Topics in Forecasting Macroeconomic Time Series

Zhang, Z.

2017

document version
Publisher's PDF, also known as Version of record

Topics in Forecasting Macroeconomic Time Series

ISBN 978 90 3160 477 7

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **687** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

VRIJE UNIVERSITEIT

Topics in Forecasting Macroeconomic Time Series

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Economische Wetenschappen en Bedrijfskunde
op maandag 23 januari 2017 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door
Zhaokun Zhang
geboren te Shanghai, China

promotor: prof.dr. S.J. Koopman
copromotor: dr. F. Blasques

Acknowledgements

The latest three years have been intellectually challenging and great fun. Completing the PhD and writing this thesis was an amazing journey that would not have been possible without the support and encouragement of many outstanding people. First and foremost I would like to express my special appreciation and thanks to my supervisors, Prof. Siem Jan Koopman and Francisco Blasques. Their patience in guiding me through all phases during my PhD greatly encouraged me to finish this thesis. Especially during the very difficult times of my PhD, they gave me enormous advice, patience, and trust. Without their guidance, support, and encouragement, my three-year PhD would not have been so enjoyable.

I would also like to thank my colleagues at the Department of Econometrics, VU University Amsterdam to provide a stimulating research environment and also a very pleasant working environment. My gratitude extends to colleagues and friends . It is really a wonderful experience to know you all and work together with you during the past five years in Tinbergen Institute,especially Zhiling Wang, Shawny Xiao and Yang Liu. Last but not least,I would like to thank Xuehan Zhang,Yue Liu and Te Li with whom I spent a lot of weekends together. Thanks to Hao Fang for your willingness to be my paranimfen during the defense.

A special thank to my parents,my wife and my parents-in-law for giving me a lot of support and encouragement.

Contents

Chapter 1

Introduction

Forecasting macroeconomic time series is a fundamental part of time series econometrics. New forecasting models and methodologies are proposed every year. However, since the mid of the 1980s it has become much more difficult for a forecaster to provide value added beyond a univariate model (Stock and Watson 2007). The recent financial crisis of 2008 also sets new challenges to econometricians as certain dynamic features of the data seem to be different from past recessions. Thanks to computational advances, it is becoming increasingly easy for econometricians to develop sophisticated forecasting models and techniques. Parameter instability, stochastic volatility, and time-varying trends are frequently considered in forecasting macroeconomic time series. Models with time-varying parameters are increasingly used in empirical studies. The use of complex time-varying models allows researchers to capture the features of the business cycle and changing dynamics of the economy.

In this thesis, we firstly compare the performance of simple linear models against complex non-linear models in terms of forecasting accuracy. In particular, we investigate the predictive ability of the Phillips curve model against other complex time-varying models for quarterly U.S. inflation. A large number of time-varying models have been introduced for inflation forecasting. For instance, Koop and Potter (2007) provided a time-varying parameter model with a changing point, Groen et al. (2013) developed a Bayesian model that averages across different Phillips curve models with time-varying parameters, Harvey (2011) introduced an unobserved random walk component into the Phillips curve model and Stock and Watson (2007) provided an unobserved component model with stochastic volatility.

1. Introduction

In the 1990s, the Phillips curve model was a very popular model that researchers discussed. The same model is also regarded as a good benchmark for inflation forecasting. In the very first part of this thesis, we study the predictive ability of the Phillips curve model against its extensions with time-varying parameters. The results from this empirical study suggest that there is no strong evidence supporting the complicated extensions of the Phillips curve model. At least, the more complex model does not seem to systematically provide more accurate forecasts than the classic one. These findins are in line with existing literature. For example, Marcellino (2008) also considers a large variety of forecasting models, including simple linear models and complicated models with time-varying parameters, for the Gross Domestic Product (GDP) growth rate and inflation rate. The results of Marcellino (2008) suggest that simple linear models are not always dominated by complicated models in terms of forecasting accuracy. In effect, Marcellino (2008) finds that simple linear models are able to provide forecasts with great accuracy when they are carefully specified. Ferrara et al. (2013) draw a similar conclusion. In this recent paper, the authors compare the forecasting abilities of some popular univariate non-linear models to a linear autoregressive model. They focused on the Great Recession period and considered a large group of macroeconomic variables as the target of their forecasting models. The forecasting experiment provided no evidence that the non-linear models can systematically provide more accurate forecasts than the benchmark. Indeed, the results were rather mixed and depended strongly on the forecasting target and evaluation period.

A general conclusion that might be drawn is this that the forecasting ability of a model may depend greatly on whether the model is correctly specified during the forecasting period. If the simple linear model is approximately well specified during the forecast period, then it forecasts well. If it is not, then the complex models become useful and provide more accurate forecasts. Motivated by this argument, this thesis proposes a weighted maximum likelihood estimation (WMLE) method which improves the forecasting accuracy of simple linear models, by acknowledging that these models may be misspecified. The WMLE method is a simple method which can lead to significant gains in forecasting accuracy. The WMLE estimator can be defined on simple linear models, but it deliver parameter estimates that can capture the complexities of time-varying parameter models.

Our WMLE method gives different weights to the sample observations and obtains parameter estimates that are optimal for forecasting. For example, the WMLE can be used to give higher weight to more recent observations compared to observations far in the past. This may be desirable as various political, institutional and technological changes may render past observations increasingly obsolete for forecasting today's economic variables. The WMLE can also be used to give higher weight to historical periods of greater relevance. For example, if we are interested in forecasting the growth rate of GDP during the recent global recession, then it may be interesting to give higher weight to past observations coming episodes of economic recession. In essence, data from past recessions may be more informative about the dynamics of GDP during the current global recession

periods. In general, at any given point in time, the WMLE weights are automatically adjusted to reflect the importance of past observations for improving forecast accuracy. The WMLE is designed to find and emphasize the weight of the most relevant data for a given forecasting period.

Many other examples can be considered. In the context of inflation, it is only natural to suppose that forecasts for inflation during an international oil crisis can be improved by paying special attention to price dynamics during past oil crises. Just as forecasts for unemployment during a period of strong fiscal austerity can benefit from emphasizing observations that lie in past episodes of strong fiscal austerity.

The idea of weighted maximum likelihood estimation can also be applied to multivariate time-series models. For the purpose of forecasting key macroeconomic or financial variables from a set of time series variables, we consider a WMLE estimator that splits the log likelihood function into two sets of variables. The first set is composed of key variables that we are interested in the second part is associated with the related variables which may contribute to the forecasting of key variables. If the multivariate model is estimated by MLE, then all variables are deemed to be 'equally important' from an estimation perspective. In other words, the MLE tries to fit all variables equally well. When the mutivariate model is correctly specified, this approach is appropriate as there exists a unique parameter that delivers the best fit for all variables. This unique parameter will also deliver the best forecasting accuracy for all variables. However, if the model is misspecified, then this approach is not optimal. Our proposed weighted maximum likelihood estimator gives more weight to the likelihood contribution from the key variables. As such the WMLE recognizes that we are more interested in fitting well the first set of variables than the second set of variables. In fact, the WMLE recognizes that we are interested in fitting the second set of variables only to the extent that this might help us forecast the first set of variables. As an example, suppose that we are interested in forecasting the quarterly growth rate of GDP, and that we have at our disposal a set of variables that may help us with this task. The WMLE will automatically find the weights accross variables that deliver the best forecasting accuracy for the variable of interest.

