 to 3.16 we present detailed results for four countries selected to represent diversity in coverage levels.

There are several findings. Firstly, for $T = 50$, the performance of the prediction intervals is short of the 95% nominal coverage. In this regard, Clements and Kim (2007) explain that a small-sample deterioration of the results of high-order models (they employ an AR(6) for United States industrial production data) in comparison to low-order models (like those employed in the simulations) is to be expected. The coverage, C_m, for this empirical example, is highly responsive to the size of the historical data (T) considered in the rolling windows: while for $T = 50$ we obtain C_m deteriorates to C_m =56.50% ($h = 12$, *none*), Table 3.13 shows that, for $T = 200$, C_m is closer to the 95% nominal coverage (the worst coverage is C_m =78.50%, for $h = 12$ in *none*). For $h = 1$, comparing Tables 3.11, 3.12, and 3.13 we can see that the mean coverage, C_m, increases from 89.22 (averaging *none*, *BC*, and *RF* for $h = 1$) for $T = 50$, to 92.31 for $T = 100$, and almost reaches the nominal value for $T = 200$ (94.74 on average for $h = 1$). In other words, large sample sizes, though not always available in practice, contribute to more accurate forecasting intervals.

Secondly, in line with the results obtained in simulations, for one-step-ahead forecasts the improvements of *BC* and *RF* appear small; however, as the forecasting horizon increases, C_m and MAE reveal an evident advantage of employing the corrections, especially *RF*. For *RF*, the advantage in comparison to *none* reaches up to 10.38 percentage points (see Table 3.11, $h = 12$). Last, interval lengths L_m tend to be greater for *BC* and *RF* than for *none*.

TABLE 3.11: IPI forecasting results for 13 European countries. The average of the series is obtained for Interval Coverage C_m (in %), Mean Absolute Error MAE, and Interval's Length L_m. Standard Errors are provided between parenthesis. $T = 50$.

Horizon	Correction	C_m (se)	MAE (se)	L_m (se)
h=1	none	88.82 (1.53)	1.47 (0.02)	6.03 (0.06)
	BC	89.56 (1.48)	1.46 (0.02)	6.07 (0.06)
	RF	89.29 (1.50)	1.46 (0.02)	6.06 (0.06)
h=6	none	71.74 (2.17)	2.78 (0.07)	7.67 (0.11)
	BC	76.50 (2.05)	2.65 (0.06)	8.17 (0.13)
	RF	77.60 (2.02)	2.58 (0.06)	8.21 (0.13)
h=12	none	56.50 (2.37)	4.29 (0.14)	8.46 (0.17)
	BC	64.90 (2.28)	4.12 (0.13)	9.68 (0.21)
	RF	66.52 (2.26)	3.85 (0.10)	9.74 (0.20)

Contrary to C_m, for this application the MAE does not seem to respond as much to sample size. This may seem odd, but it must be considered that the number of windows included in the estimation is smaller for Tables 3.12 and 3.13 because they have longer historical data-sets, than in Table 3.11.

In Tables 3.14 to 3.16 we present the results for Denmark, Finland, Luxembourg and Spain. Finland is the country with the highest coverage C_m in the short term while Luxembourg has the lowest coverage for the first

Table 3.12: IPI forecasting results for 13 European countries. The average of the series is obtained for Interval Coverage C_m (in %), Mean Absolute Error MAE, and Interval's Length L_m. Standard Errors are provided between parenthesis. Rolling windows of size $T = 100$.

Horizon	Correction	C_m (se)	MAE (se)	L_m (se)
h=1	none	92.20 (1.37)	1.47 (0.02)	6.59 (0.05)
	BC	92.41 (1.34)	1.46 (0.02)	6.60 (0.05)
	RF	92.31 (1.36)	1.46 (0.02)	6.61 (0.05)
h=6	none	80.97 (2.01)	2.74 (0.06)	9.19 (0.11)
	BC	83.36 (1.91)	2.67 (0.06)	9.55 (0.12)
	RF	84.04 (1.88)	2.62 (0.06)	9.59 (0.12)
h=12	none	69.91 (2.34)	4.12 (0.11)	10.75 (0.17)
	BC	74.61 (2.22)	3.99 (0.11)	11.84 (0.20)
	RF	76.55 (2.17)	3.82 (0.09)	11.90 (0.19)

Table 3.13: IPI forecasting results for 13 European countries. The average of the series is obtained for Interval Coverage C_m (in %), Mean Absolute Error MAE, and Interval's Length L_m. Standard Errors are provided between parenthesis. Rolling windows of size $T = 200$.

Horizon	Correction	C_m (se)	MAE (se)	L_m (se)
h=1	none	94.73 (1.31)	1.46 (0.03)	7.33 (0.07)
	BC	94.82 (1.30)	1.46 (0.03)	7.36 (0.07)
	RF	94.67 (1.32)	1.46 (0.03)	7.34 (0.06)
h=6	none	87.29 (1.98)	2.84 (0.08)	10.78 (0.11)
	BC	88.45 (1.88)	2.81 (0.07)	10.95 (0.11)
	RF	88.48 (1.88)	2.78 (0.07)	10.96 (0.11)
h=12	none	78.50 (2.46)	4.33 (0.13)	13.11 (0.16)
	BC	80.50 (2.38)	4.22 (0.12)	13.69 (0.17)
	RF	81.43 (2.32)	4.14 (0.12)	13.71 (0.17)

forecasting horizons. The results of Denmark and Spain are closer to the average results.

There may be some concerns regarding the way to best model this data that must be taken into account when interpreting the results. For instance, the models may be incorrectly specified (perhaps more sophisticated approaches should be used to model the factors), or there may be structural breaks in the data (in particular, this could be true for the windows containing the 2008 stock market crash) that are entirely ignored. However, these effects

TABLE 3.14: IPI results for four selected countries. Interval Coverage C_m (in %), Mean Absolute Error MAE, and Interval's Length L_m. Standard Errors are provided between parenthesis. $T = 50$.

Horizon	Correction	C_m (se)				MAE (se)				L_m (se)			
		Denmark	Finland	Luxembourg	Spain	Denmark	Finland	Luxembourg	Spain	Denmark	Finland	Luxembourg	Spain
h=1	none	87.47 (1.61)	90.54 (1.42)	85.58 (1.71)	86.76 (1.65)	2.00 (0.08)	1.04 (0.04)	2.72 (0.11)	1.37 (0.05)	8.24 (0.08)	4.76 (0.06)	10.26 (0.11)	5.67 (0.04)
	BC	88.89 (1.53)	91.49 (1.36)	85.34 (1.72)	87.23 (1.62)	1.99 (0.08)	1.02 (0.04)	2.73 (0.11)	1.36 (0.05)	8.25 (0.08)	4.79 (0.06)	10.36 (0.11)	5.69 (0.04)
	RF	88.18 (1.57)	92.43 (1.29)	84.87 (1.74)	88.18 (1.57)	2.00 (0.08)	1.01 (0.04)	2.70 (0.11)	1.35 (0.05)	8.27 (0.08)	4.81 (0.06)	10.39 (0.11)	5.71 (0.04)
h=6	none	75.65 (2.09)	75.89 (2.08)	74.23 (2.13)	73.29 (2.15)	3.60 (0.14)	2.15 (0.09)	4.24 (0.18)	2.70 (0.13)	10.37 (0.14)	6.31 (0.10)	12.32 (0.16)	7.18 (0.10)
	BC	78.25 (2.01)	79.91 (1.95)	78.72 (1.99)	76.36 (2.07)	3.43 (0.13)	2.02 (0.08)	4.08 (0.18)	2.63 (0.12)	11.13 (0.17)	6.79 (0.11)	13.10 (0.18)	7.63 (0.12)
	RF	77.78 (2.02)	82.27 (1.86)	81.09 (1.91)	76.60 (2.06)	3.44 (0.14)	1.95 (0.08)	3.89 (0.15)	2.53 (0.11)	11.20 (0.17)	6.82 (0.11)	13.31 (0.18)	7.61 (0.11)
h=12	none	60.99 (2.37)	62.88 (2.35)	60.05 (2.38)	60.05 (2.38)	5.16 (0.24)	3.44 (0.18)	6.12 (0.27)	4.36 (0.23)	11.07 (0.17)	7.10 (0.17)	13.17 (0.18)	7.99 (0.16)
	BC	68.32 (2.26)	69.27 (2.25)	70.21 (2.23)	65.48 (2.31)	5.03 (0.23)	3.25 (0.16)	5.91 (0.27)	4.27 (0.23)	12.72 (0.25)	8.22 (0.21)	14.80 (0.23)	9.14 (0.22)
	RF	69.98 (2.23)	71.87 (2.19)	73.29 (2.15)	67.38 (2.28)	4.84 (0.21)	3.03 (0.13)	5.28 (0.19)	3.84 (0.18)	12.88 (0.25)	8.33 (0.18)	15.18 (0.24)	9.08 (0.19)

TABLE 3.15: IPI forecasting results for four selected countries. Interval Coverage C_m (in %), Mean Absolute Error MAE, and Interval's Length L_m. Standard Errors are provided between parenthesis. Rolling windows of size $T = 100$.

Horizon	Correction	C_m (se)				MAE (se)				L_m (se)			
		Denmark	Finland	Luxembourg	Spain	Denmark	Finland	Luxembourg	Spain	Denmark	Finland	Luxembourg	Spain
h=1	none	90.35 (1.53)	95.17 (1.11)	86.06 (1.80)	89.54 (1.59)	2.07 (0.09)	1.03 (0.04)	3.00 (0.12)	1.33 (0.06)	8.91 (0.06)	5.32 (0.07)	11.21 (0.08)	6.12 (0.04)
	BC	90.08 (1.55)	96.51 (0.95)	84.45 (1.88)	90.62 (1.51)	2.07 (0.09)	1.00 (0.04)	3.00 (0.12)	1.32 (0.06)	8.98 (0.06)	5.40 (0.07)	11.21 (0.08)	6.15 (0.04)
	RF	88.74 (1.64)	95.98 (1.02)	86.86 (1.75)	90.35 (1.53)	2.06 (0.09)	1.01 (0.04)	3.00 (0.12)	1.31 (0.06)	8.92 (0.06)	5.37 (0.07)	11.27 (0.08)	6.14 (0.04)
h=6	none	80.43 (2.06)	83.65 (1.92)	77.75 (2.16)	76.68 (2.19)	3.64 (0.15)	2.33 (0.10)	4.59 (0.18)	3.02 (0.13)	12.16 (0.10)	8.18 (0.13)	14.48 (0.15)	8.56 (0.09)
	BC	83.38 (1.93)	86.33 (1.78)	81.77 (2.00)	78.82 (2.12)	3.54 (0.15)	2.20 (0.09)	4.56 (0.18)	2.89 (0.13)	12.68 (0.11)	8.57 (0.13)	15.05 (0.16)	8.85 (0.10)
	RF	82.84 (1.95)	87.94 (1.69)	81.23 (2.02)	79.62 (2.09)	3.53 (0.15)	2.12 (0.08)	4.47 (0.18)	2.79 (0.12)	12.72 (0.11)	8.59 (0.13)	15.27 (0.17)	8.87 (0.09)
h=12	none	75.60 (2.23)	67.02 (2.44)	67.29 (2.43)	56.03 (2.57)	4.99 (0.22)	4.01 (0.16)	6.47 (0.24)	5.05 (0.20)	13.79 (0.16)	9.94 (0.19)	16.36 (0.19)	10.18 (0.16)
	BC	79.89 (2.08)	73.99 (2.27)	73.19 (2.30)	64.34 (2.48)	4.76 (0.21)	3.77 (0.15)	6.40 (0.23)	4.81 (0.20)	15.08 (0.19)	11.10 (0.21)	17.92 (0.22)	11.12 (0.18)
	RF	80.70 (2.05)	77.48 (2.17)	78.28 (2.14)	67.02 (2.44)	4.57 (0.21)	3.56 (0.14)	6.12 (0.21)	4.53 (0.18)	15.17 (0.19)	11.11 (0.20)	18.33 (0.24)	11.16 (0.17)

TABLE 3.16: IPI forecasting results for four selected countries. Interval Coverage C_m (in %), Mean Absolute Error MAE, and Interval's Length L_m. Standard Errors are provided between parenthesis. Rolling windows of size $T = 200$.

Horizon	Correction	C_m (se)				MAE (se)				L_m (se)			
		Denmark	Finland	Luxembourg	Spain	Denmark	Finland	Luxembourg	Spain	Denmark	Finland	Luxembourg	Spain
h=1	none	91.58 (1.68)	97.44 (0.96)	90.11 (1.81)	94.14 (1.42)	2.08 (0.10)	1.11 (0.06)	3.00 (0.16)	1.25 (0.06)	9.83 (0.09)	6.57 (0.08)	12.40 (0.09)	6.96 (0.04)
	BC	91.94 (1.65)	97.44 (0.96)	89.01 (1.90)	95.24 (1.29)	2.08 (0.10)	1.10 (0.05)	2.97 (0.16)	1.23 (0.06)	9.85 (0.08)	6.59 (0.08)	12.51 (0.09)	6.96 (0.05)
	RF	91.58 (1.68)	97.44 (0.96)	89.74 (1.84)	95.24 (1.29)	2.08 (0.10)	1.10 (0.05)	2.99 (0.16)	1.23 (0.06)	9.83 (0.09)	6.57 (0.08)	12.44 (0.09)	7.01 (0.05)
h=6	none	88.64 (1.92)	89.01 (1.90)	79.85 (2.43)	79.12 (2.46)	3.38 (0.20)	2.82 (0.13)	5.12 (0.26)	3.02 (0.16)	13.61 (0.08)	10.86 (0.17)	17.29 (0.16)	9.99 (0.06)
	BC	86.81 (2.05)	92.67 (1.58)	83.15 (2.27)	79.49 (2.45)	3.40 (0.20)	2.66 (0.12)	4.92 (0.25)	2.95 (0.15)	13.71 (0.07)	11.14 (0.16)	17.51 (0.16)	10.14 (0.06)
	RF	87.18 (2.03)	93.77 (1.47)	83.15 (2.27)	82.42 (2.31)	3.39 (0.20)	2.57 (0.11)	4.89 (0.25)	2.85 (0.15)	13.74 (0.08)	11.19 (0.17)	17.47 (0.16)	10.10 (0.06)
h=12	none	85.71 (2.12)	71.06 (2.75)	73.26 (2.68)	68.86 (2.81)	4.37 (0.30)	4.83 (0.25)	7.41 (0.38)	5.04 (0.25)	15.09 (0.07)	14.05 (0.24)	20.83 (0.23)	12.13 (0.10)
	BC	86.08 (2.10)	77.66 (2.53)	79.49 (2.45)	71.06 (2.75)	4.41 (0.30)	4.55 (0.22)	6.86 (0.36)	4.84 (0.24)	15.53 (0.08)	14.97 (0.27)	21.65 (0.24)	12.65 (0.10)
	RF	85.71 (2.12)	77.29 (2.54)	80.22 (2.42)	71.79 (2.73)	4.36 (0.30)	4.42 (0.20)	6.63 (0.36)	4.69 (0.24)	15.62 (0.08)	14.98 (0.26)	21.68 (0.23)	12.65 (0.10)

are out of the scope of this work and the empirical illustration still serves the purpose of demonstrating how bias-correcting the models for the common factors improves forecasting outcomes.