From the point of view of the applications, this thesis focuses on the forecasting of the U.S. inflation rate and the U.S. GDP growth rate. Forecasting inflation is one of the most important and difficult exercises in macroeconomic time series analysis. Inflation forecasting usually plays an important role in a monetary decision process. A large amount of literature focuses on this topic. The most representative and influential papers include King and Watson (1994), Stock and Watson (1999), Atkeson and Ohanian. (2001), Stock and Watson (2007), Ang et al. (2007) and Harvey (2011). In the meantime, the forecasting of the GDP growth rate is also of key importance for economic policy makers. Reliable forecasts are especially in high demand when the economic environment is uncertain as we have witnessed in the years during and after the great crisis. Many different model-based approaches exist for this purpose, ranging from basic time series models to sophisticated structural dynamic macroeconomic models. For example, Stock and Wat-

son (2002b) and Giannone et al. (2008) for the U.S.,Marcellino et al. (2003) and Rünstler et al. (2009) for the euro area, and Schumacher and Breitung (2008) for Germany. Other empirical studies of important macroeconomic time series are also included in this thesis such as Industrial Production Index (IPI) and the unemployment rate. The industrial production index is important for economic policy makers as it is a high frequent indicator which is regarded as a key reference for manufacturing activities and indicates the dynamic of the business cycle. The unemployment rate provides a straightforward measure of the state of the labor market and reflects the aggregate economic activity.

Outline of this thesis

The outline of this thesis is as follows. In the second chapter, we examine the predicting ability of the Phillips curve model against other forecasting models for quarterly U.S. inflation. In particular, we consider the Phillips curve model and its extension with unobserved component and score-driven volatility. We compare the predicting ability with a set of univariate time series models, such as unobserved component models and autoregressive models. We consider the consumer price indexes (CPI) and the personal consumption expenditure (PCE) index as the inflation measures. The Diebold and Mariano (1995a) test and the Hansen et al. (2011) model confidence set test are introduced in the empirical study to show the significance of the forecasting performance. We consider several forecast horizons and a rolling window evaluation.

In the third chapter, we derive a novel weighted maximum likelihood estimation (WMLE) that provides optimal forecasting accuracy for linear autoregressive models. The WMLE reduces to the classic MLE when the weights are uniformly set to one. We show how the optimal weights can be estimated by a cross-validation technique. A Monte Carlo study reveals that the WMLE can significantly improve the forecasting accuracy of autoregressive models. The WMLE method is investigated for the forecasting of U.S. industrial production index (IPI) and other key macroeconomic time series during the recent global recession.

In the fourth chapter, we adopt the weighted maximum likelihood estimator (ML) method to the class of dynamic factor models that is specified by means of a multivariate model with a low frequency quarterly time index. This weighted ML estimator introduces variable-specific weights in the likelihood function to let some variable equations be of more importance during the estimation process. We derive asymptotic properties, including consistency and asymptotic normality, of the weighted maximum likelihood estimator. We show that this estimator outperforms the standard likelihood-based estimator in approximating the true unknown distribution of the data as well as in out-of-sample forecasting accuracy. We verify the new estimation method in a Monte Carlo study and investigate the role of different weights in different settings. In the context of forecasting gross domestic product growth, this key variable is typically observed at a low

(quarterly) frequency while the supporting variables are observed at a high (monthly) frequency. We adopt a low frequency representation of the mixed frequency dynamic factor model and discuss the computational efficiencies of this approach. We present improvements in nowcasting and forecasting accuracy when the weighted likelihood-based estimation procedure is adopted.

Chapter 2

Forecasting U.S. Inflation Using Phillips Curve Model

2.1 Introduction

The availability of accurate inflation forecasts is of major importance to policy makers as future or expected inflation typically provides the basis to advocate and discuss appropriate changes in economic and monetary policies. In the 1990s, the Phillips curve has provided the key mechanism for inflation forecasting; see, for example, the discussions in King and Watson (1994) and Fuhrer (1995). The Phillips curve model is a linear model that relates current inflation with its past observations and aggregated economic activity indices such as the unemployment rate. The seminal work of Stock and Watson (1999) has provided the leading study in investigating the Phillips curve model for the purpose of inflation forecasting. This early research examined the predictive accuracy of the Phillips curve model using various macroeconomic indexes, such as interest rate, and commodity prices. They concluded that the Phillips curve model generally can provide accurate forecasts for U.S. inflation.

Although the Phillips curve was widely used to produce inflation forecasts in the 1990s', many researchers have found evidence that the Phillips curve model is not necessarily an accurate forecasting model. McCracken and Clark (2006) argued that the poor forecasting ability is due to two reasons: (i) the out-of-sample metrics have relatively low power, and (ii) the coefficients in the model are not stable over time. With the development of advanced time series methods, researchers have become increasingly interested in investigating forecasting models with time-varying parameters. These com-

plicated models are expected to provide accurate inflation forecasts. For instance, Koop and Potter (2007) provided a time-varying parameter model with a changing point, and Groen et al. (2013) developed a Bayesian model that averages across different Phillips curve models with time-varying parameters. However, for the purpose of our study, the most interesting contribution is presented in Harvey (2011) where an unobserved random walk coefficient for the output gap variable is introduced. It is shown that this flexible Phillips curve model fits the data well and accurately captures the salient features of the core Consumer Price Index (CPI) inflation measure.

The unobserved component (UC) model in the context of macroeconomic forecasting is widely considered in the forecasting literature. Stock and Watson (2007) provides an unobserved component model with stochastic volatility (UC-SV) and presents evidence that the UC-SV model describes the main feature of the past inflation pattern and provides good forecasts when compared to other univariate models during the great moderation. In addition, Creal et al. (2008) argue that the parameters of the UC model can also vary according to a score-driven process over time. We will show that the unobserved component model with score-driven volatility (UC-SDV) provides a similar in-sample fit as the UC-SV model. The score driven volatility is based on the generalized autoregressive score (GAS) process where the scaled score function of the observation density is used as an effective driving mechanism of the time-varying parameters; see Creal et al. (2013) for a detailed discussion. The difference between UC-SV model and UC-SDV model is that the volatility process of the former one is parameter-driven while the latter is observation-driven. The forecasting ability of these two kinds of models is discussed in detail in Koopman et al. (2015). In our study we compare the forecasting ability of UC-SV model and UC-SDV model using U.S. inflation data.

We conduct a thorough analysis of the forecasting performance of univariate time series models and Phillips curve models. We consider a large variety of models, including autoregressive (AR) models, unobserved component models and their extensions with score-driven volatility. In these comparisons, we also include the Phillips curve model together with its extensions with an unobserved component, a score-driven volatility, or both. A range of real economic activity indicators is used as predictors in the Phillips curve models. A subset of these indicators are not only individually considered to be included in the Phillips curve model but we also consider the inclusion of a set of principal components from these economic activity indicators. We consider two indexes of consumer price index (CPI) and the personal consumption expenditure index (PCE) as our measures of U.S. inflation. The model comparison exercise is conducted using rolling-window forecasts. We focus on two out-of-sample periods using quarterly data for US inflation, the great moderation period (1985Q1-2007Q4) and the period after the recent financial crisis (2008Q1-2015Q4). We take the root mean squared forecasts error (RMSE) as the criterion in our comparative analysis. Several tests are conducted to show the significance of the forecasting performances including the Diebold and Mariano (1995a) test and the model confidence set (MCS) method of Hansen et al. (2011).

The purpose of this chapter is to show that whether the Phillips curve model provides a good benchmark for U.S. inflation forecasting. We aim to investigate whether a more sophisticated extension of the Phillips curve model, such as adding an unobserved component or score-driven volatility, can provide more accurate forecasts for U.S. inflation. Our major empirical results can be summarised as follows. Firstly, we have learned that the Phillips curve can still provide accurate forecasts for U.S. inflation. The various extensions of the Phillips curve cannot always outperform the basic Phillips curve model as far as forecasting accuracy is concerned. Secondly, the UC Phillips curve model provides the most accurate forecasts overall, among all considered competing models, when forecasting PCE inflation. Thirdly, univariate time series models appear to outperform the Phillips curve models in forecasting core-CPI inflation. The results are mixed but this also illustrates the many challenges of forecasting inflation. Finally, in comparing the forecasting results of UC-SV model and UC-SDV model, we have learned that the forecasting accuracy of these two models is highly similar. This result is consistent with the finding of Koopman et al. (2015). It suggests that GAS models provide a competitive alternative to parameter driven models from a forecasting perspective. Given that inference procedures for UC-SV models are somewhat involved as they rely on simulation-based procedures while those for GAS models are straightforward and fast, this finding is of key importance to the practitioners responsible for inflation forecasting.

Another key finding of our study is that the classical Phillips curve model is not always inferior to its basic version with the time-varying parameter extensions. This finding is consistent with the discussions presented in Marcellino (2008) and where it is concluded that a more involved time series model (for example, a model with time-varying parameters) does not necessarily provide more accurate forecasts than a simple time series model. There is no strong evidence that more involved forecasting models are superior to simple time series models when the interest is mainly directed towards forecasting accuracy.

The outline of the chapter is as follows, In Section 2 we present the time series models and the Phillips curves models which we adopt in our empirical study. This section also reviews the methodology of parameter estimation for these models and the computation of the forecasts. In Section 3, we present and explore the results of our empirical study for U.S. inflation forecasting. Finally, the conclusions are presented in Section 4.

2.2 Forecasting models and methodology

2.2.1 Forecasting models

As mentioned in the introduction, the forecasting exercise is based on evaluating the performance of several univariate models and Phillips curve model together with its extensions. In all the out-of-sample forecasting exercises we consider the quarterly inflation

rate, which is formally defined as $\pi_t = 100 \ln(P_t/P_{t-1})$, where P_t is the quarterly price index. We begin with presenting a number of univariate forecasting models, including autoregressive (AR) model and unobserved component (UC) model. Then we discuss the next group of models, Phillips curve model with its extensions of unobserved component or score drive variances. Each model discussed below contains at least one of the four components we considered. The four different components we considered in this chapter are: (i) autoregressive component, (ii) unobserved component, (iii) exogenous predictor component and (iv) score-driven variance. Notice that the UC Phillips curve model with score driven variance contains all the considered components. Table (2.1) presents a full list of the models we considered.

Abbreviation	Specification	*(i)*	*(ii)*	*(iii)*	*(iv)*
	Time series models				
AR(AIC)	Autoregressive model	√			
AR-SDV	AR model with score-driven variances	√			√
UC	Unobserved component(UC) model		√		
UC-SV	Unobserved component(UC) model with stochastic volatility		√		
UC-SDV	Unobserved component(UC) model with score-driven variances		√		√
	Phillips curve models				
PC	Classic Phillips curve model	√		√	
PC-SDV	Phillips curve with score-driven variances	√		√	√
UC-PC	Phillips curve with unobserved component	√	√	√	
UC-PC-SDV	Phillips curve with unobserved component and score-driven variances	√	√	√	√

Table 2.1: List of considered forecasting models. We present the labels and definitions of the considered forecasting models for U.S. quarterly inflation. The last four columns indicate whether components (i),(ii),(iii) or(iv) are included in each forecasting model.

Autoregressive(AR) model. The Linear autoregressive (AR) model may provide good forecasting precision and be able to outperform larger and more complex models. AR models have been the workhorse of time-series econometrics for several decades. Such model allows econometricians to describe relatively complex dynamics in a remarkably simple way. The AR model is a univariate autoregression relating current variables with its past observations with k lags. Researchers usually use Akaike information criterion

(AIC) to determine the optimal lag length, k, for the AR model. The AIC deals with the trade-off between the goodness-of-fit and the complexity of the model and provides a way of model selection. For this kind of model, we adopt a direct multistep AR (AIC) model, proposed by Stock and Watson (1999), as our benchmark. In the AR (AIC) model, the forecast is directly calculated using a univariate regression according to k past observations. By using the direct multistep model, the benchmark forecasts are nested within the Phillips curve model discussed below. The optimal lag length k is selected using AIC with a rolling-window estimation. Then the h-step ahead forecasts are computed in a direct way using the following model:

$$y_{t+h} = \mu^h + \sum_{j=1}^{k} \phi_j^h y_{t+1-j} + v_t^h, \quad v_t^h \sim N(0, \sigma_v^{h\,2}) \tag{2.1}$$

where μ^h is a constant and ϕ_j^hs are coefficients that are fixed, v_t^h is the h-step ahead error term and h represents the length of forecast horizon. The AR(AIC) model is estimated by the maximum likelihood (ML) method and the forecasts are computed directly using Equation (2.1).

AR model with score-driven volatility (AR-SDV). In the AR (AIC) model, the h-step ahead error term is usually assumed to be serial independent. However, there is a large volume of research stated that the variance of the current error could be related to the previous. Besides the generalized autoregressive conditional heteroskedasticity (GARCH) model of Bollerslev (1986), many researchers also assume that the logarithm of the standard variance of the current error equals that of the previous plus a white noise. Such model is called AR-EGARCH model. However, following the study of Creal et al. (2008), the time-varying variance of the error term can also be estimated under a score-driven framework that is also called a generalize autoregressive score (GAS) model. The difference between the AR-SDV model and the ordinary AR-EGARCH model is that the variances generated under GAS framework vary more moderately. Moreover, Koopman et al. (2015) further established that GAS models often lead to forecasting gains over GARCH model. In the score driven framework, the score-driven variance is updated by an autoregressive updating equation including a scaled score of the log-likelihood, which is specified as:

$$\ln \sigma_{v,t}^h = w + a \ln \sigma_{v,t-1}^h + b s_{t-1}^h, \tag{2.2}$$

where $\sigma_{v,t}^h$ is the standard variance of h-step ahead error term v_t^h, w is a constant and a and b are coefficients that are fixed. Furthermore, s_t^h is the scaled score of the log-likelihood of y_t with respect to $\ln \sigma_{v,t}^h$. We further assume that the h-step ahead error terms v_t^hs are t-distributed with a zero mean. Combining Equation (2.1) and Equation (2.2) gives the AR-SDV model. Before calculating the forecast for time $t = T + h$ made at time T, all the parameters in the model for each horizon h are estimated by the

maximum-likelihood estimation (MLE) method using the data available up to time T, then the forecast is computed using the direct model Equation (2.1).

Unobserved component (UC) model. Unobserved component (UC) model is also widely used to forecast macroeconomic time series when there is apparent unit root in the time series of inflation (see Stock and Watson (2007), Koop and Potter (2007), Chan et al. (2013). In general, UC model can contain trend, seasons and cycles components. In this chapter, we focus on the simplest case, which is a UC model including only a trend component that follows a simple driftless random walk. In the rest of this chapter, we investigate the forecasting performances of both the simplest UC model and the UC model with time-varying variance. In the UC model, inflation is modelled by a stochastic trend τ_t and a serially uncorrelated disturbance ϵ_t:

$$y_t = \tau_t + \epsilon_t, \quad \text{where } \epsilon_t \sim N(0, \sigma^2_{\epsilon,t}), \tag{2.3}$$

$$\tau_t = \tau_{t-1} + \eta_t, \quad \text{where } \eta_t \sim N(0, \sigma^2_{\eta,t}), \tag{2.4}$$

where ϵ_t and η_t are serially and mutually uncorrelated. In the UC model, two standard variances, $\sigma_{\epsilon,t}$ and $\sigma_{\eta,t}$ are time-invariant ($\sigma_{\epsilon,t} = \sigma_\epsilon$ and $\sigma_{\eta,t} = \sigma_\eta$). The situation of time-varying variances is discussed in the rest of this section. When estimating the model, Equation (2.3)- (2.4) are applied to data up to time T to obtain filtered estimates of the trend component by Kalman filter. Then the filtered estimated mean of the trend component $\hat{\tau}_{T+1|T}$ is used as the forecasts for time $T + h$ ($\hat{y}_{t+h} = \hat{\tau}_{T+1|T}$).