3.6 Concluding Remarks

Following Clements and Kim (2007) we have studied the behavior, in small samples, of three estimators for the AR parameters in a context of highly persistent models. Taking the applications of the methodology one step further, we have employed it in AR models for common factors, when we believe there are underlying unobserved factors driving the behavior of several time series. In all the cases we use the same bootstrap procedure to obtain prediction intervals (Alonso et al., 2008), so the only divergence originates in the estimation of the aforementioned AR parameters.

To evaluate this methodology, we carried out several Monte Carlo simulations, with alternative settings. These consisted of alternative sample sizes (in the time dimension) $T = 50, 100, 200$, different models for the behavior of the common factors (AR(1) and AR(2) factors), and various assumptions in regard to the information and tools available to the researcher, such as previous knowledge (or not) of the number of common factors to obtain and their AR order and employing AICc or BIC criteria to select p, as well as the possibility of having non-Gaussian residuals.

Our most important finding is that in all the settings considered, the two techniques BC and RF succeed at obtaining improved coverage rates in comparison with the situation when no correction is performed. Furthermore, RF tends to be the most advantageous.

Another outcome of the simulations is that, as expected, the smaller the sample (T), the greater the improvement due to bias correction. Therefore, it is more effective to use correction techniques when the sample size is small (which we have represented with $T = 50$) than for larger samples (in particular, we have worked with $T = 200$).

Additionally, the edge of the techniques employed over *none* augments for longer forecasting horizons, another result in line with Clements and Kim (2007).

Lastly, though the empirical results turn out to be rather modest measured by coverage rates, still reveal large differences in performance of the corrected methods vs. *none*. In agreement with the simulations' outcomes, the improvements are more noticeable as the forecasting horizon increases.

Possible extensions include exploring the bias when the common factors follow alternative specifications. For instance, MA terms could be included to the AR models hereby studied. Also, seasonality may be modelled if needed. Another option would be to include VAR specifications for the common factors instead of AR.

Chapter 4

Conclusions and Possible Extensions

4.1 Conclusions

In this dissertation we explore some of the questions that arise with the use of factor models. In particular, we take a different approach to the questions of the number of common factors to use and the model they follow. This approach consists of combining the forecasts of different models for the common factors. These models vary depending on the number of common factors they include and their ARIMA specification. We apply this approach to two data-sets for electricity prices and obtained improved results with the median-based combination, the mean-based combination, and the mean BIC-based combination in comparison to the benchmark (best model according to the BIC). Additionally, we find that using longer historic data-sets especially improve results for the longer forecasting horizons we consider (one to two months). Also, we obtain that the forecasting error is reduced when we include MA terms for modeling the unobserved factors rather than using only AR alternatives.

Another issue we address concerns the bias in the AR coefficient of the common factors when these are close to the unit root. Using Monte Carlo simulations, we find that the bias is stronger for samples that have a small time dimension (low values of T). The techniques we use to correct this bias are the Bootstrap Bias-Correction of Clements and Kim (2007) and Roy-Fuller's estimator by Roy and Fuller (2001). These bias corrections obtain more accurate forecasts than if the bias is not corrected, and this conclusion is maintained in alternative settings such as varying data lengths or having non-Gaussian residuals. Furthermore, we find that RF tends to perform better than BC.

4.2 Possible Extensions

Selection of the Data-Set: The availability of large amounts of data may tempt us to include a very large number of variables N. However, not all the time series will be equally informative of the underlying factors and "extracting information from a large data-set (...) can be suboptimal because of oversampling and error correlation" (Caggiano et al., 2011). One possibility is to verify the importance of the common component on each time series and proceed to exclude the least critical ones. Another option in the same line consists of selecting a subset of N to obtain the commonality. Boivin and Ng (2006) obtain improvements in efficiency by doing so. They introduce two rules for reducing the number of series included in the N used to estimate the factors. One drops those series which idiosyncratic error is most correlated to some other series and the other one reduces the data even further by dropping also the series which error is second most correlated to others. Caggiano et al. (2011) also obtain forecasts improvements by pre-selecting variables according to Boivin and Ng (2006) in an empirical investigation with European data.

Outliers: Weron (2014) points out a disagreement in the literature regarding the inclusion of price spikes in the estimation of statistical models. It is not clear whether we should intervene outliers directly in the original data, or if we should handle outliers once the factors are estimated and when we intend to estimate their models. Lee and Carter (1992) employ a factor model for mortality rates in the United States and address the issue of how to treat the 1918 epidemic (an extreme behavior of the series). Their discussion starts once the common factor has been estimated, not before. They decide to intervene the outlier, but recognize that this may not be desirable, because it is necessary to recognize this uncertainty in the forecasting intervals, if it is believed that such an event may re-occur.

A further problem caused by the presence of outliers in the data is the reliability of the sample variance-covariance matrix, and thus of the results of the SVD. There are alternative approaches to robust estimation. An interesting approach is that of Xu et al. (2012). These authors work with convex optimization to obtain the exact low-dimensional subspace of the common factors. Their technique may be employed even when all the time series are corrupted at a point t in time, which can be beneficial in many real-life settings, and not only produces robust principal components estimates, but it also identifies the outliers. Peña and Prieto (2007) also work with outlier detection and robust estimation of the covariance of high dimensional data basing their procedure on projections.

Alternative Models for the Common Factors: including alternative specifications could reduce the forecasting error even more. As extensions to Chapter 2, alternative VARIMA models for the common factors could enter the combinations as well (like the SeaDFA). On the other hand, other techniques for forecast combination could be tested. One example of this is to employ Bayesian weights, which could be calculated based on prior distributions that incorporate knowledge of the particular data to model. Also, though we work with weights that are fixed for each window of time, weights tailored to the forecasting horizon may be successful.

With regard to the bias correction techniques of Chapter 3, the models for the common factors could be extended to ARMA instead of only AR models. If the common factors were correlated, the bias of VAR models could be studied as well. A further extension would be to compare the small sample bias when Principal Components are used to estimate the factors, vs. other techniques, such as the generalized Dynamic Factor Model of Forni et al. (2005).

Appendix A

Appendix to Chapter 2

A.1 Details of ANOVA for a Comparison of the Alternatives for Modeling. Results for the Iberian Electricity Market.

In this section, we describe in detail the results for the ANOVA performed for the forecasting horizons $h = 1,\ 7,\ 30,\ 60$, for the data-set of prices in the Iberian electricity market. In order to be robust against departures from the Gaussianity assumption, we employ bootstrap to calculate the ANOVA's p-values, as well as the confidence intervals for the means of the DOE's factors.

Table A.1 presents a summary of the results described in the following subsections. For each DOE's factor it indicates the best level. For instance, for $h = 1$, *Historical Length*= 308 days outperforms the alternative *Historical Length*= 548 days. We work with ln(MAE) as dependent variable in order to eliminate heteroskedasticity.

TABLE A.1: Summary of results of the Analysis of Variance for ln(MAE) for all forecasting horizons. Iberian market.

	$h = 1$	$h = 7$	$h = 30$	$h = 60$
Logarithm	ns	ns	ns	ns
Hist Length (days)	308	308	548	548
MA	Yes	Yes	Yes	Yes
Combinations	{6}	{2, 3, 6}	{2, 3, 6}	{2, 3, 6}

Notes: Significance level of at least α =0.01. ns: not significant.

A.1.1 Minimizing Forecasting Error for One-Day-Ahead Forecasts

To assess the results for one-day-ahead forecasts ($h = 1$), see Figure A.1 and Table A.2. In Figure A.1, the horizontal axis presents the alternative values of the DOE's factors (*Logarithm*, *Historical Length*, *Moving Average*, *Forecast Combinations*) and the vertical axis shows the logarithm of MAE corresponding to the means and 95% confidence intervals.

The effect of DOE factor *Logarithm* is not significant. On the contrary, when considering *Historical Length*, we can see a significant difference: using the short historic window gives significantly better forecasts in terms of forecasting accuracy - a smaller MAE - than using the long historic window (548 days). Regarding the *Moving Average* component, significantly better results are obtained when incorporating an MA term in the model of the unobserved common factors (the alternative being modeling as ARI). Last, *Forecast Combination* 6 (mean BIC-based combination) outperforms most of the others.

Table A.2: Analysis of Variance for ln(MAE). Main effects. Forecast horizon $h = 1$, Iberian market.

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Logarithm	1	0.02	0.02	0.67	0.4331
Hist Length	1	1.56	1.56	59.45	0.0020
MA	1	111.24	111.24	4240.25	0.0020
Combinations	5	0.35	0.07	2.70	0.0140
Day	1766	22166.97	12.55	478.46	0.0020
Residuals	83041	2178.51	0.03		

Notes: DF stands for Degrees of Freedom. All F-ratios are based on the residual mean square error. P-values are estimated by bootstrap.

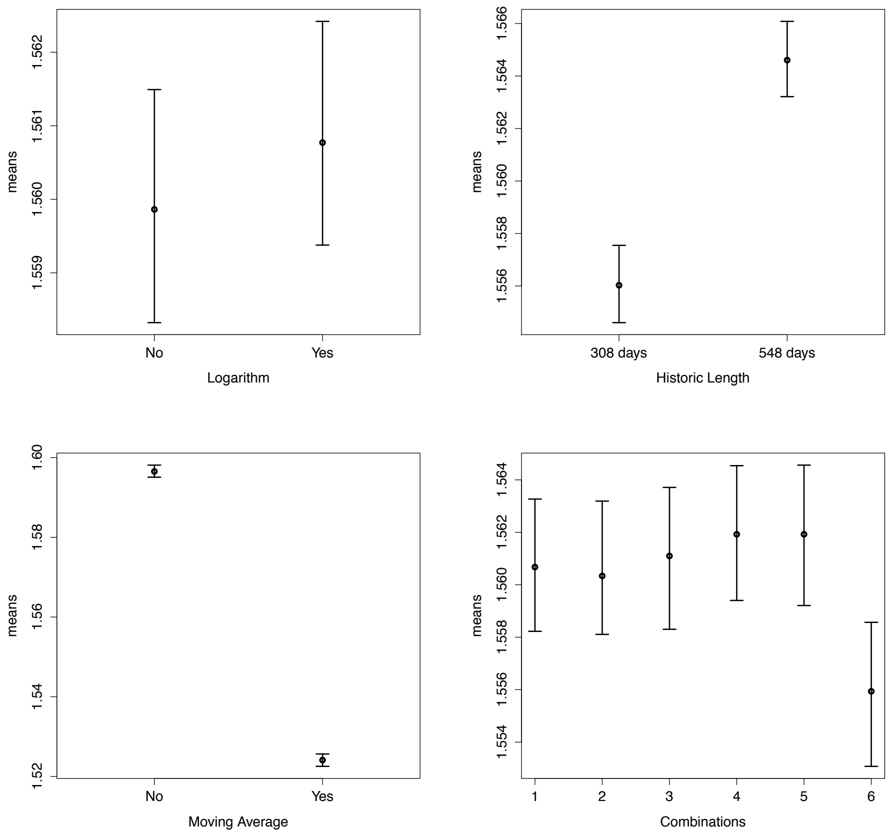


Figure A.1: Bootstrap confidence intervals for the mean ln(MAE) of the factors *Logarithm*, *Historical Length*, *Moving Average*, and *Forecast Combinations*. *Forecast Combinations* include: (1) benchmark BIC-selected model, (2) median-based combination, (3) mean-based combination, (4) BIC-based combination, (5) BIC-50% combination, (6) mean BIC-based combination. Forecast horizon $h = 1$, Iberian market.

TABLE A.3: Analysis of Variance for ln(MAE). Main effects. Forecast horizon $h = 7$, Iberian market.

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Logarithm	1	0.07	0.07	2.66	0.1218
Hist Length	1	0.60	0.60	23.07	0.0020
MA	1	7.71	7.71	295.44	0.0020
Combinations	5	8.37	1.67	64.18	0.0020
Day	1766	25363.03	14.36	550.59	0.0020
Residuals	83041	2166.09	0.03		

Notes: DF stands for Degrees of Freedom. All F-ratios are based on the residual mean square error. P-values are estimated by bootstrap.

A.1.2 Minimizing Forecasting Error for Seven-Day-Ahead Forecasts

Table A.3 and Figure A.2 present the results for seven-day-ahead forecasts ($h = 7$). There are three *Forecast Combinations* which outperform the benchmark: 2, 3 and 6. Furthermore, there is a significant difference between the two values for *Historical Length*: employing historic data-sets of 308 days provides with significantly better forecasts than 548 days. Significantly better results are obtained when incorporating a *Moving Average* component in the unobserved common factors' models. Last, the effect of *Logarithm* continues to be not significant.

A.1.3 Minimizing Forecasting Error for One-Month-Ahead Forecasts

Details on the results for thirty-day-ahead forecasts ($h = 30$) can be found in Table A.4 and Figure A.3. For *Forecast Combinations*, we obtain similar results to $h = 7$. Contrary to shorter forecasting horizons, the *Historical Length* of 548 days presents significantly better forecasts than the shorter window, an intuitive result. Again, significantly better results are obtained

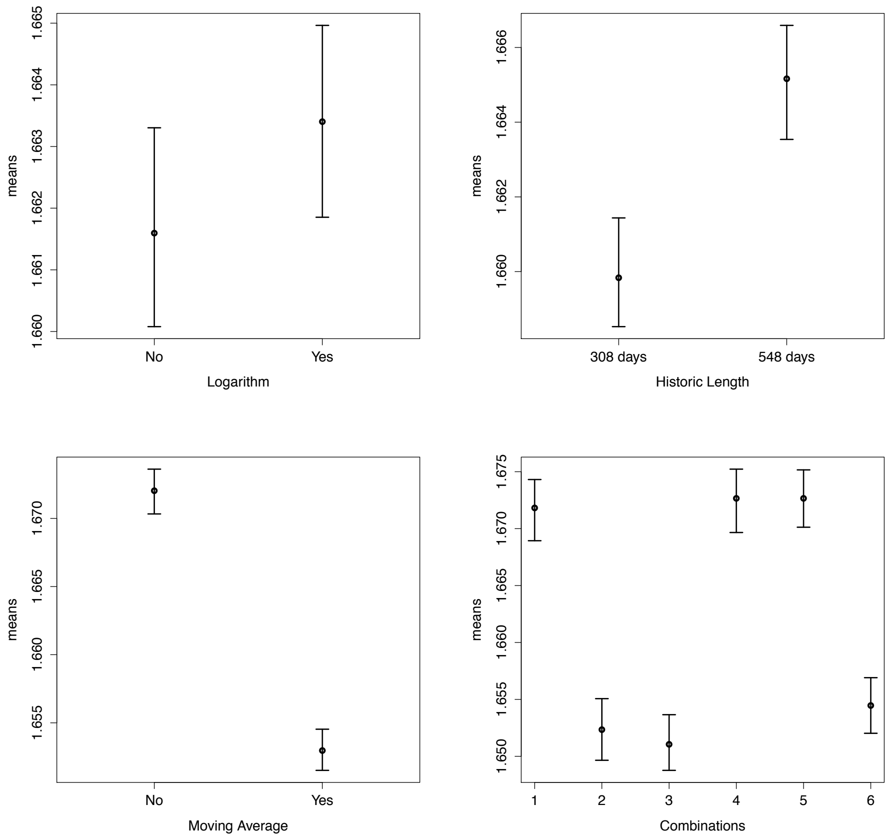


FIGURE A.2: Bootstrap confidence intervals for the mean ln(MAE) of the factors *Logarithm*, *Historical Length*, *Moving Average*, and *Forecast Combinations*. *Forecast Combinations* include: (1) benchmark BIC-selected model, (2) median-based combination, (3) mean-based combination, (4) BIC-based combination, (5) BIC-50% combination, (6) mean BIC-based combination. Forecast horizon $h = 7$, Iberian market.

with a *Moving Average* component in the model for unobserved common factors. The effect of *Logarithm* continues to be not significant.

A.1.4 Minimizing Forecasting Error for Two-Month-Ahead Forecasts

The longest forecasting horizon considered in this assessment is sixty-day-ahead ($h = 60$). The results are presented in Table A.5 and Figure A.4. As

TABLE A.4: Analysis of Variance for ln(MAE). Main effects. Forecast horizon $h = 30$, Iberian market.

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Logarithm	1	0.01	0.01	0.33	0.5669
Hist Length	1	19.90	19.90	458.58	0.0020
MA	1	10.34	10.34	238.25	0.0020
Combinations	5	11.35	2.27	52.31	0.0020
Day	1766	24319.88	13.77	317.31	0.0020
Residuals	83041	3604.00	0.04		

Notes: DF stands for Degrees of Freedom. All F-ratios are based on the residual mean square error. P-values are estimated by bootstrap.

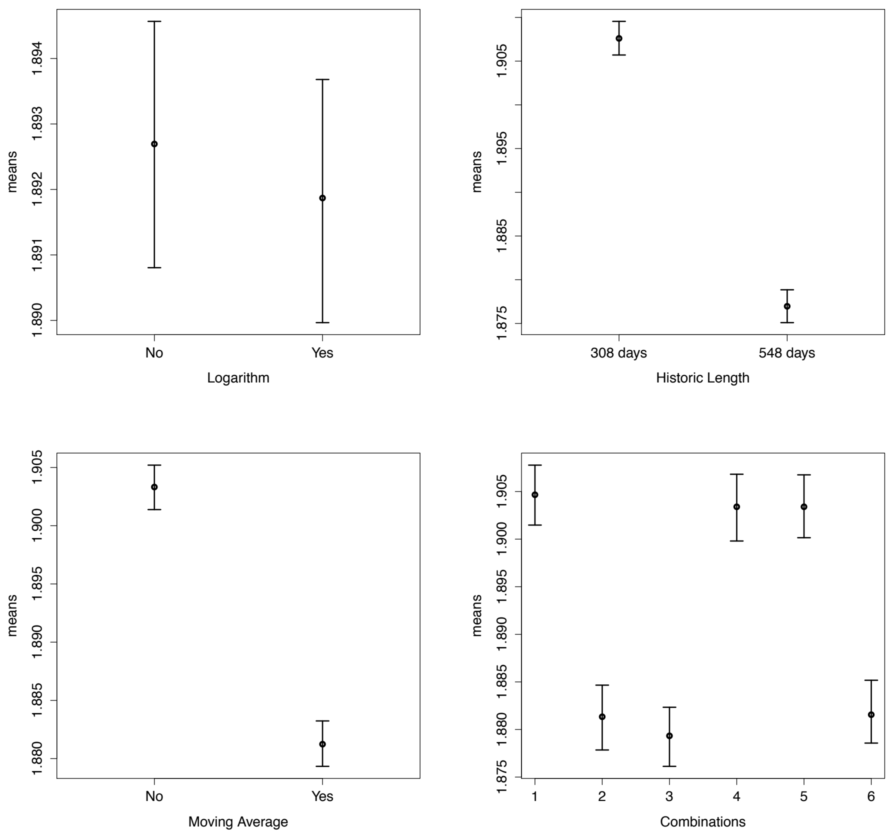


FIGURE A.3: Bootstrap confidence intervals for the mean ln(MAE) of the factors *Logarithm*, *Historical Length*, *Moving Average*, and *Forecast Combinations*. *Forecast Combinations* include: (1) benchmark BIC-selected model, (2) median-based combination, (3) mean-based combination, (4) BIC-based combination, (5) BIC-50% combination, (6) mean BIC-based combination. Forecast horizon $h = 30$, Iberian market.

TABLE A.5: Analysis of Variance for ln(MAE). Main effects. Forecast horizon $h = 60$, Iberian market.

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Logarithm	1	0.01	0.01	0.15	0.6946
Hist Length	1	126.98	126.98	2359.20	0.0020
MA	1	6.73	6.73	125.11	0.0020
Combinations	5	20.82	4.16	77.37	0.0020
Day	1766	25660.02	14.53	269.97	0.0020
Residuals	83041	4469.42	0.05		

Notes: DF stands for Degrees of Freedom. All F-ratios are based on the residual mean square error. P-values are estimated by bootstrap.

for $h = 7$ and $h = 30$, we obtain that the three most successful *Forecast Combinations* are 2, 3 and 6. Employing a *Historical Length* of 548 days provides with significantly better forecasts in terms of forecasting accuracy than using the shorter option. Additionally, significantly better results are obtained when incorporating a *Moving Average* component to model the unobserved common factors. Last, there is no change with respect to the conclusions for *Logarithm*.

A.2 Details of ANOVA for a Comparison of the Alternatives for Modeling. Results for the Italian Electricity Market

In this section, we describe in detail the results for the ANOVA performed for the forecasting horizons $h = 1,\ 7,\ 30,\ 60$, for the data-set of prices in the Italian electricity market. In order to be robust against departures from the Gaussianity assumption, we employ bootstrap to calculate the p-values of the analysis, as well as the confidence intervals for the means of the DOE's factors.

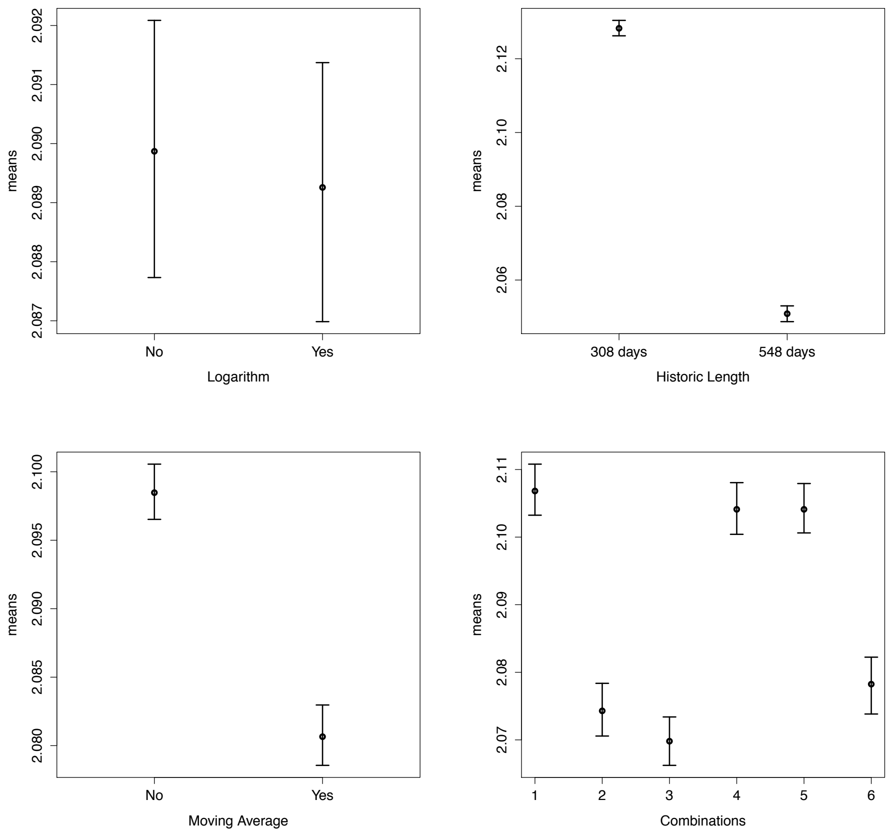


FIGURE A.4: Bootstrap confidence intervals for the mean ln(MAE) of the factors *Logarithm*, *Historical Length*, *Moving Average*, and *Forecast Combinations*. *Forecast Combinations* include: (1) benchmark BIC-selected model, (2) median-based combination, (3) mean-based combination, (4) BIC-based combination, (5) BIC-50% combination, (6) mean BIC-based combination. Forecast horizon $h = 60$, Iberian market.

Table A.6 presents a summary of the results described in the following sections. Notice that we work with ln(MAE) for dependent variable in order to eliminate heteroskedasticity.

Table A.6: Summary of results of the Analysis of Variance for ln(MAE) for all forecasting horizons. Italian market.

	$h = 1$	$h = 7$	$h = 30$	$h = 60$
Logarithm	Yes	Yes	Yes	Yes
Hist Length (days)	308	308	308	548
MA	Yes	Yes	Yes	Yes
Combinations	$\{2:6\}$	$\{2:6\}$	$\{2,\ 3,\ 6\}$	$\{6\}$

Notes: Significance level of at least α =0.01. ns: not significant.

Table A.7: Analysis of Variance for ln(MAE). Main effects. Forecast horizon $h = 1$, Italian market.

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Logarithm	1	49.61	49.61	1092.13	0.0020
Hist Length	1	25.78	25.78	567.39	0.0020
MA	1	220.14	220.14	4846.00	0.0020
Combinations	5	5.66	1.13	24.91	0.0020
Day	1218	12469.53	10.24	225.36	0.0020
Residuals	57285	2602.32	0.05		

Notes: DF stands for Degrees of Freedom. All F-ratios are based on the residual mean square error. P-values are estimated by bootstrap.

A.2.1 Minimizing Forecasting Error for One-Day-Ahead Forecasts

For the shortest forecasting horizon considered ($h = 1$), we obtain that all the factors included in the DOE affect the forecasting error measured by ln(MAE) (see Table A.7). Figure A.5 presents 95% confidence intervals showing the effect of the DOE factors in the mean of ln(MAE). Better results are obtained when the dependent variable is ln(Prices) rather than *Prices* and also when the short *Historical Length* is used. Employing ARIMA models for the common factors instead of ARI models also results in a smaller forecasting error. We find that all the proposed *Forecast Combinations*, specially 6, turn out to provide significantly better results than the benchmark (presented as *Combination* 1 in the plot).

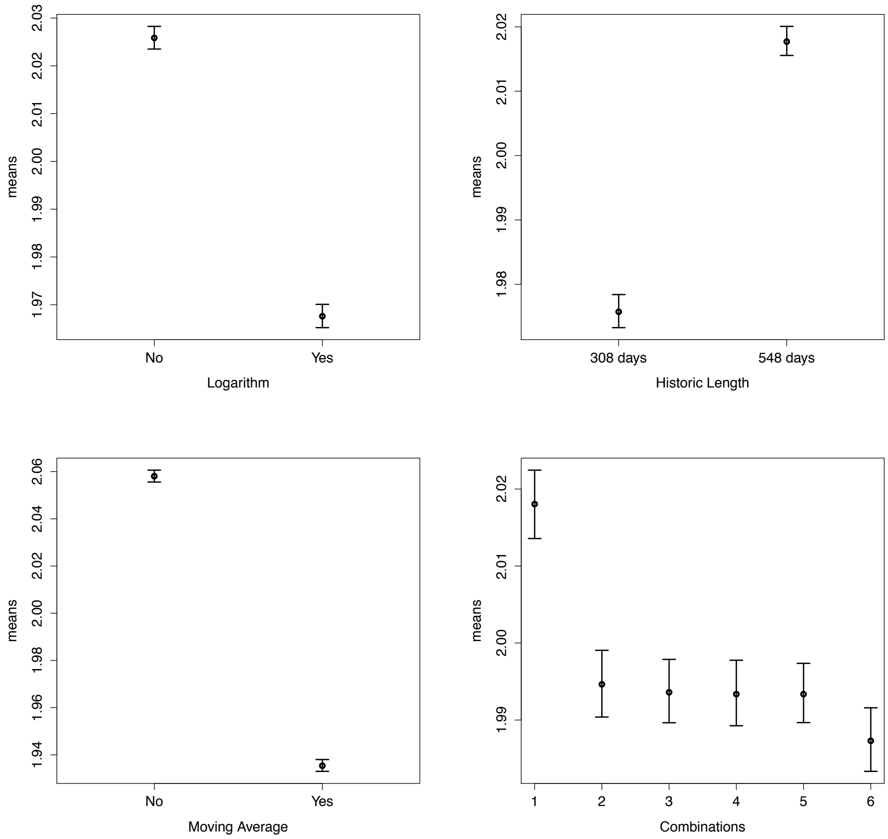


FIGURE A.5: Bootstrap confidence intervals for the mean ln(MAE) of the factors *Logarithm*, *Historical Length*, *Moving Average*, and *Forecast Combinations*. *Forecast Combinations* include: (1) benchmark BIC-selected model, (2) median-based combination, (3) mean-based combination, (4) BIC-based combination, (5) BIC-50% combination, (6) mean BIC-based combination. Forecast horizon $h = 1$, Italian market.

A.2.2 Minimizing Forecasting Error for Seven-Day-Ahead Forecasts

For $h = 7$, all the factors included in the DOE affect the forecasting error measured by ln(MAE) (see Table A.8). The confidence intervals showing the effect of the DOE's factors in the mean of ln(MAE) are presented in Figure A.6. Similar results to $h = 1$ are obtained for *Logarithm*, *Historical Length*, *Moving Average*, and *Forecast Combinations*. Especially *Forecast*

TABLE A.8: Analysis of Variance for ln(MAE). Main effects. Forecast horizon $h = 7$, Italian market.

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Logarithm	1	17.17	17.17	530.14	0.0020
Hist Length	1	10.42	10.42	321.78	0.0020
MA	1	60.84	60.84	1878.39	0.0020
Combinations	5	10.03	2.01	61.94	0.0020
Day	1218	13982.28	11.48	354.44	0.0020
Residuals	57285	1855.34	0.03		

Notes: DF stands for Degrees of Freedom. All F-ratios are based on the residual mean square error. P-values are estimated by bootstrap.

TABLE A.9: Analysis of Variance for ln(MAE). Main effects. Forecast horizon $h = 30$, Italian market.

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Logarithm	1	21.23	21.23	453.13	0.0020
Hist Length	1	6.21	6.21	132.49	0.0020
MA	1	140.73	140.73	3004.28	0.0020
Combinations	5	2.36	0.47	10.09	0.0020
Day	1218	13557.33	11.13	237.62	0.0020
Residuals	57285	2683.40	0.05		

Notes: DF stands for Degrees of Freedom. All F-ratios are based on the residual mean square error. P-values are estimated by bootstrap.

Combinations 2,3 and 6, turn out to provide significantly better results than the benchmark *Forecast Combination* 1.

A.2.3 Minimizing Forecasting Error for One-Month-Ahead Forecasts

For $h = 30$, we obtain similar results to those presented for shorter horizons (see Table A.9). This is different from the Iberian case, in which a longer *Historical Length* results beneficial for this forecasting horizon. Figure A.7 presents the confidence intervals showing the effect of the factors in the mean of ln(MAE). In the case of *Forecast Combinations*, 4 and 5 no longer provide a significant advantage.