Unobserved component model with stochastic volatility (UC-SV). Stock and Watson (2007) argued that there is significant parameter instability in the univariate inflation series and this could be related to time varying parameter or changing regimes in the parameter. In their paper, they presented an unobserved component model where the logarithm of the variances of observed and level disturbances evolve as independent random walks. It is called an unobserved component model with stochastic volatility (UC-SV). In the UC-SV model, two log-variances are specified by:

$$\ln \sigma^2_{\epsilon,t} = \ln \sigma^2_{\epsilon,t-1} + e_{\epsilon,t}, \tag{2.5}$$

$$\ln \sigma^2_{\eta,t} = \ln \sigma^2_{\eta,t-1} + e_{\eta,t}, \tag{2.6}$$

where $e_t = (e_{\epsilon,t}, e_{\eta,t})$ is normally distributed with mean zero and variance A and A is a two by two diagonal matrix with element $A_{1,1}$ and $A_{2,2}$ on the diagonal. Combining Equations (2.3)-(2.4) with Equations (2.5)-(2.6) gives the UC-SV model. The UC-SV model are estimated using Markov Chain Monte Carlo (MCMC) method according to Stock and Watson (2007). The sampled trend mean given data up to T, $\hat{\tau}_{T|T}$, is used as the h-step ahead forecast. ($\hat{y}_{t+h} = \hat{\tau}_{T|T}$)

Unobserved component model with score-driven volatility (UC-SDV). According to Creal et al. (2008), the parameters of the UC model can also vary according to GAS process over time, and that such model can capture the overall evolution of U.S. inflation. Furthermore, the patterns of estimated time-varying parameters are similar to those obtained from UC-SV model. We call it the UC-SDV model here. The difference between UC-SV model and UC-SDV model is that the former is a parameter driven model while the latter is an observation driven model. We include the UC-SDV model into the comparison because Koopman et al. (2015) concluded that GAS models provide a competitive alternative to the parameter driven model and they are more robust to model misspecification. On the other hand, the likelihood function for GAS model is available in closed-form, such that estimating the GAS model does not require the use of simulation method. The GAS process uses the scaled score function of the model density as an effective choice for the driving mechanism of the time-varying parameters as we do in AR-SDV model. The GAS updating function for the variances is specified as:

$$\ln \sigma^2_{\epsilon,t} = w_1 + a_1 \ln \sigma^2_{\epsilon,t-1} + b_1 s_{\epsilon,t}, \tag{2.7}$$

$$\ln \sigma^2_{\eta,t} = w_2 + a_2 \ln \sigma^2_{\eta,t-1} + b_2 s_{\eta,t}, \tag{2.8}$$

where w_1, w_2 are constants, a_1, a_2, b_1 and b_2 are coefficients that are fixed and $s_t = (s_{\epsilon,t}, s_{\eta,t})$ is the scaled score of the likelihood function of y_t with respect to $\ln \sigma^2_{\epsilon,t}$ and $\ln \sigma^2_{\eta,t}$. Combining Equation (2.3)-(2.4) with Equation (2.7)-(2.8) gives the UC-SDV model. In order to calculate the forecasts , the UC-SDV model is first applied to data up to time T to obtain filtered estimates of the trend component according to Creal et al. (2008). Then the filtered estimated mean of the trend component $\hat{\tau}_{T+1|T}$ is used as the forecasts for time $T + h$ ($\hat{y}_{t+h} = \hat{\tau}_{T+1|T}$).

Phillips curve model. We consider the Phillips curve model of Stock and Watson (1999) which directly calculate the forecasts using a multistep function. Stock and Watson (1999) also suggested that the Phillips curve model should include measures of real activity as a predictor in the model. Those economic indicators which measure the real activities are introduced in Section 4.1. A direct multistep forecasting model is considered. The lag number of k is also chosen according to Akaike information criterion (AIC). The model is specified as:

$$y_{i,t+h} = \mu^{i,h} + \sum_{j=1}^{k} \phi_j^{i,h} y_{t+1-j} + \beta^{i,h} x_{i,t-1} + w_t^{i,h}, \tag{2.9}$$

where $\mu^{i,h}$ is a constant, $\phi^{i,h}$ and $\beta^{i,h}$ are coefficients, $w^{i,h}$ is the h-step ahead error term and $x_{i,t-1}$ is the predictor variable. Similar to the AR-SDV model, we also consider a Phillips curve model with score-driven volatility under a GAS framework by combining Equation (2.2) and Equation (2.9), which is abbreviated to PC-SDV model. For both

models, when calculating h-step ahead forecasts, the Phillips curve model is first estimated using the data up to time T. Then the forecast is made using the direct function Equation (2.9).

UC Phillips curve model. Harvey (2011) claimed that the inflation forecasting model can be modelled more effectively by including an unobserved component with predictor index rather than using the Phillips curve model and such model provides a better fit to the data. This is called the UC Phillips curve model. The dynamic properties of the unobserved component can match the pattern of the core inflation and the inflation forecasts can be written as the combination of the expectation of the core inflation, the output gap and a residual. In the original paper, the UC Phillips curve model is modelled only with GDP gap as a predictor. Here a wider range of economic indicators (see Table (2.2)) is included. The UC Phillips curve model is specified as follows:

$$y_{t+h} = \tau_t^{i,h} + \beta^{i,h} x_{i,t} + \epsilon_{i,t}, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{2.10}$$

$$\tau_t^{i,h} = \tau_{t-1}^{i,h} + \eta_{i,t}, \quad \eta_t \sim N(0, \sigma_\eta^2) \tag{2.11}$$

where ϵ_t^is and η_t^is are all mutually independent for $t = 1, \ldots, n$. We also consider the UC Phillips curve model where the variances $\epsilon_{i,t}$s are time varying under GAS framework, which is combining Equations (2.10)-(2.11) with Equations (2.2). For both UC Phillips curve models, the forecast for horizon h is calculated by using the direct Equations (2.10) with filtered trend mean $\hat{\tau}_{T+1|T}^{i,h}$ given observation up to time T and predictor variable $x_{i,T}$.

2.2.2 Extracting information from Large Data Sets

Principal Component Model. We consider two ways of extracting information from large data sets. Firstly, many studies have found that using principal components as predictors in the Phillips curve model is a useful way to forecast macroeconomic series. Stock and Watson (2002c) proved that forecasts of a single time series using principal components are asymptotically efficient in an approximate factor model. Gavin and Kliesen (2008) found that the data-rich model using principal components outperform autoregressive and random walk models in forecasting real output and inflation index. Therefore, we also add a regression including the principal component of the considered predictors which are discussed in Section 4.1. By calculation, the first three principal components explain 46,92% , 25,04% and 13,45% of the total variation in the nine series respectively. Since the first component already explains almost half of the variation, it is interesting to include only the first principal component to the Phillips curve model. We also add a regression where the first three principal components are included as they explain 85,41% of the total variation. Thus, we introduce two principal component models, namely PC1 and PC2. PC1 is the principal component model, which only includes

the first principal component while PC2 is the principal component model including the first three principal components.

Forecasting Combination. Pooling the single indicator Phillips curve forecasts or forecasting combination can be considered as another way for extracting information from large data sets. Such kind of method also helps to improve macroeconomic index forecasts (e.g. Stock and Watson (2004) and Timmermann (2006)). We do not consider a complicated method of determining pooling weights. Instead, we only consider the equal weighted average (EWA) of all the single indicator forecasts. In many inflation forecasting studies, equal weighting of many forecasts has been used widely and find hard to be beaten by other kinds of averaging method. These two kinds of data-rich models can help us to compare the overall forecasting ability of different Phillips curve model. As mentioned in many articles, it is difficult to find a leading indicator which provides inflation forecasts with the best accuracy. It is difficult to comparing different Phillips curve models, by looking at the forecasting results individually. Thus, an overall measurement of different Phillips curve models is needed.

2.3 Empirical study

2.3.1 Data

In our empirical analysis, we consider three different measures of quarterly inflation as our target forecasting series. The first two measures are the consumer price indexes (CPI) including both the all-CPI for all consumers and all items, as well as the core-CPI for all consumers and all items less food and energy. The last inflation index is the personal consumption expenditure (PCE). All inflation indexes are observed at a quarterly frequency. The two CPI measures are collected from U.S. Department of Labor: Bureau of Labor Statistics and the PCE measure is collected from U.S. Department of Commerce: Bureau of Economic Analysis. Inflation rate is calculated as $\pi_t = 100 \ln(P_t/P_{t-1})$, where P_t is quarterly price level mentioned above. We consider two different inflation measures (CPI and PCE) because both two measures are important for different decision-making processes. The CPI measure is used to adjust social security payments and more importantly it is also the reference for some financial contracts and inflation swaps. On the other hand, the PCE measure is used to set the goal for inflation by the Federal Reserve.

We also consider 9 predictor variables to be used in Phillips curve model. Such predictors are economic indexes which present aggregate activity in the real economy. These 9 U.S. economic time series are observed at a monthly frequency except the series "GDP" that is observed at a quarterly frequency. These 9 economic indexes are collected from three data sources. They are U.S. Department of Commerce, U.S. Department of Labor, and Board of Governors of the Federal Reserve System. The 9 predictor variables are: the seasonally adjusted real GDP (GDP), real personal consumption expenditures on services

(RCON), the number of employees on non-agricultural payrolls (NFPR), new privately owned housing units started (HSFR), civilian unemployment rate for people above 16 (UNEMP), interest rate of 3-month US treasury bill (YFGM3), industrial production index (IPI), effective federal fund interest rate (FFR) and real disposable personal income (RDPI). Some of the data are transformed before we use them as predictors in the model in order to remove possible trends from the series. When the data are observed at a monthly level , we select the level at the last month of a quarter as the quarterly observations. The definition and transformation of each series are summarized in Table (2.2).

2.3.2 Design of Empirical Study

In our empirical study, we investigate the forecasting performance of different univariate time series models and Phillips curve models. We aim to establish whether the models based on Phillips curve are dominant by the univariate time series model such as the AR type model and UC type model in the sense of forecasting precision. We are also interested to see whether adding time-varying features, such as unobserved component and score-driven volatility, can improve the forecasting accuracy of the Phillips curve model. Moreover, it is well known that the U.S. inflation was harder to forecast during the great moderation and we see a great increase in volatility of the inflation series after 2008. We would like to investigate whether there is a difference when forecasting inflation after 2008. We focus on two out-of-sample forecasts period. The first period starts from 1985Q1 till 2008Q4 and the second one starts from 2008Q1 till 2015Q4. By doing so, we can distinguish the forecasting ability of different models before and after the recent big crisis. Forecasts are calculated based on parameters estimated over a rolling window of 25 years starting from 1960Q1.

We use the root mean squared forecasts error (RMSE) as the criterion for evaluating forecast accuracy. The ratios of RMSEs of the competing models relative to the benchmark model AR (AIC) are reported. Furthermore, in order to show the significance of the forecast accuracy of the competing model, the Diebold and Mariano (1995a) test is performed. Furthermore, the model confidence set (MCS) test by Hansen et al. (2011) is carried out among the forecasting models. Such procedure is performed in order to determine the model set that consists of the best forecasting models given a level of confidence. The p-values of such test are also presented. These p-values are analogous to the classical p-values thus, cannot be interpreted as the probability that a particular model is the best against others. It is constructed in a way such that it reflects the order of the model to be removed from the best model set. The higher the p-value is, the later it is removed from the best model set. The p-values can also easily determine whether a particular forecasting model is in the best model set of confidence level α, especially if the p-value is bigger than the confidence level then the corresponding model is in the best model set. Hansen et al. (2011) also stated the MSC procedure can also address

Variables	Transformation	Definition
Inflation		
ALL-CPI	100△ ln	Consumer Price Index: Total All Items for the United States
CORE-CPI	100△ ln	Consumer Price Index for All Urban Consumers: All Items Less Food and Energy
PCE	100△ ln	Consumer Price Index for All Urban Consumers: All Items Less Food and Energy
Predictor variables		
GDP	100△ ln	Real Gross Domestic Product
RCONS	100△ ln	Real personal consumption expenditures: Services (chain-type quantity index)
NFPR	ln	All Employees: Total Non-farm, commonly known as Total Non-farm Payroll,
SHFR	100△ ln	Housing Starts: Total: New Privately Owned Housing Units Started
UNEMP	level	Unemployment Rate: Aged 15-64: All Persons for the United States
YFGM3	△	3-Month Treasury Bill in United States: Secondary Market Rate
IPI	100△ ln	Industrial Production Index
FFR	△	3-Month Treasury Bill in United States: Secondary Market Rate
RDPI	100△ ln	Real Disposable Personal Income

Table 2.2: Database variables. We present the labels and definitions of the quarterly variables that are used in our empirical study for the U.S. inflation forecasting. All data are in quarterly frequency and download from "FRED2". Transformations are: (i) ln:X_t = ln(P_t), and (ii) 100 △ ln: $X_t = 100ln(P_t/P_{t-1})$. (iii)△: $X_t = P_t - P_{t-1}$. (iv)level: $X_t = P_t$

whether the benchmark model is outperformed by observing whether the benchmark is in the best model set.

2.3.3 Empirical Results

In this section, we investigate the forecasting performance for U.S. inflation between different univariate time series models and Phillips curve models. In Section 2.3.3.1, we compare the forecasting performance for U.S. inflation among the univariate time series models listed in Table (2.3). In Section 2.3.3.2 and Section 2.3.3.3, we study whether adding an unobserved component or score-driven volatility to the classic Phillips curve model can help to forecast U.S. inflation. In Section 2.3.3.4, we compare the forecasting accuracy of the different Phillips curve models against the univariate time series models by comparing the forecasts from the best forecasting models from each category.

2.3.3.1 Time series model

The forecasting results for our univariate time series models listed in Table (2.1) during the two evaluation periods (post 1985 and post 2008) for the three inflation measures are presented in Table (2.3). The forecast accuracy is measured by the root mean squared forecasting errors (RMSE). Entries in Table (2.3) report the relative RMSE ratios of the competing models to the benchmark. A value smaller than 1 implies that the competing model provides more accurate forecasts (in terms of a smaller RMSE) than the benchmark. The most accurate model for each forecast horizon determined by the smallest RMSE ratio is highlighted. We focus on comparing the forecasting accuracy among different univariate time series models. In the first panel of Table (2.3), it is clear that the AR models provide smaller RMSEs when forecasting the core-CPI inflation. In most cases, amongst all forecast horizons, AR-SDV model provide more accurate forecasts than the benchmark. For Example, in the post-1985 period and four-step ahead forecasting, h=4, the RMSE of AR-SDV model is 25% smaller than the benchmark. Even the simple AR(4) model can outperform the benchmark in most cases. The AR-SDV model still leads to small RMSE ratios in most cases when forecasting all-CPI and PCE inflation. However, the RMSE ratios of the unobserved component (UC) models are even smaller in some cases. The UC model is selected as the best model among the competitors at forecast horizon, $h = 1, 2, 4$, when forecasting all-CPI inflation after 2008 and is also selected to be the best model among the competitors at forecast horizon, $h = 1, 2, 8$, when forecasting PCE inflation during 1985-2007. Stock and Watson (2007) argued that the UC-SV model can outperform the UC model in terms of forecasting accuracy as there is significant parameter instability in the univariate inflation series. When we compare the forecasting performance between UC model and its extensions with stochastic volatility (UC-SV) and score-driven volatility (UC-SDV), we cannot conclude that the UC-SV model and UC-SDV model provide more accurate forecasts than the UC model. To summarize, we

may conclude from Table (2.3) that the AR models (AR(4) or AR-SDV) provide most accurate inflation forecasts for the core CPI inflation. While the UC model can provide more accurate forecasts than the AR models for all-CPI and PCE inflation.