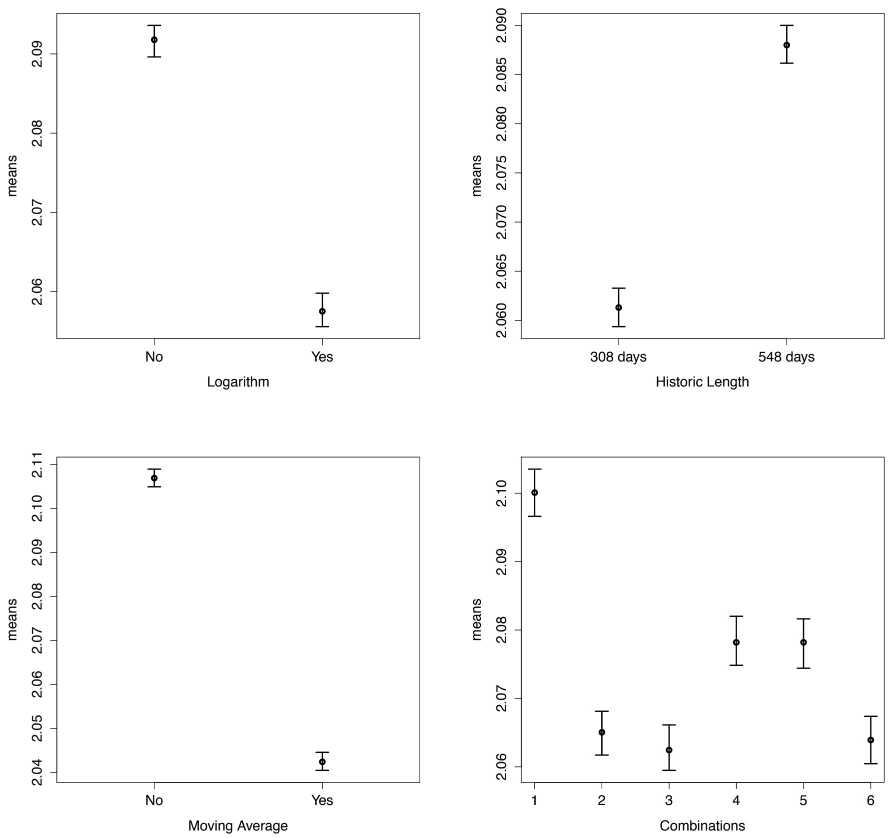


FIGURE A.6: Bootstrap confidence intervals for the mean ln(MAE) of the factors *Logarithm*, *Historical Length*, *Moving Average*, and *Forecast Combinations*. *Forecast Combinations* include: (1) benchmark BIC-selected model, (2) median-based combination, (3) mean-based combination, (4) BIC-based combination, (5) BIC-50% combination, (6) mean BIC-based combination. Forecast horizon $h = 7$, Italian market.

A.2.4 Minimizing Forecasting Error for Two-Month-Ahead Forecasts

For the largest forecasting horizon considered ($h = 60$) we obtain that all the factors included in the DOE are statistically significant (see Table A.10). Figure A.8 presents the confidence intervals showing the effect of the factors in the mean of ln(MAE). Better results are obtained when the dependent variable is ln(Prices) instead of 'Prices' and also when the long *Historical*

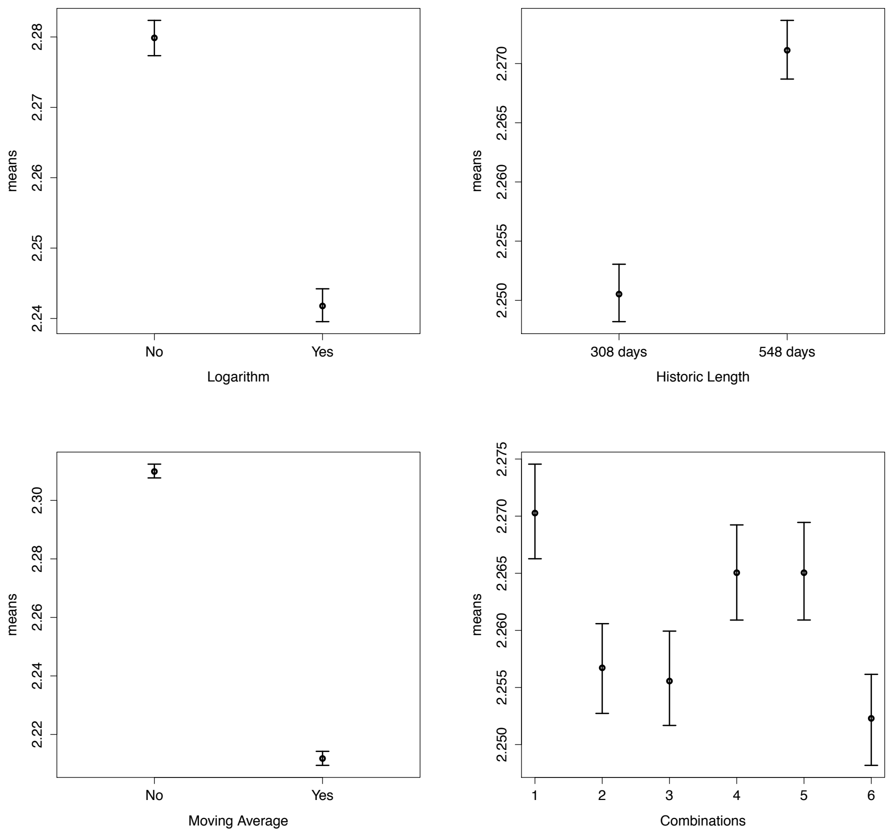


FIGURE A.7: Bootstrap confidence intervals for the mean ln(MAE) of the factors *Logarithm*, *Historical Length*, *Moving Average*, and *Forecast Combinations*. *Forecast Combinations* include: (1) benchmark BIC-selected model, (2) median-based combination, (3) mean-based combination, (4) BIC-based combination, (5) BIC-50% combination, (6) mean BIC-based combination. Forecast horizon $h = 30$, Italian market.

Length is used, contrary to what was found in the previous forecasting horizons but supported in the literature. Setting *Moving Average*= Yes still improves the results. We continue to find that *Forecast Combination* 6 provides significantly better results than the benchmark (*Combination* 1 in the figure), but the benefits of using 2 and 3 are not as strong as with $h < 60$.

Table A.10: Analysis of Variance for ln(MAE). Main effects. Forecast horizon $h = 60$, Italian market.

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Logarithm	1	5.21	5.21	110.18	0.0020
Hist Length	1	1.93	1.93	40.86	0.0020
MA	1	83.70	83.70	1770.80	0.0020
Combinations	5	1.69	0.34	7.17	0.0020
Day	1218	13013.62	10.68	226.05	0.0020
Residuals	57285	2707.58	0.05		

Notes: DF stands for Degrees of Freedom. All F-ratios are based on the residual mean square error. P-values are estimated by bootstrap.

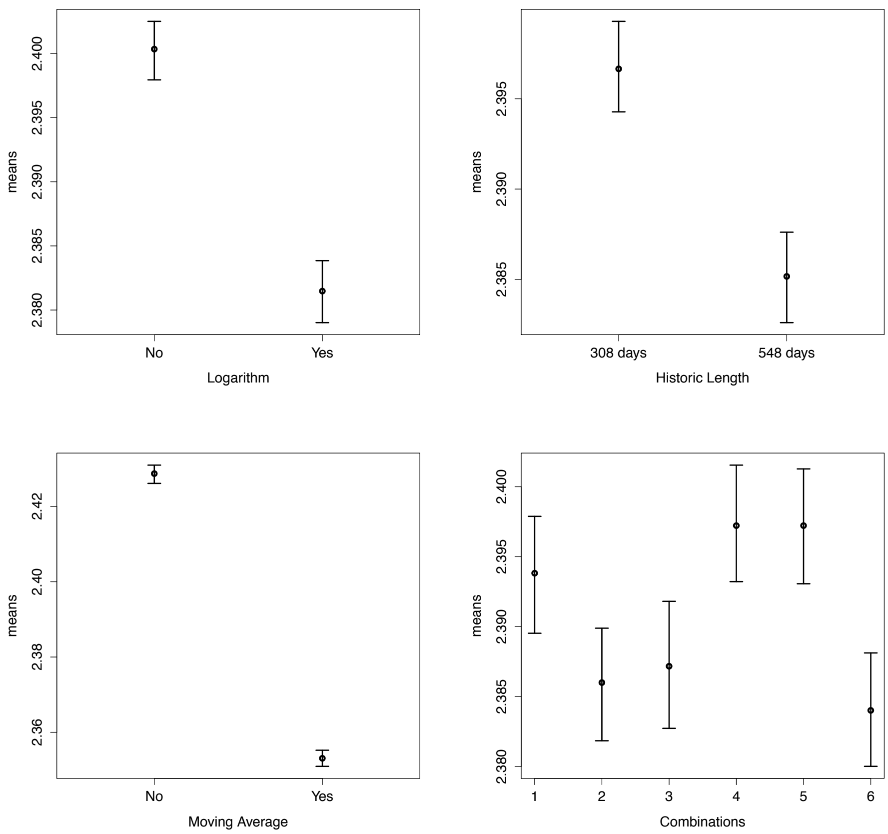


Figure A.8: Bootstrap confidence intervals for the mean ln(MAE) of the factors *Logarithm*, *Historical Length*, *Moving Average*, and *Forecast Combinations*. *Forecast Combinations* include: (1) benchmark BIC-selected model, (2) median-based combination, (3) mean-based combination, (4) BIC-based combination, (5) BIC-50% combination, (6) mean BIC-based combination. Forecast horizon $h = 60$, Italian market.

A.3 ANOVA Bootstrap Procedures

Davison and Hinkley (1997) indicate that when we have doubts about the accuracy of the normal approximation, the empirical distribution can be a fairer approximation of the distribution of the parameter of interest. As an advantage, we do not need to know the distribution of the underlying parameter in order to employ bootstrap procedures.

A.3.1 Bootstrap Procedure to Calculate ANOVA P-values

Usually, the ANOVA relies on the assumption that the residuals u_{ijkld} of expression (2.11) are normally distributed. When the data is normally distributed, the Sum of Squares has a χ^2 distribution and the quotient of Mean Squares follows a distribution Fisher-Snedecor (we call this statistic F-ratio). The lack of normality in u_{ijkld} results in an F-ratio that will most likely not have a Fisher-Snedecor distribution; therefore, the p-value, determined as $Pr(F-\text{ratio} \geq F_{n_1,n_2,\alpha})$[1], is no longer accurate. To avoid complicating notation, we keep using the denomination F-ratio even if this statistic does not have a Fisher-Snedecor distribution.

To be robust to departures from the Gaussianity assumption, we calculate the ANOVA p-values employing bootstrap, according to the following steps. We use the DOE factor *Logarithm* for illustrative purposes, but the procedure is the same for all the factors considered.

1. We estimate model (2.11), obtaining estimates $\widehat{u}_{ijkld}$ and the F-ratio of the ANOVA.

[1] $F_{n_1,n_2,\alpha}$ is such that $\Pr(F > F_{n1,n2,\alpha}) = \alpha$ when F follows the Fisher-Snedecor distribution with n_1 and n_2 degrees of freedom.

2. We would like to test if there is an effect to taking logarithm of prices. In other words, is there a difference in the forecasting error of setting *Logarithm*= No vs. *Logarithm*= Yes? This translates in the null hypothesis $H0: \ \alpha_{\text{No}} = \alpha_{\text{Yes}}$. We estimate (2.11) under null hypothesis $H0$ and obtain estimates $\hat{\mu}^R$, $\hat{\beta}_j^R$, $\hat{\gamma}_k^R$, $\hat{\delta}_l^R$ and $\hat{\epsilon}_d^R$, where ^{R} stands for 'restricted' model.

3. We generate synthetic samples for the ln(MAE); each bootstrap replication needs to satisfy the null hypothesis, though employing a random sample of the unrestricted model's estimated residuals, $\widehat{u^*}_{ijkld}$. The replications are independent and the random samples (with replacement) of $\widehat{u^*}_{ijkld}$ have the same size as the original data-set.

$$ln(\text{MAE}^*_{ijkld}) = \hat{\mu}^R + \hat{\beta}_j^R + \hat{\gamma}_k^R + \hat{\delta}_l^R + \hat{\epsilon}_d^R + u^*_{ijkld}. \tag{A.1}$$

4. For each bootstrap replication we re-estimate model (2.11) to the synthetic data $\ln(\text{MAE}^*_{ijkld})$ and obtain the statistic F_b^*-ratio, where the sub-index $b = 1, ..., B$ represents each of the B bootstrap replicas.

5. We obtain the Monte Carlo p-value as indicated in Davison and Hinkley (1997),
$$\hat{p}^* = \frac{1 + \#\{F_b^* \geq F\}}{B+1}, \tag{A.2}$$
where $\#\{\cdot\}$ indicates the number of times the event in braces occurs. We employ $B = 500$ bootstrap replications.

A.3.2 Bootstrap Procedure to Calculate ANOVA Main Effects Confidence Intervals

Also to allow departures from Gaussianity, the confidence intervals for the main effects need not be estimated with the parametric formula employed for normal data. Instead, we recur to a Monte Carlo simulation of the bootstrap. The procedure is the following.

1. We estimate model (2.11), obtaining estimates for the residuals, $\widehat{u}_{ijkld}$.

2. For each bootstrap replication, we obtain a random sample of $\widehat{u}_{ijkld}$ and generate a data-set

$$ln(\mathrm{MAE}^{*}_{ijkld}) = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\delta}_l + \hat{\epsilon}_d + u^{*}_{ijkld}. \tag{A.3}$$

3. We use each simulated sample to estimate the ANOVA and obtain the estimates for each value of the factors *Logarithm*, *Historical Length*, *MA*, and *Forecast Combinations*.

4. We work with $B = 500$ replications to obtain a bootstrap distribution for the mean ln(MAE) at each level of the factors of the DOE. We use the bootstrap percentile interval (Davison and Hinkley, 1997, chapter 5) to calculate 95% confidence intervals for the mean levels of each factor, by employing the 2.5 and 97.5 percentiles of the estimates of all the bootstrap replications.

A.4 Combinations and Benchmark Comparison Using the RMSE

In this section, we present the results analogous to Tables 2.3 and 2.4, but employing a different accuracy metric, the Root Mean Squared Error. We obtain similar results to those for the MAE: for both data-sets, *Forecast Combinations* median-based, mean-based, and specially mean-BIC-based, report noticeable improvements with respect to the benchmark BIC-selected model.

TABLE A.11: Weekly, monthly, and bi-monthly RMSE for the Iberian Market.

	BIC-selected model	Median-based Combination	Mean-based Combination	BIC-based Combination	BIC 50% Combination	Mean-BIC-based Combination
Weekly						
RMSE	7.1468	7.0810	7.1103	7.1389	7.1389	7.0393
Monthly						
RMSE	8.1668	7.9612	7.9721	8.1533	8.1533	7.9052
Bi-Monthly						
RMSE	9.1352	8.8689	8.8705	9.1189	9.1189	8.7971

TABLE A.12: Weekly, monthly, and bi-monthly RMSE for the Italian Market.