We adopt the Diebold and Mariano (1995a) (DM) test to verify whether the forecasts obtained from the competing model are significantly more accurate compared to those obtained from the benchmark. The p-value of Diebold and Mariano (1995a) test for each model is also presented in Table (2.3). Entry smaller than 1 with a p-value smaller than 0.1 indicates that such competing model provides significantly more accurate forecasts than the benchmark in terms of RMSE. In almost all cases, we can find at least one competing model with RMSE ratio smaller than unity, but for a number of entries we have forecasts from the competing model which are significantly better than those from the benchmark, at 10% confidence level. We also conduct the Hansen et al. (2011) model confidence set(MCS) test to determine the best model set. The results for model confidence set method are reported in Table (2.4). Entry with an asterisk means the corresponding model is included in the best model set, at 10% confidence level. When forecasting core-CPI inflation, either AR(4) model or AR-SDV model is included in the best model set. When forecasting all-CPI and PCE inflation, unobserved component (UC) model is always included in the best model set. This implies that when forecasting core-CPI inflation the AR models provide more accurate forecasts and when forecasting all-CPI and PCE inflation, the UC model performs better. Such argument is consistent with the conclusions we draw from Table (2.3).

2.3.3.2 Phillips curve model

In this section, we study whether adding an unobserved component or score-driven volatility to the classic Phillips curve model can help to forecast U.S. inflation. We consider four different kinds of Phillips curve models: the classic Phillips curve model (PC), the Phillips curve model with score -driven variance (PC-SDV), the UC Phillips curve model (UC-PC) and the UC Phillips curve model with score-driven variance (UC-PC-SDV). For each version of the Phillips curve models, 9 economic time series are included in the model individually. Thus, 9 forecasting series are calculated from the 9 economic time series. Addition to these, two extra forecasting series are calculated using the principal components of these economic time series, namely PC1 and PC2. PC1 is the principal component model, which only includes the first principal component while PC2 is the principal component model including the first three principal components. The equal weighted average (EWA) of all the single indicator forecasts is also calculated.

The best predictors which provide the smallest RMSEs when the same Phillips curve model is considered are presented in Table (2.5). It is difficult to conclude which economic time series can be considered as the leading predictor to forecast inflation. We can only conclude that RCONS (real personal consumption expenditure) is very useful when forecasting all CPI and PCE inflation. Thus, it is difficult to compare the forecasting

	Post-1985 sample				Post-2008 sample			
Horizon:	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
				I: CORE-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AR4	0.730***	0.987**	1.003	0.984	0.669***	0.999	0.988	1.021
	(0.000)	(0.044)	(0.571)	(0.131)	(0.000)	(0.471)	(0.351)	(0.646)
AR-SDV	0.922*	1.306	0.750***	0.780***	0.913*	1.419	0.986	1.022
	(0.053)	(0.974)	(0.003)	(0.002)	(0.098)	(0.983)	(0.332)	(0.783)
UC	0.945**	1.214	1.135	0.882	1.046	1.403	1.370	1.366
	(0.017)	(0.972)	(0.867)	(0.182)	(0.676)	(0.998)	(0.997)	(1.000)
UC-SV	0.976	1.218	1.093	0.853	1.128	1.393	1.254	1.260
	(0.238)	(0.973)	(0.826)	(0.107)	(0.824)	(0.998)	(0.976)	(1.000)
UC-SDV	1.008	1.217	1.068	0.846*	1.172	1.401	1.225	1.252
	(0.588)	(0.964)	(0.752)	(0.089)	(1.000)	(0.996)	(0.948)	(1.000)
				II: ALL-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AR4	1.006	1.031	1.038	1.016	1.008	0.996	1.009	1.007
	(0.751)	(0.981)	(0.998)	(0.998)	(0.813)	(0.404)	(0.775)	(0.643)
AR-SDV	1.042	1.035	0.955***	0.901***	0.944**	1.030	1.005	0.913
	(0.886)	(0.969)	(0.006)	(0.010)	(0.036)	(0.755)	(0.586)	(0.234)
UC	1.038	1.001	0.961	0.924*	0.936*	0.954***	0.999	0.969*
	(0.869)	(0.521)	(0.121)	(0.062)	(0.059)	(0.003)	(0.467)	(0.096)
UC-SV	1.115	1.034	0.963	0.912**	1.152	1.254	1.278	1.753
	(0.885)	(0.767)	(0.246)	(0.028)	(0.864)	(0.865)	(0.915)	(0.877)
UC-SDV	1.109	1.101	0.948	0.911**	0.974**	1.003	1.022	1.083
	(0.928)	(0.853)	(0.106)	(0.046)	(0.026)	(0.539)	(0.776)	(0.778)
				III: PCE Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AR4	0.974	0.985**	0.989	0.994	1.009	1.004	0.998	1.093
	(0.110)	(0.012)	(0.109)	(0.134)	(0.869)	(0.975)	(0.265)	(0.880)
AR-SDV	0.983	0.993	1.087	1.109	1.002	0.979	1.019	1.049
	(0.190)	(0.365)	(0.978)	(0.999)	(0.531)	(0.201)	(1.000)	(0.786)
UC	0.897**	0.920**	0.967**	0.979	0.991	1.004	1.020	0.996
	(0.012)	(0.012)	(0.016)	(0.253)	(0.422)	(0.528)	(0.868)	(0.476)
UC-SV	0.935	0.929*	0.958	0.999	0.966	1.123	1.257	2.224
	(0.130)	(0.100)	(0.154)	(0.491)	(0.436)	(0.845)	(0.828)	(0.870)
UC-SDV	1.049	1.005	1.026	1.074	0.896	1.054	1.169	2.020
	(0.729)	(0.518)	(0.657)	(0.901)	(0.141)	(0.742)	(0.791)	(0.870)

Table 2.3: We forecast the quarterly inflation of U.S. over 1985Q1-2015Q4 and 2008Q1-2015Q4. Entries are RMSEs relative to the benchmark. Entries in brackets denoted the DM test p-values, where *,** and *** indicate rejection of the null at 10%,5%, and 1% level. The smallest RMSE ratios for each inflation measure at different forecast horizons are highlighted.

	Post-1985 sample				Post-2008 sample			
Horizon:	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
				I: CORE-CPI Index				
Benchmark	0.000	0.311*	0.000	0.000	0.038	0.932*	0.874*	1.000*
AR4	1.000*	1.000*	0.000	0.001	1.000*	1.000*	0.945*	0.734*
AR-SDV	0.001	0.002	1.000*	1.000*	0.038	0.147*	1.000*	0.591*
UC	0.000	0.012	0.000	0.086	0.038	0.147*	0.126*	0.138*
UC-SV	0.000	0.012	0.000	0.158*	0.038	0.147*	0.266*	0.223*
UC-SDV	0.000	0.012	0.000	0.158*	0.038	0.147*	0.313*	0.223*
				II: ALL-CPI Index				
Benchmark	1.000*	1.000*	0.098	0.004	0.329*	0.198*	0.965*	0.747*
AR4	0.539*	0.198*	0.025	0.004	0.329*	0.381*	0.900*	0.747*
AR-SDV	0.459*	0.458*	0.964*	1.000*	0.725*	0.381*	0.954*	1.000*
UC	0.459*	0.962*	0.964*	0.732*	1.000*	1.000*	1.000*	0.747*
UC-SV	0.459*	0.633*	0.964*	0.838*	0.364*	0.381*	0.593*	0.677*
UC-SDV	0.459*	0.548*	1.000*	0.838*	0.362*	0.381*	0.854*	0.747*
				III: PCE Index				
Benchmark	0.025	0.081	0.307*	0.742*	0.900*	0.703*	0.726*	0.948*
AR4	0.055	0.081	0.396*	0.742*	0.792*	0.109*	1.000*	0.260*
AR-SDV	0.055	0.088	0.084	0.000	0.900*	1.000*	0.060	0.260*
UC	1.000*	1.000*	0.779*	1.000*	0.900*	0.726*	0.392*	1.000*
UC-SV	0.237*	0.740*	1.000*	0.742*	0.900*	0.703*	0.726*	0.497*
UC-SDV	0.025	0.128*	0.151*	0.019	1.000*	0.679*	0.726*	0.447*

Table 2.4: We forecast the quarterly inflation of U.S. over 1985Q1-2015Q4 and over 2008Q1-2015Q4. Entries are p-value of the model confidence set test, where * indicate that the corresponding model is selected in the best model set at 10% level.

abilities among different versions of the Phillips curve model by looking at the individual forecasting results. Instead, we can compare the overall forecasting performance by investigating the results from the data-rich models (principal component model and the equal weighted average of all the single indicator forecasts). In the following analysis, we compare the forecasting ability of different versions of the Phillips curve model by evaluating the forecasting results from the data-rich model.