	BIC-selected model	Median-based Combination	Mean-based Combination	BIC-based Combination	BIC 50% Combination	Mean-BIC-based Combination
Weekly						
RMSE	10.6280	10.5466	10.5937	10.6194	10.6194	10.4675
Monthly						
RMSE	11.9479	11.8175	11.8685	11.9353	11.9353	11.7096
Bi-Monthly						
RMSE	12.9861	12.6656	12.6981	12.9637	12.9637	12.5709

Appendix B

Appendix to Chapter 3

B.1 Details About the Results for the Common Factors

Because we work with simulated data, we generate and know the values of the underlying factors, contrary to the situation of working with empirical data. In this section we take advantage of this setting to understand better the circumstances surrounding the estimation of the models for the common factors $\boldsymbol{F_1}$ and $\boldsymbol{F_2}$. We provide details for the case of two factors that follow AR(1) processes, both r and p are assumed to be known.

In our simulation, it is straight forward to check the bias for factors that are AR(1). This is done in Table B.1. The bias of BC and RF is much smaller than the one for *none*. Notice however that the estimation rendered by *none* gets closer to the true value of the coefficients ϕ_{F1}, ϕ_{F2} as the sample gets larger. Thus, the emphasis in that the correction techniques employed in this work are particularly beneficial for small samples.

Table B.1: Bias for common factors following AR(1) processes with normal errors. Between parenthesis, the variance of the AR estimated coefficients. Monte Carlo simulations of model with coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. 10 000 MC replications.

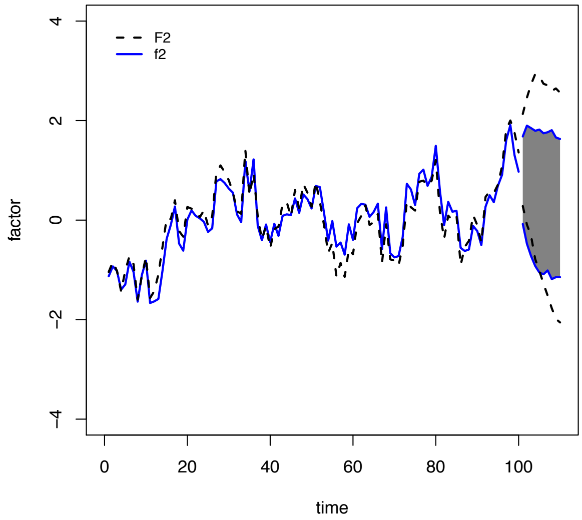
Factor	Correction	T=50	T=100	T=200
F1	none	0.086 (0.006)	0.045 (0.002)	0.023 (0.001)
	BC	0.028 (0.005)	0.011 (0.002)	0.003 (0.001)
	RF	0.018 (0.005)	0.005 (0.002)	-0.001 (0.001)
F2	none	0.157 (0.013)	0.078 (0.005)	0.040 (0.002)
	BC	0.092 (0.015)	0.042 (0.005)	0.021 (0.002)
	RF	0.078 (0.017)	0.037 (0.005)	0.021 (0.002)

Tables B.2, B.3, and B.4 present the results of the same simulation, this time for the factors (instead of the selected series) for each indicated sample size. We can see that the interval coverage C_m is oftentimes far from the theoretical 95%, especially for $\boldsymbol{F}_2$. Notwithstanding, the performance in terms of coverage of the series (studied in Subsection 3.4.1) is much better. Take for instance the estimation of $\boldsymbol{F}_2$ of a sample of size $T = 100$. In Figure B.1 we present $\boldsymbol{F}_2$ and its estimate $\boldsymbol{f}_2$ for a random sample, of the model in which factors follow AR(1) processes. Notice that for the last "observed" value (for time $t = 100$), $\boldsymbol{F}_2$ and the factor estimation $\boldsymbol{f}_2$ are slightly different; in this case $\boldsymbol{f}_2$ is smaller than the actual value for $\boldsymbol{F}_2$. As a consequence, the forecasting interval for $\boldsymbol{f}_2$ does not match exactly what would be the interval for the actual values of $\boldsymbol{F}_2$.[1] In this case the forecasting interval of $\boldsymbol{f}_2$ (blue solid lines) is narrower than the equivalent interval for the continuations (black dotted lines). This breach has a negative effect in C_m for the factor, as we can see in Table B.3, with coverage values around 86.50% for $h = 1$, and even more when the samples are smaller like in Table B.2.

Besides this point, we observe the same patterns for the factors than for the

[1]The interval for the actual values of $\boldsymbol{F}_2$ is obtained using continuations given the known value of $F_2(T = 100)$ and the known value ϕ_{F2}.

FIGURE B.1: Example of the estimation of $\boldsymbol{F}_2$ for a window with $T = 100$. Simulation of model with coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90.



series. This is to be expected given that, as previously explained, the time series are mainly the product of the common factors times some weights.

TABLE B.2: Results of Monte Carlo simulation, 10 000 replications. Two common factors following AR(1) models with normal errors. Model with coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. $T = 50$. Nominal coverage 95%.

Factor	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
F1	h=1	none	90.38 (0.079)	3.73 (0.006)	3.88 (0.002)	0.09
		BC	90.68 (0.081)	3.78 (0.006)	3.88 (0.002)	0.07
		RF	90.94 (0.079)	3.78 (0.006)	3.88 (0.002)	0.07
	h=5	none	85.33 (0.087)	6.87 (0.012)	8.27 (0.003)	0.27
		BC	89.02 (0.078)	7.69 (0.013)	8.27 (0.003)	0.13
		RF	90.76 (0.071)	7.87 (0.014)	8.27 (0.003)	0.09
	h=10	none	79.38 (0.116)	8.16 (0.020)	11.01 (0.005)	0.42
		BC	85.93 (0.111)	10.05 (0.025)	11.01 (0.005)	0.18
		RF	89.21 (0.100)	10.51 (0.025)	11.01 (0.005)	0.11
F2	h=1	none	82.34 (0.198)	2.02 (0.003)	1.94 (0.001)	0.17
		BC	82.42 (0.200)	2.03 (0.003)	1.94 (0.001)	0.18
		RF	82.40 (0.201)	2.03 (0.004)	1.94 (0.001)	0.18
	h=5	none	84.28 (0.097)	3.02 (0.006)	3.60 (0.002)	0.27
		BC	87.50 (0.093)	3.35 (0.008)	3.60 (0.002)	0.15
		RF	87.99 (0.094)	3.46 (0.008)	3.60 (0.002)	0.11
	h=10	none	82.28 (0.096)	3.18 (0.008)	4.17 (0.002)	0.37
		BC	86.58 (0.099)	3.77 (0.012)	4.17 (0.002)	0.18
		RF	87.44 (0.101)	4.01 (0.014)	4.17 (0.002)	0.12

TABLE B.3: Results of Monte Carlo simulation, 10 000 replications. Two common factors following AR(1) models with normal errors. Model with coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. $T = 100$. Nominal coverage 95%.

Factor	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
F1	h=1	none	92.85 (0.040)	3.83 (0.004)	3.88 (0.002)	0.04
		BC	92.92 (0.040)	3.85 (0.004)	3.88 (0.002)	0.03
		RF	93.08 (0.040)	3.86 (0.004)	3.88 (0.002)	0.03
	h=5	none	90.23 (0.054)	7.52 (0.009)	8.27 (0.003)	0.14
		BC	92.18 (0.046)	8.06 (0.010)	8.27 (0.003)	0.06
		RF	92.85 (0.043)	8.18 (0.010)	8.27 (0.003)	0.03
	h=10	none	86.75 (0.080)	9.32 (0.016)	11.02 (0.005)	0.24
		BC	90.60 (0.070)	10.68 (0.019)	11.02 (0.005)	0.08
		RF	91.90 (0.064)	10.98 (0.019)	11.02 (0.005)	0.04
F2	h=1	none	86.50 (0.158)	2.02 (0.002)	1.94 (0.001)	0.13
		BC	86.49 (0.160)	2.02 (0.002)	1.94 (0.001)	0.13
		RF	86.38 (0.161)	2.03 (0.002)	1.94 (0.001)	0.13
	h=5	none	89.68 (0.062)	3.32 (0.005)	3.60 (0.001)	0.13
		BC	91.39 (0.058)	3.53 (0.005)	3.60 (0.001)	0.06
		RF	91.40 (0.059)	3.57 (0.006)	3.60 (0.001)	0.05
	h=10	none	88.38 (0.066)	3.61 (0.006)	4.17 (0.002)	0.20
		BC	90.96 (0.063)	4.01 (0.008)	4.17 (0.002)	0.08
		RF	91.11 (0.065)	4.10 (0.009)	4.17 (0.002)	0.06

TABLE B.4: Results of Monte Carlo simulation, 10 000 replications. Two common factors following AR(1) models with normal errors. Model with coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. $T = 200$. Nominal coverage 95%.

Factor	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
F1	h=1	none	93.95 (0.024)	3.89 (0.003)	3.88 (0.002)	0.01
		BC	93.96 (0.024)	3.89 (0.003)	3.88 (0.002)	0.01
		RF	93.97 (0.025)	3.90 (0.003)	3.88 (0.002)	0.01
	h=5	none	92.76 (0.032)	7.92 (0.007)	8.27 (0.003)	0.07
		BC	93.73 (0.028)	8.25 (0.007)	8.27 (0.003)	0.02
		RF	93.88 (0.027)	8.32 (0.007)	8.27 (0.003)	0.02
	h=10	none	91.06 (0.048)	10.13 (0.012)	11.01 (0.005)	0.12
		BC	93.04 (0.041)	11.00 (0.013)	11.01 (0.005)	0.02
		RF	93.40 (0.040)	11.21 (0.014)	11.01 (0.005)	0.04
F2	h=1	none	89.96 (0.111)	2.01 (0.002)	1.94 (0.001)	0.09
		BC	89.94 (0.112)	2.02 (0.002)	1.94 (0.001)	0.09
		RF	89.97 (0.112)	2.02 (0.002)	1.94 (0.001)	0.09
	h=5	none	92.59 (0.038)	3.50 (0.003)	3.59 (0.002)	0.05
		BC	93.41 (0.035)	3.61 (0.004)	3.59 (0.002)	0.02
		RF	93.41 (0.036)	3.62 (0.004)	3.59 (0.002)	0.02
	h=10	none	91.81 (0.042)	3.91 (0.005)	4.17 (0.002)	0.10
		BC	93.19 (0.039)	4.13 (0.006)	4.17 (0.002)	0.03
		RF	93.18 (0.040)	4.14 (0.006)	4.17 (0.002)	0.03

B.2 Results for Non-Gaussian Errors

In this section we re-run the simulations for AR(1) common factors introducing non normal errors. In particular, the innovations $\eta_{i,t}$ (in (3.3)) follow a centred $\chi^2(5)$ distribution (as in Clements and Kim, 2007). The main difference of this distribution with the normal is that it is not symmetrical. Table B.5 is the analogous to Table B.1. Notice that the bias does not seem to worsen with the new distribution of $\eta_{i,t}$. *BC* and *RF* continue to improve upon *none*, in a similar measure to the case of normally distributed errors. Still, as expected, the estimation without any bias correction gets closer to the true AR coefficients (ϕ_{F1}, ϕ_{F2}) as the sample (in the time dimension, T) gets larger.

TABLE B.5: Bias for common factors following AR(1) processes with centred $\chi^2(5)$ errors. Between parenthesis, the variance of the AR estimated coefficients. Monte Carlo simulations of model with coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. 10 000 MC replications.

factor	method	T=50	T=100	T=200
F1	none	0.083 (0.004)	0.042 (0.001)	0.022 (0.001)
	BC	0.015 (0.003)	0.004 (0.001)	0.001 (0.001)
	RF	0.002 (0.003)	-0.003 (0.001)	-0.004 (0.001)
F2	none	0.177 (0.013)	0.078 (0.004)	0.035 (0.001)
	BC	0.111 (0.015)	0.042 (0.004)	0.016 (0.002)
	RF	0.096 (0.017)	0.036 (0.005)	0.015 (0.002)

The results in terms of coverage, interval length, and CQ_m, are similar to those for the process with normal errors, revealing that the bias-corrections for the underlying non observed factors improve forecasting results even when they are not normally distributed. For the five selected series, results are presented in Tables B.6-B.8.

Table B.6: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created using two common factors, both following AR(1) models with centred $\chi^2(5)$ errors. Model with coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. $T = 50$. Nominal coverage 95%.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	91.89 (0.044)	4.60 (0.008)	4.79 (0.003)	0.07
		BC	92.26 (0.042)	4.64 (0.008)	4.79 (0.003)	0.06
		RF	92.34 (0.041)	4.66 (0.008)	4.79 (0.003)	0.06
	h=5	none	85.02 (0.088)	7.36 (0.015)	9.05 (0.004)	0.29
		BC	88.64 (0.081)	8.31 (0.018)	9.05 (0.004)	0.15
		RF	89.44 (0.079)	8.58 (0.019)	9.05 (0.004)	0.11
	h=10	none	81.15 (0.106)	8.23 (0.021)	10.99 (0.005)	0.40
		BC	87.09 (0.103)	10.24 (0.030)	10.99 (0.005)	0.15
		RF	88.61 (0.100)	10.86 (0.032)	10.99 (0.005)	0.08
Y2	h=1	none	92.26 (0.056)	2.29 (0.004)	2.35 (0.001)	0.05
		BC	93.05 (0.048)	2.31 (0.004)	2.35 (0.001)	0.04
		RF	93.23 (0.047)	2.32 (0.004)	2.35 (0.001)	0.03
	h=5	none	82.44 (0.105)	3.93 (0.008)	5.02 (0.002)	0.35
		BC	87.93 (0.091)	4.45 (0.010)	5.02 (0.002)	0.19
		RF	89.08 (0.089)	4.57 (0.010)	5.02 (0.002)	0.15
	h=10	none	75.73 (0.129)	4.53 (0.012)	6.70 (0.003)	0.53
		BC	84.74 (0.122)	5.67 (0.016)	6.70 (0.003)	0.26
		RF	86.88 (0.119)	5.99 (0.017)	6.70 (0.003)	0.19
Y5	h=1	none	91.75 (0.042)	5.98 (0.010)	6.25 (0.003)	0.08
		BC	92.23 (0.039)	6.05 (0.010)	6.25 (0.003)	0.06
		RF	92.35 (0.038)	6.07 (0.010)	6.25 (0.003)	0.06
	h=5	none	83.94 (0.093)	9.77 (0.020)	12.28 (0.005)	0.32
		BC	88.36 (0.082)	11.04 (0.023)	12.28 (0.005)	0.17
		RF	89.25 (0.081)	11.37 (0.024)	12.28 (0.005)	0.13
	h=10	none	78.45 (0.116)	11.07 (0.029)	15.65 (0.007)	0.47
		BC	85.89 (0.112)	13.82 (0.039)	15.65 (0.007)	0.21
		RF	87.70 (0.109)	14.62 (0.042)	15.65 (0.007)	0.14
Y10	h=1	none	92.79 (0.051)	6.21 (0.011)	6.29 (0.004)	0.04
		BC	93.04 (0.049)	6.26 (0.011)	6.29 (0.004)	0.02
		RF	93.09 (0.048)	6.29 (0.011)	6.29 (0.004)	0.02
	h=5	none	85.78 (0.089)	9.92 (0.021)	11.92 (0.005)	0.26
		BC	88.89 (0.083)	11.20 (0.025)	11.92 (0.005)	0.12
		RF	89.62 (0.080)	11.57 (0.027)	11.92 (0.005)	0.09
	h=10	none	83.10 (0.102)	11.01 (0.028)	13.96 (0.006)	0.34
		BC	88.00 (0.101)	13.67 (0.041)	13.96 (0.006)	0.09
		RF	89.35 (0.095)	14.53 (0.045)	13.96 (0.006)	0.10
Y25	h=1	none	92.52 (0.053)	8.77 (0.015)	8.99 (0.005)	0.05
		BC	92.76 (0.051)	8.84 (0.015)	8.99 (0.005)	0.04
		RF	92.80 (0.050)	8.86 (0.015)	8.99 (0.005)	0.04
	h=5	none	85.56 (0.089)	14.22 (0.029)	17.13 (0.008)	0.27
		BC	88.77 (0.084)	16.06 (0.036)	17.13 (0.008)	0.13
		RF	89.52 (0.081)	16.58 (0.038)	17.13 (0.008)	0.09
	h=10	none	82.68 (0.105)	15.85 (0.040)	20.19 (0.009)	0.34
		BC	87.80 (0.103)	19.68 (0.058)	20.19 (0.009)	0.10
		RF	89.22 (0.097)	20.93 (0.064)	20.19 (0.009)	0.10

Table B.7: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created using two common factors, both following AR(1) models with centred $\chi^2(5)$ errors. Model with coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. $T = 100$. Nominal coverage 95%.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	93.27 (0.029)	4.68 (0.006)	4.79 (0.003)	0.04
		BC	93.36 (0.028)	4.70 (0.006)	4.79 (0.003)	0.04
		RF	93.41 (0.028)	4.71 (0.006)	4.79 (0.003)	0.03
	h=5	none	89.87 (0.055)	8.11 (0.011)	9.04 (0.004)	0.16
		BC	91.67 (0.050)	8.69 (0.013)	9.04 (0.004)	0.07
		RF	91.87 (0.051)	8.83 (0.013)	9.04 (0.004)	0.06
	h=10	none	87.47 (0.073)	9.36 (0.017)	10.98 (0.005)	0.23
		BC	90.73 (0.067)	10.65 (0.021)	10.98 (0.005)	0.07
		RF	91.22 (0.067)	10.99 (0.023)	10.98 (0.005)	0.04
Y2	h=1	none	93.88 (0.036)	2.33 (0.003)	2.34 (0.001)	0.02
		BC	94.17 (0.034)	2.34 (0.003)	2.34 (0.001)	0.01
		RF	94.29 (0.033)	2.35 (0.003)	2.34 (0.001)	0.01
	h=5	none	89.13 (0.065)	4.46 (0.006)	5.01 (0.002)	0.17
		BC	91.80 (0.054)	4.79 (0.007)	5.01 (0.002)	0.08
		RF	92.37 (0.052)	4.88 (0.007)	5.01 (0.002)	0.05
	h=10	none	85.04 (0.089)	5.44 (0.010)	6.69 (0.003)	0.29
		BC	90.05 (0.077)	6.28 (0.012)	6.69 (0.003)	0.11
		RF	91.10 (0.073)	6.49 (0.013)	6.69 (0.003)	0.07
Y5	h=1	none	93.24 (0.027)	6.11 (0.007)	6.25 (0.003)	0.04
		BC	93.38 (0.026)	6.14 (0.007)	6.25 (0.003)	0.03
		RF	93.43 (0.026)	6.15 (0.007)	6.25 (0.003)	0.03
	h=5	none	89.56 (0.057)	10.93 (0.015)	12.27 (0.006)	0.17
		BC	91.76 (0.049)	11.74 (0.017)	12.27 (0.006)	0.08
		RF	92.12 (0.048)	11.94 (0.017)	12.27 (0.006)	0.06
	h=10	none	86.32 (0.080)	12.99 (0.024)	15.63 (0.007)	0.26
		BC	90.43 (0.070)	14.88 (0.029)	15.63 (0.007)	0.10
		RF	91.19 (0.068)	15.37 (0.031)	15.63 (0.007)	0.06
Y10	h=1	none	94.15 (0.033)	6.30 (0.008)	6.29 (0.004)	0.01
		BC	94.18 (0.033)	6.32 (0.008)	6.29 (0.004)	0.01
		RF	94.15 (0.034)	6.33 (0.008)	6.29 (0.004)	0.02
	h=5	none	90.17 (0.057)	10.81 (0.016)	11.91 (0.005)	0.14
		BC	91.74 (0.054)	11.57 (0.017)	11.91 (0.005)	0.06
		RF	91.81 (0.055)	11.75 (0.018)	11.91 (0.005)	0.05
	h=10	none	88.42 (0.070)	12.24 (0.022)	13.95 (0.006)	0.19
		BC	91.01 (0.068)	13.85 (0.028)	13.95 (0.006)	0.05
		RF	91.29 (0.068)	14.26 (0.031)	13.95 (0.006)	0.06
Y25	h=1	none	93.90 (0.036)	8.89 (0.010)	8.99 (0.005)	0.02
		BC	93.92 (0.036)	8.92 (0.011)	8.99 (0.005)	0.02
		RF	93.87 (0.037)	8.93 (0.010)	8.99 (0.005)	0.02
	h=5	none	89.97 (0.058)	15.48 (0.022)	17.11 (0.008)	0.15
		BC	91.67 (0.054)	16.60 (0.025)	17.11 (0.008)	0.07
		RF	91.76 (0.054)	16.85 (0.026)	17.11 (0.008)	0.05
	h=10	none	88.19 (0.071)	17.62 (0.032)	20.18 (0.009)	0.20
		BC	90.97 (0.068)	19.98 (0.041)	20.18 (0.009)	0.05
		RF	91.28 (0.068)	20.59 (0.044)	20.18 (0.009)	0.06

Table B.8: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created using two common factors, both following AR(1) models with centred $\chi^2(5)$ errors. Model with coefficients ϕ_{F1} =0.975, ϕ_{F2} =0.90. $T = 200$. Nominal coverage 95%.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	93.84 (0.022)	4.72 (0.004)	4.79 (0.003)	0.03
		BC	93.88 (0.021)	4.72 (0.004)	4.79 (0.003)	0.03
		RF	93.85 (0.022)	4.73 (0.004)	4.79 (0.003)	0.03
	h=5	none	92.35 (0.034)	8.55 (0.008)	9.05 (0.004)	0.08
		BC	93.21 (0.031)	8.88 (0.009)	9.05 (0.004)	0.04
		RF	93.22 (0.032)	8.92 (0.009)	9.05 (0.004)	0.03
	h=10	none	91.13 (0.046)	10.10 (0.013)	10.99 (0.005)	0.12
		BC	92.73 (0.042)	10.84 (0.014)	10.99 (0.005)	0.04
		RF	92.85 (0.042)	10.97 (0.015)	10.99 (0.005)	0.02
Y2	h=1	none	94.21 (0.029)	2.33 (0.002)	2.34 (0.001)	0.01
		BC	94.33 (0.029)	2.34 (0.002)	2.34 (0.001)	0.01
		RF	94.32 (0.029)	2.34 (0.002)	2.34 (0.001)	0.01
	h=5	none	92.25 (0.039)	4.75 (0.005)	5.01 (0.002)	0.08
		BC	93.45 (0.034)	4.96 (0.005)	5.01 (0.002)	0.03
		RF	93.58 (0.033)	5.00 (0.005)	5.01 (0.002)	0.02
	h=10	none	90.23 (0.056)	6.05 (0.008)	6.69 (0.003)	0.15
		BC	92.57 (0.048)	6.59 (0.009)	6.69 (0.003)	0.04
		RF	92.90 (0.046)	6.72 (0.009)	6.69 (0.003)	0.03
Y5	h=1	none	93.82 (0.020)	6.16 (0.005)	6.25 (0.003)	0.03
		BC	93.87 (0.020)	6.17 (0.005)	6.25 (0.003)	0.03
		RF	93.86 (0.020)	6.17 (0.005)	6.25 (0.003)	0.02
	h=5	none	92.38 (0.033)	11.61 (0.011)	12.27 (0.006)	0.08
		BC	93.37 (0.030)	12.09 (0.011)	12.27 (0.006)	0.03
		RF	93.44 (0.030)	12.17 (0.012)	12.27 (0.006)	0.02
	h=10	none	90.78 (0.048)	14.26 (0.018)	15.64 (0.007)	0.13
		BC	92.74 (0.043)	15.44 (0.020)	15.64 (0.007)	0.04
		RF	92.96 (0.042)	15.68 (0.021)	15.64 (0.007)	0.02
Y10	h=1	none	94.56 (0.026)	6.32 (0.006)	6.29 (0.004)	0.01
		BC	94.58 (0.026)	6.32 (0.006)	6.29 (0.004)	0.01
		RF	94.53 (0.027)	6.32 (0.006)	6.29 (0.004)	0.01
	h=5	none	92.41 (0.038)	11.31 (0.012)	11.92 (0.005)	0.08
		BC	93.19 (0.036)	11.72 (0.012)	11.92 (0.005)	0.04
		RF	93.15 (0.038)	11.76 (0.013)	11.92 (0.005)	0.03
	h=10	none	91.37 (0.047)	12.99 (0.017)	13.97 (0.006)	0.11
		BC	92.70 (0.045)	13.82 (0.019)	13.97 (0.006)	0.03
		RF	92.69 (0.047)	13.92 (0.020)	13.97 (0.006)	0.03
Y25	h=1	none	94.34 (0.028)	8.95 (0.008)	8.99 (0.005)	0.01
		BC	94.36 (0.028)	8.95 (0.008)	8.99 (0.005)	0.01
		RF	94.32 (0.029)	8.96 (0.008)	8.99 (0.005)	0.01
	h=5	none	92.27 (0.038)	16.22 (0.016)	17.12 (0.008)	0.08
		BC	93.09 (0.037)	16.79 (0.017)	17.12 (0.008)	0.04
		RF	93.07 (0.038)	16.86 (0.018)	17.12 (0.008)	0.04
	h=10	none	91.18 (0.048)	18.71 (0.024)	20.20 (0.009)	0.11
		BC	92.59 (0.046)	19.94 (0.028)	20.20 (0.009)	0.04
		RF	92.62 (0.048)	20.11 (0.028)	20.20 (0.009)	0.03

B.3 Number of Factors and AR Orders Unknown for T=50 and T=200

This section complements the results presented in Section 3.4.2, considering the alternative sample sizes $T = 50$ (Tables B.9 and B.10) and $T = 200$ (Tables B.11 and B.12). It continues to be the case that the indicators C_m, L_m, and CQ are equal or better for BC and RF than for *none*. Consistently with the results obtained in previous sections, the results are enhanced as the forecasting horizon h increases.

Additionally, for all the sample sizes (T), the results of employing BIC in the selection of p outperform those of employing AICc. The performance of these criteria is recorded in Tables B.13 and B.14.

Last, it must be acknowledged that there is a sharp deterioration of results when the time frame reduces to $T = 50$. In this regard, it should be reminded that the common factors are estimates themselves and their accuracy improves with the sample size.

Table B.9: Results of Monte Carlo simulation, 10.000 replications. Five representative time series created with both common factors following AR(2) processes with normal errors and coefficients ϕ_1^{F1} =1.475, $\phi_2^{F1} = -0.4875$, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. T = 50. Nominal coverage 95%. IC_3 used in the estimation of R, BIC used in the selection of $\hat{p}$.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	88.37 (0.071)	0.96 (0.001)	1.05 (0.000)	0.15
		BC	89.00 (0.067)	0.97 (0.002)	1.05 (0.000)	0.14
		RF	89.39 (0.065)	0.98 (0.002)	1.05 (0.000)	0.13
	h=5	none	80.58 (0.132)	2.67 (0.006)	3.29 (0.001)	0.34
		BC	84.91 (0.119)	2.97 (0.007)	3.29 (0.001)	0.20
		RF	86.27 (0.106)	2.95 (0.007)	3.29 (0.001)	0.19
	h=10	none	73.98 (0.158)	3.31 (0.010)	4.61 (0.002)	0.50
		BC	81.20 (0.155)	4.08 (0.013)	4.61 (0.002)	0.26
		RF	84.26 (0.128)	4.11 (0.013)	4.61 (0.002)	0.22
Y2	h=1	none	84.65 (0.078)	0.69 (0.001)	0.84 (0.000)	0.29
		BC	85.30 (0.075)	0.69 (0.001)	0.84 (0.000)	0.28
		RF	85.75 (0.071)	0.70 (0.001)	0.84 (0.000)	0.27
	h=5	none	79.37 (0.139)	2.07 (0.005)	2.65 (0.001)	0.38
		BC	83.66 (0.131)	2.29 (0.006)	2.65 (0.001)	0.25
		RF	85.60 (0.113)	2.29 (0.006)	2.65 (0.001)	0.24
	h=10	none	71.74 (0.170)	2.68 (0.009)	3.89 (0.002)	0.56
		BC	79.31 (0.172)	3.31 (0.012)	3.89 (0.002)	0.31
		RF	83.17 (0.144)	3.33 (0.011)	3.89 (0.002)	0.27
Y5	h=1	none	88.57 (0.068)	1.50 (0.002)	1.65 (0.001)	0.16
		BC	89.17 (0.065)	1.52 (0.002)	1.65 (0.001)	0.14
		RF	89.61 (0.062)	1.52 (0.002)	1.65 (0.001)	0.13
	h=5	none	80.03 (0.134)	4.42 (0.011)	5.55 (0.002)	0.36
		BC	84.34 (0.124)	4.90 (0.012)	5.55 (0.002)	0.23
		RF	86.08 (0.108)	4.89 (0.012)	5.55 (0.002)	0.21
	h=10	none	72.55 (0.165)	5.64 (0.018)	8.08 (0.003)	0.54
		BC	80.08 (0.164)	6.97 (0.024)	8.08 (0.003)	0.29
		RF	83.70 (0.136)	7.02 (0.023)	8.08 (0.003)	0.25
Y10	h=1	none	90.11 (0.074)	1.13 (0.002)	1.14 (0.000)	0.06
		BC	90.67 (0.070)	1.14 (0.002)	1.14 (0.000)	0.05
		RF	90.95 (0.067)	1.14 (0.002)	1.14 (0.000)	0.05
	h=5	none	81.74 (0.136)	2.85 (0.007)	3.35 (0.001)	0.29
		BC	85.80 (0.123)	3.17 (0.008)	3.35 (0.001)	0.15
		RF	86.55 (0.112)	3.15 (0.008)	3.35 (0.001)	0.15
	h=10	none	77.00 (0.151)	3.33 (0.010)	4.37 (0.002)	0.43
		BC	82.93 (0.149)	4.10 (0.014)	4.37 (0.002)	0.19
		RF	84.97 (0.125)	4.13 (0.014)	4.37 (0.002)	0.16
Y25	h=1	none	90.88 (0.070)	1.66 (0.003)	1.64 (0.001)	0.06
		BC	91.42 (0.066)	1.68 (0.003)	1.64 (0.001)	0.06
		RF	91.66 (0.064)	1.68 (0.003)	1.64 (0.001)	0.06
	h=5	none	81.76 (0.134)	4.28 (0.011)	5.06 (0.002)	0.29
		BC	85.85 (0.121)	4.77 (0.012)	5.06 (0.002)	0.15
		RF	86.75 (0.109)	4.74 (0.012)	5.06 (0.002)	0.15
	h=10	none	76.47 (0.154)	5.08 (0.015)	6.72 (0.003)	0.44
		BC	82.62 (0.151)	6.26 (0.021)	6.72 (0.003)	0.20
		RF	84.93 (0.126)	6.31 (0.021)	6.72 (0.003)	0.17