Table (2.6) compares the forecasting performance of data-rich models from four different versions of the Phillips curve model. Details of the forecasting performances of different Phillips curve models using individual predictors can be found in the extra tables in Section 2.5. Table (2.9), Table (2.11), Table (2.13) and Table (2.16) report the relative RMSE ratios of the competing models comparing to the benchmark and the p-value from DM test. Table (2.10) ,Table (2.12),Table (2.14) and Table (2.16) report the test results from MSC test.

Table (2.6) presents the RMSE ratios of the forecasts for different Phillips curves models with respect to the benchmark. The most accurate models with smallest RMSE ratios in each comparison are highlighted. We also consult the Diebold and Mariano

	Post-1985 sample				Post-2008 sample			
Horizon:	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
				I: CORE-CPI Index				
PC	NFPR	GDP	NFPR	NFPR	NFPR	RCONS	RCONS	YFGM3
PC-SDV	RCONS	UNEMP	HSFR	YFGM3	YFGM3	RDPI	IPI	UNEMP
UC-PC	HSFR	RDPI	UNEMP	GDP	RDPI	RDPI	NFPR	RDPI
UC-PC-SDV	HSFR	GDP	NFPR	IPI	PMI	GDP	HSFR	PMI
				II: ALL-CPI Index				
PC	IPI	RCONS	NFPR	NFPR	RDPI	NFPR	RCONS	NFPR
PC-SDV	PMI	RCONS	RDPI	RDPI	RDPI	RCONS	RCONS	NFPR
UC-PC	HSFR	RCONS	UNEMP	PMI	RDPI	RCONS	NFPR	NFPR
UC-PC-SDV	YFGM3	UNEMP	NFPR	YFGM3	YFGM3	UNEMP	NFPR	YFGM3
				III: PCE Index				
PC	RCONS	UNEMP	RCONS	NFPR	RCONS	RCONS	NFPR	NFPR
PC-SDV	RCONS	HSFR	HSFR	NFPR	RCONS	RCONS	HSFR	NFPR
UC-PC	RCONS	UNEMP	UNEMP	UNEMP	RCONS	RCONS	NFPR	NFPR
UC-PC-SDV	RCONS	IPI	PMI	GDP	RCONS	RCONS	NFPR	UNEMP

Table 2.5: This table reports the best predictor which provide the smallest RMSE when the same Phillips curve model is considered but different economic time series is used as the predictor in the model.

(1995a) test to verify whether forecast accuracy differences are significant. From the results of Table (2.6) we learn that even the classic Phillips curve model can provide accurate inflation forecasts in some cases. For example, the forecasting accuracy of the classic Phillips curve model is higher than the alternatives at one-step ahead and two-step ahead horizons, $h = 1, 2$, when the forecasting target is all-CPI inflation and in the post-1985 period. When the forecasting target is PCE inflation, the classic Phillips curve model does not provide the best forecasts with smallest RMSE. However, it still provides inflation forecasts which are significantly better than the benchmark using an equal weighted average method in most cases.

When the forecasting target is PCE inflation, the UC Phillips curve model outperforms other versions of the Phillips curves model in most cases. The UC Phillips curve is selected as the best forecasting models in the post-1985 period and the forecasts are significantly better than the benchmark. On the other hand, the PC-SDV model can provide more accurate forecasts in some cases when the forecasting target is CPI inflation. For example, its forecasts are significantly more accurate than the benchmark at forecast horizon $h = 4, 8$ for both core-CPI and all-CPI forecasts.

By comparing the forecasting accuracy of different Phillips curve models using data-rich models, we conclude that the considered Phillips curve models can outperform the benchmark in forecasting accuracy in most cases. However, there are still some cases that none of the Phillips curve models can provide more accurate forecasts with smaller

RMSE than the benchmark when forecasting core-CPI inflation.

2.3.3.3 Comparing different versions of Phillips curve model

In this chapter, we propose four different versions of Phillips curve model: the standard Phillips curve, the Phillips curve model with score-driven volatility, the UC Phillips curve model, and the UC Phillips curve model with score-driven volatility. The later three versions of Phillips curve model are adding different time-varying component to the classic Phillips curve model. In order to show how the forecasts can be improved by adding different components, Figure (2.1) reports the number of individual forecasts from later three versions of Phillips curve models which have smaller RMSEs than the corresponding individual forecasts from the classic Phillips curve model.

First, we see in most cases, the alternative Phillips curve models cannot provide more than five more accurate individual forecasts series when the forecasting target is core CPI inflation. We may conclude that none of the alternative Phillips curve models can provide better forecasts than the classic Phillips curve model. Secondly, when the forecasting target is all-CPI inflation, both the Phillips curve with score-driven volatility (black bar) and the UP Phillips curve model (grey bar) can improve the classic Phillips curve model in most cases (the number of better individual forecasts of the alternative model is larger than 5) except at forecast horizon $h = 1, 2$ in post 1985 period and at forecast horizon $h = 4$ in post 2008 period. Thirdly, when the forecasting target is PCE inflation, the UC Phillips curve model (gray bar) has the highest number of better individual forecasts beating the standard Phillips curve model in post-1985 and post-2008 period. On the other hand, the Phillips curve model with score-driven volatility(black bar) cannot improve the classic Phillips curve model because in most cases the number of better individual forecasts of this model is not larger than 5. Finally, the UC Phillips curve model with score-driven variances(white bar), which includes both unobserved component and score-driven volatility, does not have better individual forecasts than the UC Phillips curve model in most cases, which shows that the unobserved component plays an important role in forecasting PCE inflation. The results above show that adding time-varying components, such as unobserved component and score-driven volatility, does not always lead to a better forecasting model.

2.3.3.4 Summary

In this section, we compare the forecasting performance of different Phillips curve models with univariate time series model to get an overall picture of the performances of various models. we report the forecasting results of the best models from five different categories: the univariate time series model, the classic Phillips curve model (PC), the Phillips curve model with score-driven volatility (PC-SDV), the UC Phillips curve model (CU-PC) and the UC Phillips curve model with score-driven volatility (UC-PC-SDV). In Table (2.7),