TABLE B.10: Results of Monte Carlo simulation, 10.000 replications. Five representative time series created with both common factors following AR(2) processes with normal errors and coefficients ϕ_1^{F1} =1.475, $\phi_2^{F1} = -0.4875$, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. $T = 50$. Nominal coverage 95%. R estimated with IC_3, AICc used in the selection of $\hat{p}$.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	87.43 (0.078)	0.95 (0.001)	1.05 (0.000)	0.18
		BC	88.07 (0.073)	0.96 (0.002)	1.05 (0.000)	0.16
		RF	88.47 (0.070)	0.96 (0.002)	1.05 (0.000)	0.16
	h=5	none	79.30 (0.140)	2.64 (0.006)	3.29 (0.001)	0.36
		BC	83.75 (0.128)	2.93 (0.007)	3.29 (0.001)	0.23
		RF	85.26 (0.111)	2.91 (0.007)	3.29 (0.001)	0.22
	h=10	none	72.43 (0.166)	3.29 (0.010)	4.61 (0.002)	0.52
		BC	79.80 (0.164)	4.06 (0.014)	4.61 (0.002)	0.28
		RF	83.05 (0.135)	4.05 (0.013)	4.61 (0.002)	0.25
Y2	h=1	none	83.66 (0.082)	0.68 (0.001)	0.84 (0.000)	0.31
		BC	84.39 (0.081)	0.69 (0.001)	0.84 (0.000)	0.30
		RF	84.89 (0.076)	0.69 (0.001)	0.84 (0.000)	0.29
	h=5	none	78.06 (0.141)	2.03 (0.005)	2.65 (0.001)	0.41
		BC	82.65 (0.133)	2.26 (0.006)	2.65 (0.001)	0.28
		RF	84.60 (0.116)	2.25 (0.005)	2.65 (0.001)	0.26
	h=10	none	70.13 (0.174)	2.64 (0.009)	3.88 (0.002)	0.58
		BC	78.04 (0.177)	3.27 (0.012)	3.88 (0.002)	0.34
		RF	81.97 (0.148)	3.26 (0.011)	3.88 (0.002)	0.30
Y5	h=1	none	87.64 (0.075)	1.48 (0.002)	1.65 (0.001)	0.18
		BC	88.29 (0.072)	1.50 (0.002)	1.65 (0.001)	0.16
		RF	88.76 (0.067)	1.50 (0.002)	1.65 (0.001)	0.15
	h=5	none	78.73 (0.140)	4.36 (0.011)	5.55 (0.002)	0.39
		BC	83.23 (0.130)	4.84 (0.012)	5.55 (0.002)	0.25
		RF	85.09 (0.113)	4.82 (0.012)	5.55 (0.002)	0.24
	h=10	none	70.98 (0.171)	5.60 (0.018)	8.07 (0.003)	0.56
		BC	78.76 (0.171)	6.92 (0.024)	8.07 (0.003)	0.31
		RF	82.49 (0.142)	6.90 (0.023)	8.07 (0.003)	0.28
Y10	h=1	none	89.10 (0.081)	1.11 (0.002)	1.14 (0.000)	0.09
		BC	89.64 (0.078)	1.12 (0.002)	1.14 (0.000)	0.07
		RF	89.98 (0.074)	1.12 (0.002)	1.14 (0.000)	0.07
	h=5	none	80.35 (0.145)	2.81 (0.007)	3.35 (0.001)	0.32
		BC	84.45 (0.135)	3.13 (0.008)	3.35 (0.001)	0.18
		RF	85.37 (0.120)	3.09 (0.008)	3.35 (0.001)	0.18
	h=10	none	75.51 (0.160)	3.31 (0.010)	4.37 (0.002)	0.45
		BC	81.48 (0.160)	4.06 (0.014)	4.37 (0.002)	0.21
		RF	83.64 (0.133)	4.04 (0.014)	4.37 (0.002)	0.19
Y25	h=1	none	89.76 (0.077)	1.62 (0.003)	1.64 (0.001)	0.06
		BC	90.36 (0.074)	1.64 (0.003)	1.64 (0.001)	0.05
		RF	90.62 (0.070)	1.64 (0.003)	1.64 (0.001)	0.05
	h=5	none	80.14 (0.143)	4.20 (0.010)	5.05 (0.002)	0.32
		BC	84.41 (0.131)	4.68 (0.012)	5.05 (0.002)	0.19
		RF	85.41 (0.118)	4.63 (0.012)	5.05 (0.002)	0.19
	h=10	none	74.61 (0.160)	5.01 (0.015)	6.72 (0.003)	0.47
		BC	81.12 (0.159)	6.16 (0.021)	6.72 (0.003)	0.23
		RF	83.49 (0.132)	6.13 (0.020)	6.72 (0.003)	0.21

TABLE B.11: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created with both common factors following AR(2) processes with normal errors and coefficients ϕ_1^{F1} =1.475, ϕ_2^{F1} =-0.4875, ϕ_1^{F2} =1.4,ϕ_2^{F2} =-0.45. $T = 200$. Nominal coverage 95%. R estimated with IC_3, BIC used in the selection of $\hat{p}$.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	92.44 (0.030)	1.01 (0.001)	1.05 (0.000)	0.07
		BC	92.44 (0.030)	1.01 (0.001)	1.05 (0.000)	0.07
		RF	92.47 (0.030)	1.01 (0.001)	1.05 (0.000)	0.07
	h=5	none	92.25 (0.040)	3.15 (0.004)	3.29 (0.001)	0.07
		BC	93.02 (0.036)	3.24 (0.004)	3.29 (0.001)	0.04
		RF	92.89 (0.036)	3.23 (0.004)	3.29 (0.001)	0.04
	h=10	none	90.51 (0.055)	4.25 (0.006)	4.61 (0.002)	0.13
		BC	92.32 (0.048)	4.55 (0.007)	4.61 (0.002)	0.04
		RF	92.29 (0.047)	4.56 (0.007)	4.61 (0.002)	0.04
Y2	h=1	none	90.45 (0.031)	0.75 (0.001)	0.84 (0.000)	0.16
		BC	90.50 (0.030)	0.75 (0.001)	0.84 (0.000)	0.16
		RF	90.52 (0.030)	0.75 (0.001)	0.84 (0.000)	0.16
	h=5	none	92.05 (0.039)	2.51 (0.003)	2.65 (0.001)	0.08
		BC	92.79 (0.036)	2.58 (0.003)	2.65 (0.001)	0.05
		RF	92.74 (0.035)	2.58 (0.003)	2.65 (0.001)	0.05
	h=10	none	90.24 (0.057)	3.55 (0.005)	3.89 (0.002)	0.14
		BC	92.08 (0.051)	3.82 (0.006)	3.89 (0.002)	0.05
		RF	92.20 (0.048)	3.83 (0.006)	3.89 (0.002)	0.04
Y5	h=1	none	93.32 (0.027)	1.62 (0.001)	1.65 (0.001)	0.04
		BC	93.34 (0.026)	1.62 (0.001)	1.65 (0.001)	0.03
		RF	93.37 (0.026)	1.62 (0.001)	1.65 (0.001)	0.03
	h=5	none	92.31 (0.039)	5.31 (0.006)	5.55 (0.002)	0.07
		BC	93.05 (0.036)	5.48 (0.006)	5.55 (0.002)	0.03
		RF	92.97 (0.035)	5.46 (0.006)	5.55 (0.002)	0.04
	h=10	none	90.38 (0.057)	7.42 (0.011)	8.08 (0.003)	0.13
		BC	92.24 (0.050)	7.97 (0.012)	8.08 (0.003)	0.04
		RF	92.32 (0.047)	8.00 (0.012)	8.08 (0.003)	0.04
Y10	h=1	none	93.15 (0.030)	1.13 (0.001)	1.13 (0.000)	0.02
		BC	93.18 (0.030)	1.13 (0.001)	1.13 (0.000)	0.02
		RF	93.15 (0.030)	1.13 (0.001)	1.13 (0.000)	0.02
	h=5	none	92.50 (0.038)	3.24 (0.004)	3.35 (0.001)	0.06
		BC	93.24 (0.035)	3.34 (0.004)	3.35 (0.001)	0.02
		RF	93.05 (0.036)	3.32 (0.004)	3.35 (0.001)	0.03
	h=10	none	91.00 (0.051)	4.07 (0.006)	4.37 (0.002)	0.11
		BC	92.54 (0.047)	4.34 (0.006)	4.37 (0.002)	0.03
		RF	92.37 (0.048)	4.33 (0.006)	4.37 (0.002)	0.04
Y25	h=1	none	93.84 (0.029)	1.67 (0.001)	1.64 (0.001)	0.03
		BC	93.89 (0.028)	1.67 (0.001)	1.64 (0.001)	0.03
		RF	93.87 (0.028)	1.67 (0.001)	1.64 (0.001)	0.03
	h=5	none	92.52 (0.038)	4.89 (0.006)	5.05 (0.002)	0.06
		BC	93.21 (0.035)	5.03 (0.006)	5.05 (0.002)	0.02
		RF	93.08 (0.036)	5.02 (0.006)	5.05 (0.002)	0.03
	h=10	none	90.92 (0.052)	6.24 (0.009)	6.72 (0.003)	0.11
		BC	92.51 (0.047)	6.67 (0.010)	6.72 (0.003)	0.03
		RF	92.40 (0.047)	6.67 (0.010)	6.72 (0.003)	0.03

TABLE B.12: Results of Monte Carlo simulation, 10 000 replications. Five representative time series created with both common factors following AR(2) processes with normal errors and coefficients ϕ_1^{F1} =1.475, ϕ_2^{F1} =-0.4875, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. T = 200. Nominal coverage 95%. R estimated with IC_3, AICc used in the selection of $\hat{p}$.

Series	Horizon	Correction	C_m (se)	L_m (se)	L_t (se)	CQ_m
Y1	h=1	none	92.03 (0.032)	1.00 (0.001)	1.05 (0.000)	0.08
		BC	92.08 (0.031)	1.00 (0.001)	1.05 (0.000)	0.08
		RF	92.06 (0.032)	1.00 (0.001)	1.05 (0.000)	0.08
	h=5	none	91.98 (0.042)	3.14 (0.004)	3.29 (0.001)	0.08
		BC	92.76 (0.038)	3.23 (0.004)	3.29 (0.001)	0.04
		RF	92.66 (0.037)	3.22 (0.004)	3.29 (0.001)	0.05
	h=10	none	90.36 (0.058)	4.26 (0.006)	4.61 (0.002)	0.12
		BC	92.11 (0.051)	4.55 (0.007)	4.61 (0.002)	0.04
		RF	92.05 (0.049)	4.54 (0.007)	4.61 (0.002)	0.05
Y2	h=1	none	89.99 (0.033)	0.74 (0.001)	0.84 (0.000)	0.17
		BC	90.06 (0.032)	0.74 (0.001)	0.84 (0.000)	0.17
		RF	90.13 (0.032)	0.74 (0.001)	0.84 (0.000)	0.17
	h=5	none	91.73 (0.041)	2.49 (0.003)	2.65 (0.001)	0.09
		BC	92.53 (0.038)	2.57 (0.003)	2.65 (0.001)	0.06
		RF	92.46 (0.036)	2.56 (0.003)	2.65 (0.001)	0.06
	h=10	none	90.04 (0.060)	3.55 (0.005)	3.88 (0.002)	0.14
		BC	91.86 (0.054)	3.81 (0.006)	3.88 (0.002)	0.05
		RF	91.92 (0.049)	3.80 (0.006)	3.88 (0.002)	0.05
Y5	h=1	none	92.90 (0.028)	1.60 (0.001)	1.65 (0.001)	0.05
		BC	92.97 (0.028)	1.61 (0.001)	1.65 (0.001)	0.05
		RF	92.98 (0.028)	1.61 (0.001)	1.65 (0.001)	0.05
	h=5	none	92.01 (0.041)	5.29 (0.006)	5.55 (0.002)	0.08
		BC	92.78 (0.037)	5.45 (0.006)	5.55 (0.002)	0.04
		RF	92.69 (0.037)	5.43 (0.006)	5.55 (0.002)	0.05
	h=10	none	90.21 (0.060)	7.44 (0.011)	8.07 (0.003)	0.13
		BC	92.03 (0.053)	7.96 (0.012)	8.07 (0.003)	0.05
		RF	92.05 (0.049)	7.94 (0.012)	8.07 (0.003)	0.05
Y10	h=1	none	92.85 (0.032)	1.12 (0.001)	1.14 (0.000)	0.03
		BC	92.89 (0.032)	1.12 (0.001)	1.14 (0.000)	0.03
		RF	92.86 (0.032)	1.12 (0.001)	1.14 (0.000)	0.03
	h=5	none	92.32 (0.040)	3.25 (0.004)	3.35 (0.001)	0.06
		BC	93.03 (0.038)	3.34 (0.004)	3.35 (0.001)	0.03
		RF	92.86 (0.039)	3.32 (0.004)	3.35 (0.001)	0.03
	h=10	none	90.86 (0.054)	4.09 (0.006)	4.37 (0.002)	0.11
		BC	92.36 (0.050)	4.35 (0.007)	4.37 (0.002)	0.03
		RF	92.17 (0.050)	4.33 (0.007)	4.37 (0.002)	0.04
Y25	h=1	none	93.54 (0.031)	1.66 (0.001)	1.64 (0.001)	0.03
		BC	93.59 (0.030)	1.66 (0.001)	1.64 (0.001)	0.03
		RF	93.59 (0.030)	1.67 (0.001)	1.64 (0.001)	0.03
	h=5	none	92.34 (0.040)	4.89 (0.006)	5.05 (0.002)	0.06
		BC	93.05 (0.037)	5.03 (0.006)	5.05 (0.002)	0.03
		RF	92.87 (0.038)	5.00 (0.006)	5.05 (0.002)	0.03
	h=10	none	90.77 (0.054)	6.27 (0.009)	6.72 (0.003)	0.11
		BC	92.38 (0.048)	6.68 (0.010)	6.72 (0.003)	0.03
		RF	92.15 (0.049)	6.65 (0.010)	6.72 (0.003)	0.04

TABLE B.13: Comparison of relative frequencies in the estimation of $\hat{p}$ by BIC and AICc. The values correspond to a Monte Carlo simulation with 10 000 replications. Two common factors, both following AR(2) models with normal errors. Model with coefficients ϕ_1^{F1} =1.475, ϕ_2^{F1} =-0.4875, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. $T = 50$. Nominal coverage 95%.

Factor	$\hat{p} = 1$	$\hat{p} = 2$	$\hat{p} = 3$	$\hat{p} = 4$	$\hat{p} = 5$	$\hat{p} = 6$
BIC						
F1	4.84	64.87	11.91	7.32	5.64	5.42
F2	10.09	59.03	13.18	7.46	5.04	5.20
AICc						
F1	1.24	37.05	15.65	13.51	13.58	18.97
F2	2.70	34.17	16.71	13.86	13.81	18.75

TABLE B.14: Comparison of relative frequencies in the estimation of $\hat{p}$ by BIC and AICc. The values correspond to a Monte Carlo simulation with 10 000 replications. Two common factors, both following AR(2) models with normal errors. Model with coefficients ϕ_1^{F1} =1.475, ϕ_2^{F1} =-0.4875, ϕ_1^{F2} =1.4, ϕ_2^{F2} =-0.45. $T = 200$. Nominal coverage 95%.

Factor	$\hat{p} = 1$	$\hat{p} = 2$	$\hat{p} = 3$	$\hat{p} = 4$	$\hat{p} = 5$	$\hat{p} = 6$
BIC						
F1	0.00	89.33	7.56	2.07	0.77	0.27
F2	0.03	85.96	10.44	2.39	0.77	0.41
AICc						
F1	0.00	36.91	16.35	14.14	13.68	18.92
F2	0.00	32.78	18.89	14.00	13.98	20.35

Bibliography

S.C. Ahn and A.R. Horenstein. Eigenvalue ratio test for the number of factors. *Econometrica*, 81:1203–1227, 2013.

A.M. Alonso, D. Peña, and J. Romo. Forecasting time series with sieve bootstrap. *Journal of Statistical Planning and Inference*, 100:1–11, 2002.

A.M. Alonso, D. Peña, and J. Romo. Introducing model uncertainty in time series bootstrap. *Statistica Sinica*, 14:155–174, 2004.

A.M. Alonso, D. Peña, and J. Rodríguez. A methodology for population projections: an application to Spain. *Working Paper 08 - 45 Statistics and Econometrics Series 12*, Available online: `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.584.2593&rep=rep1&type=pdf`, 2008.

A.M. Alonso, C. García-Martos, J. Rodríguez, and M.J. Sánchez. Seasonal dynamic factor analysis and bootstrap inference: Application to electricity market forecasting. *Technometrics*, 53:137–151, 2011.

A.M. Alonso, G. Bastos, and C. García-Martos. Electricity price forecasting by averaging dynamic factor models. *Energies*, 9:1–21, 2016.

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

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

J. Bai and S. Ng. Large dimensional factor analysis. *Foundations and Trends in Econometrics*, 3:89–163, 2008.

A. Bello, J. Reneses, and A. Muñoz. Medium-term probabilistic forecasting of extremely low prices in electricity markets: Application to the Spanish case. *Energies*, 9:193–219, 2016.

M. Billio, R. Casarin, F. Ravazzolo, and H.K. Van Dijk. Combining predictive densities using nonlinear filtering with applications to US economics data. *Tinbergen Institute Discussion Paper*, 2011. Available at: `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1967435`.

H.C. Bjørnland, K. Gerdrup, A.S. Jore, C. Smith, and L.A. Thorsrud. Does forecast combination improve Norges Bank inflation forecasts? *CAMAR Working Paper Series*, 2:1–32, 2010.

J. Boivin and S. Ng. Are more data always better for factor analysis? *Journal of Econometrics*, 132:169–194, 2006.

S. Bordignon, D.W. Bunn, F. Lisi, and F. Nan. Combining day-ahead forecasts for British electricity prices. *Energy Economics*, 35:88–103, 2013.

G. Caggiano, G. Kapetanios, and V. Labhard. Are more data always better for factor analysis? results for the Euro Area, the six largest Euro Area countries and the UK. *Journal of Forecasting*, 30:736–752, 2011.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51:1281–1304, 1983.

H. Chipman, E.I. George, and R.E. McCulloch. The practical implementation of Bayesian model selection. *IMS Lecture Notes*, 38:70–134, 2001.

R.T. Clemen. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5:559–583, 1989.

M.P. Clements and J.H. Kim. Bootstrap prediction intervals for autoregressive time series. *Computational Statistics & Data Analysis*, 51:3580 – 3594, 2007.

M.P. Clements and N. Taylor. Bootstrapping prediction intervals for autoregressive models. *International Journal of Forecasting*, 17:247 – 267, 2001.

A.J. Conejo, J. Contreras, R. Espínola, and M.A. Plazas. Forecasting electricity prices for a day-ahead poll-based electric energy market. *International Journal of Forecasting*, 21:435–462, 2005.

A.J. Conejo, M. Carrión, and J.M. Morales. *Decision Making Under Uncertainty in Electricity Markets.*, chapter 1 and 5. Springer, New York, US, 2010a.

A.J. Conejo, J. Nogales, M. Carrión, and J.M. Morales. Electricity pool prices: long-term uncertainty characterization for futures-market trading and risk management. *Journal of the Operational Research Society*, 61: 235–245, 2010b.

V. Di Cosmo. Forward price, renewables and the electricity price: The case of Italy. *The Economic and Social Research Institute Working Paper*, 511:1–24, 2015. Available online: `https://www.esri.ie/publications/forward-price-renewables-and-the-electricity-price-the-case-of-italy/`.

K.J.M. Cremers. Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies*, 15:1223–1249, 2002.

A.C. Davison and D.R. Hinkley. *Bootstrap Methods and their Application.* Cambridge University Press, 1997.

C. Doz, D. Giannone, and L. Reichlin. A quasi-maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics*, 94:1014 – 1024, 2012.

E. Egrioglu, C.H. Aladag, and C. Kadilar. The CSS and the two-staged methods for parameter estimation in SARFIMA models. *Journal of Probability and Statistics*, 2011:1–11, 2011.

G. Elliott, C.W.J. Granger, and A. Timmermann. *Handbook of Economic Forecasting vol.1*, Oxford: North-Holland edition, 2006.

G. Fiorentini, A. Galesi, and E. Sentana. A spectral EM algorithm for dynamic factor models. *Documentos de Trabajo, Banco de España*, 1619: 1–62, 2016.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. Reference cycles : the NBER methodology revisited. *NBER*, 1999.

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

D. Fresoli, E. Ruiz, and L. Pascual. Bootstrap multi-step forecasts of non-Gaussian VAR models. *International Journal of Forecasting*, 31:834—-848, 2015.

A. García-Ferrer, E. González-Prieto, and D. Peña. Exploring ICA for time series decomposition. *Working Paper 11-16. Statistics and Econometrics Series 11. Departamento de Estadística Universidad Carlos III de Madrid. May 2011*, 2011.

A. García-Ferrer, E. González-Prieto, and D. Peña. A conditionally heteroskedastic independent factor model with an application to financial stock returns. *International Journal of Forecasting*, 28:70–93, 2012.

C. García-Martos, J. Rodríguez, and M.J. Sánchez. Mixed models for short-run forecasting of electricity prices: Application for the Spanish market. *IEEE Transactions on Power Systems*, 22:544–552, 2007.

C. García-Martos, J. Rodríguez, and M.J. Sánchez. Forecasting electricity prices by extracting dynamic common factors: application to the Iberian market. *IET Generation, Transmission & Distribution*, 6:11–20, 2012.

C. García-Martos, J. Rodríguez, and M.J. Sánchez. Modelling and forecasting fossil fuels, CO2 and electricity prices and their volatilities. *Applied Energy*, 101:363–375, 2013.

C. García-Martos, E. Caro, and M.J. Sánchez. Electricity price forecasting accounting for renewable energies: optimal combined forecasts. *Journal of the Operational Research Society*, 66:871–884, 2015.

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

J.F. Geweke and K.J. Singleton. Maximum likelihood "confirmatory" factor analysis of economic time series. *International Economic Review*, 22:37–54, 1981.

A.W. Gregory and A.C. Head. Common and country-specific fluctuations in productivity, investment, and the current account. *Journal of Monetary Economics*, pages 423–451, 1999.

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

S. Hörmann, L. Kidziński, and M. Hallin. Dynamic functional principal components. *Journal of the Royal Statistical Society*, 77:319–348, 2015.

R. Huisman, C. Huurman, and R. Mahieu. Hourly electricity prices in day-ahead markets. *Energy Economics*, 29:240–248, 2007.

C.M. Hurvich and C.L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76:297–307, 1989.

R.J. Hyndman and A. Koehler. Another look at measures of accuracy. *International Journal of Forecasting*, 22:679–688, 2006.

E.B. Iversen, J.M. Morales, J.K. Møller, and H. Madsen. Probabilistic forecasts of solar irradiance by stochastic differential equations. *Environmetrics*, 25:152–164, 2014.

E.B. Iversen, J.M. Morales, J. K. Møller, P.J. Trombe, and H. Madsen. Leveraging stochastic differential equations for probabilistic forecasting of wind power using a dynamic power curve. *Wind Energy*, Forthcoming: NA, 2016.

S.H. Jeong and B. Bienkiewicz. Application of autoregressive modeling in proper orthogonal decomposition of building wind pressure. *Journal of Wind Engineering and Industrial Aerodynamics*, 69:685–695, 1997.

L. Kilian. Small-sample confidence intervals for impulse response functions. *The Review of Economics and Statistics*, 80:218–230, 1998a.

L. Kilian. Accounting for lag-order uncertainty in autoregressions: the endogenous lag order bootstrap algorithm. *Journal of Time Series Analysis*, 19:528–531, 1998b.

C.J. Kim and C.R. Nelson. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime-switching. *Review of Economics and Statistics*, 80:188–201, 1998.

G. Koop and S. Potter. Forecasting in large macroeconomic panels using Bayesian model averaging. *Federal Reserve Bank of New York, Staff Report*, 163, 2003.

V. Kuzin, M. Marcellino, and C. Schumacher. Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics*, 28:32–411, 2012.

E.E. Leamer. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: John Wiley & Sons., 1978.

R.D. Lee and L.R. Carter. Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87:659–671, 1992.

K. Maciejowska and R. Weron. Forecasting of daily electricity prices with factor models: utilizing intra-day and inter-zone relationships. *Comput Stat*, 30:805–819, 2015.

J.G. MacKinnon and Jr.A. Smith. Approximate bias correction in econometrics. *Journal of Econometrics*, 85:205–230, 1998.

M. Marcellino, J.H. Stock, and M.W. Watson. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135:499–526, 2006.

K.V. Mardia, T. Kent, and M. Bibby. *Multivariate Analysis*, chapter 8 and 9. Academic Press, Inc. San Diego, 1979.

F. Martínez-Álvarez, A. Troncoso, G. Asencio-Cortés, and J.C. Riquelme. A survey on data mining techniques applied to electricity-related time series forecasting. *Energies*, 8:13162–13193, 2015.

C. Monteiro, L.A. Fernández-Jiménez, and I.J. Ramírez-Rosado. Explanatory information analysis for day-ahead price forecasting in the Iberian electricity market. *Energies*, 8:10464–10486, 2015.

D.C. Montgomery. *Design and Analysis of Experiments*. New York: Wiley, 1984.

J.M. Morales, M. Zugno, S. Pineda, and P. Pinson. Electricity market clearing with improved scheduling of stochastic production. *European Journal of Operational Research*, 235:765–774, 2014.

J. Nowotarskia, E. Raviv, S. Trückc, and R. Werona. An empirical comparison of alternate schemes for combining electricity spot price forecasts. *Energy Economics*, 46:395–412, 2014.

F.J. Nogales, J. Contreras, A.J. Conejo, and R. Espínola. Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, 17:342–348, 2002.

A. Onatski. A formal statistical test for the number of factors in the approximate factor models. *Economics Department, Columbia University, unpublished*, pages 1–32, 2006.

C. Otrok and C.H. Whiteman. Bayesian leading indicators: Measuring and predicting economic conditions in Iowa. *International Economic Review*, 39:997–1014, 1998.

A. Panagiotelis and M. Smith. Bayesian density forecasting of intraday electricity prices using multivariate skewed t distributions. *International Journal of Forecasting*, 24:710–727, 2008.

L. Pascual, J. Romo, and E. Ruiz. Effects of parameter estimation on prediction densities: a bootstrap approach. *International Journal of Forecasting*, 17:83–103, 2001.

L. Pascual, J. Romo, and E. Ruiz. Bootstrap predictive inference for arima processes. *Journal of Time Series Analysis*, 25:449–465, 2004.

D. Peña. Dimension reduction in time series and the dynamic factor model. *Biometrika*, 96:494–496, 2009.

D. Peña and G.E.P. Box. Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82:836–843, 1987.

D. Peña and P. Poncela. Forecasting with nonstationary dynamic factor models. *Journal of Econometrics*, 119:291–321, 2004.

D. Peña and P. Poncela. Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, 136:1237–1257, 2006.

D. Peña and F.J. Prieto. Combining random and specific directions for outlier detection and robust estimation inhigh-dimensional multivariate data. *Journal of Computational and Graphical Statistics*, 16:228–254, 2007.

D. Peña and T. Safadi. Bayesian analysis of dynamic factor models: an application to air pollution and mortality in São Paulo, Brazil. *Environmetrics*, 19:582–601, 2008.

S. Pineda, J.M. Morales, A.J. Conejo, and M. Carrión. Scenario reduction for futures market trading in electricity markets. *IEEE Transactions on Power Systems*, 24:878–888, 2009.

P. Poncela, J. Rodríguez, R. Sánchez-Mangas, and E. Senra. Forecast combination through dimension reduction techniques. *Journal of Forecasting*, 27:224–237, 2011.

P. Poncela, E. Senra, and L. Sierra. Common dynamics of nonenergy commodity prices and their relation to uncertainty. *Applied Economics*, 46: 3724–3735, 2014.

A.E. Raftery. Bayesian model selection in social research. *Sociological Methodology*, 25:111–163, 1995.

A.E. Raftery, D. Madigan, and J.A. Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92:179–191, 1997.

E. Raviv, K.E. Bouwman, and D. Van Dijk. Forecasting day-ahead electricity prices: utilizing hourly prices. *Energy Economics*, 50:227—239, 2015.

P.M. Robinson. Conditional-sum-of-squares estimation of models for stationary time series with long memory. *IMS Lecture Notes–Monograph Series. Time Series and Related Topics*, 52:130–137, 2006.

A. Roy and W.A. Fuller. Estimation for autoregressive time series with a root near 1. *Journal of Business & Economic Statistics*, 19:482–493, 2001.

E. Ruiz and P. Poncela. *Small versus big-data factor extraction in dynamic factor models: An empirical assesment, in Advance Econometrics*, chapter 35. Emerald Publishing Group, 2015a.

E. Ruiz and P. Poncela. *More is not always better: Kalman filtering in Dynamic Factor Models, in Unobserved Components and Time Series Econometrics*, chapter 11. Oxford University Press, 2015b.

I. Sánchez. Short-term prediction of wind energy production. *International Journal of Forecasting*, 22:43–56, 2006.

T.J. Sargent and C.A. Sims. Business cycle modeling without pretending to have too much a priory economic theory. *Working paper 55. Federal Reserve Bank of Minneapolis*, 1977.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.

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

J.H. Stock and M.W. Watson. Combination forecasts of output growth in a seven country dataset. *Journal of Forecasting*, 23:405–430, 2004.

J.H. Stock and M.W. Watson. Forecasting with many predictors. *Handbook of Economic Forecasting*, 1:516–542, 2006.

J.H. Stock and M.W. Watson. *Dynamic Factor Models*. Prepared for the Oxford Handbook of Economic Forecasting, edition by Michael P. Clements and David F. Hendry, Oxford University Press edition, 2010.

N.R. Swanson and T. Zeng. Choosing among competing econometric forecasts: Regression-based forecast combination using model selection. *Journal of Forecasting*, 20:425–440, 2001.

R. Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30:1030–1081, 2014.

J.H. Wright. Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics*, 146:329–341, 2008.

H. Xu, C. Caramanis, and S. Sanghavi. Robust PCA via outlier pursuit. *IEEE Transactions on Information Theory*, 58:3047–3064, 2012.

Y. Yang. Combining forecasting procedures: some theoretical results. *Econometric Theory*, 20:176–222, 2004.