		Post-1985 sample				Post-2008 sample			
Horizon:		h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
					I: CORE-CPI Index				
PC	*PC1*	1.008	1.013	0.994	0.992	1.012	1.021	1.015	1.017
	PC2	0.975	1.015	0.926*	0.885**	0.980	1.013	0.985	1.074
	EQW	0.963***	1.003	0.966**	0.932***	0.963**	0.989***	0.988**	0.934
PC-SDV	*PC1*	1.158	1.327	0.779***	0.823***	0.920	1.317	0.969	1.031
	PC2	1.185	1.321	0.799***	0.896*	0.920	1.471	1.185	0.994
	EQW	1.183	1.283	0.817***	0.752***	0.946	1.309	0.941*	0.962
UC-PC	*PC1*	0.939**	1.419	1.420	1.042	1.052	1.471	1.485	1.481
	PC2	0.940***	1.437	1.430	1.094	1.040	1.467	1.482	1.464
	EQW	0.943**	1.412	1.395	1.034	1.012	1.424	1.431	1.365
UC-PC-SDV	*PC1*	1.001	1.459	1.498	1.087	1.146	1.416	1.533	1.533
	PC2	1.095	1.504	1.458	1.039	1.171	1.619	1.586	1.524
	EQW	0.982	1.474	1.471	1.076	1.195	1.604	1.544	1.496
					II: ALL-CPI Index				
PC	*PC1*	1.004	0.993	1.002	1.004	1.028	1.029	1.006	1.063
	PC2	0.982	0.960	0.985	0.975	1.072	1.056	0.996	1.066
	EQW	0.983*	0.981**	0.974**	0.964***	1.000	1.000	0.989*	1.003
PC-SDV	*PC1*	1.153	1.142	0.948**	0.919***	0.918**	0.929**	0.977	1.059
	PC2	1.185	1.104	0.957*	0.897***	0.988	0.925**	1.000	0.940
	EQW	1.103	1.109	0.942**	0.884***	0.902*	0.929**	1.004	0.881
UC-PC	*PC1*	1.024	1.030	0.895*	0.909*	0.962**	0.960***	1.034	0.894
	PC2	1.024	1.025	0.899**	0.926	0.937***	0.961	1.035	0.894
	EQW	1.032	1.028	0.896*	0.903*	0.934**	0.941***	1.023	0.863*
UC-PC-SDV	*PC1*	1.148	1.104	0.958	0.924	0.978*	0.999	1.048	0.806**
	PC2	1.121	1.028	1.004	0.986	0.900*	0.946***	1.030	0.907
	EQW	1.100	1.079	0.962	0.929	0.957*	0.945***	1.053	0.774**
					III: PCE Index				
PC	*PC1*	1.023	0.999	1.001	0.986	1.025	1.009	1.030	1.286
	PC2	0.917**	0.989	0.981	0.991	0.926	0.965	0.995	1.189
	EQW	0.842***	0.858*	0.873**	1.053***	0.974**	0.981	0.984	1.011
PC-SDV	*PC1*	1.085	0.912**	1.044	1.073	1.109	0.950	0.988	1.084
	PC2	1.036	1.130	1.094	1.181	0.845*	0.942	1.006	1.095
	EQW	0.897	0.860	0.900	1.151	1.061	0.956	0.978	1.044
UC-PC	*PC1*	0.917*	0.917**	0.947**	0.913**	0.988	0.974	1.008	1.221
	PC2	0.815***	0.936**	0.946	0.914**	0.818*	0.895	0.969	1.187
	EQW	0.767***	0.804**	0.838**	0.990**	0.905	0.936	0.968	0.942
UC-PC-SDV	*PC1*	1.003	0.973	0.991	0.921***	0.932	0.971	1.102	1.107
	PC2	1.003	0.939**	0.978	0.958**	0.971	0.978	1.119	1.097
	EQW	0.787***	0.829	0.875	0.952**	0.849	0.935	0.917	0.994

Table 2.6: We reports out-of-sample performance of the data-rich forecasting models from different categories: the Phillips curve models (PC), the Phillips curve model with score-driven variance (PC-SDV), the UC Phillips curve model (UC-PC) and the UC Phillips curve model with score-driven variance (UC-PC-SDV). Each entry reports the RMSE ratio of the competing model to benchmark, where *,** and *** indicate rejection of the null at 10%,5%, and 1% level.

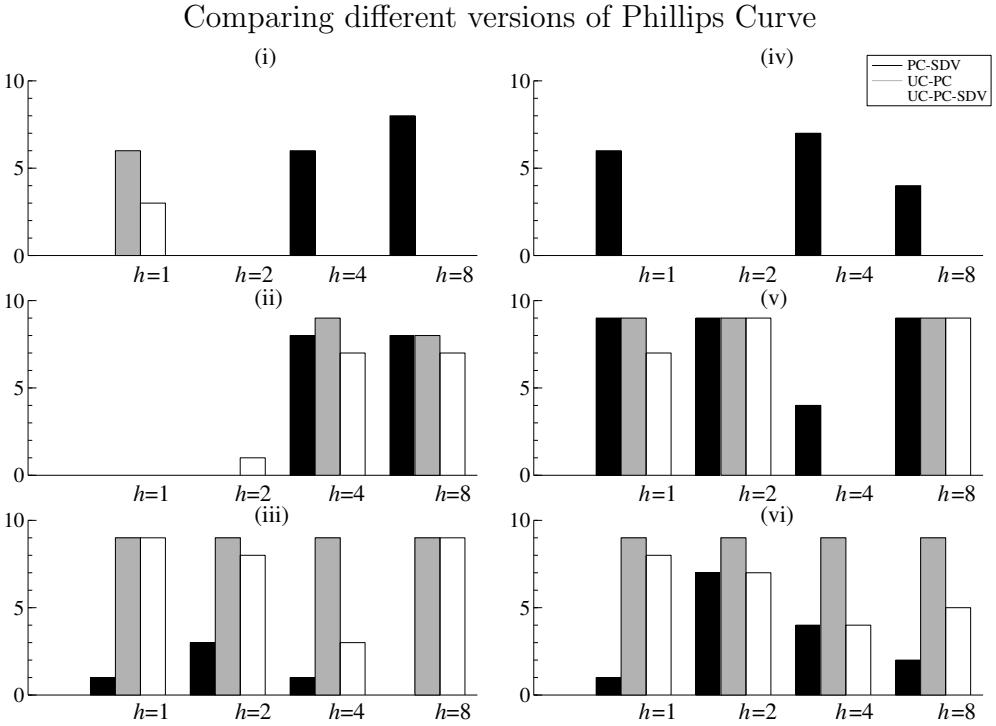

Figure 2.1: Forecasting comparison of classic Phillips curves model against its extensions. The three extensions of the Phillips curve model are: Phillips curve model with score-driven variances (PC-SDV), UC Phillips curve model (UC-PC) and the UC Phillips curve model with score-driven variances (UC-PC-SDV). We represent the number of individual forecasts which is more accurate than the correspondence of the classic Phillips curve model in terms of RMSE. The left panel shows the results for the post-1985 period using different inflation measure: (i) core-CPI inflation, (ii) al-CPI inflation, (iii) PCE inflation. The right panel is the results for the post-2008 period using different inflation measure: (iv) core-CPI inflation, (v) al-CPI inflation, (vi) PCE inflation.

the relative RMSE ratios of the best models respect to the AR (AIC) benchmark models are presented. Entries in brackets are the p-values of Diebold and Mariano (1995a) test. The most accurate model in every comparison is highlighted. Entry with RMSE ratio smaller than one and DM test p-value smaller than 0.1 indicates that such competing model is significantly better than the benchmark with respect to forecasting accuracy.

From the first panel of Table (2.7), it is difficult to conclude which model provides the most accurate forecasts for the core- CPI inflation. Different kinds of models are selected to be the best forecasting models with smallest RMSE at different horizons. However, the time series model provide consistent accurate forecasts at different horizons in big moderation (post-1985) period. During the big moderation period, the best time series model is determined as the best forecasting model amongst the alternatives at one-step ahead forecasts only, but all the RMSEs of the best time series model are significantly smaller than the benchmark at 10% confidence level. On the other hand, the Phillips curve model provides forecasts with smaller RMSE at all forecast horizons in the post-2008 period, but the forecasts are only significantly better than the benchmark at eight-step ahead forecasts.

When the forecasting target is all-CPI inflation, the univariate time series model cannot outperform the Phillips curve models by providing more accurate forecasts. Different Phillips curves models are determined to be the best forecasting model with smallest RMSE at different forecast horizons. During the big moderation period, the Phillips curve model provides more accurate forecasts than the rest competing models at forecasts horizon $h = 1, 2$. The UC Phillips curve model and the UC-PC-SDV model is selected as the best forecasting model at four-step ahead horizon and eight-step ahead horizon respectively. In the post-2008 period, the PC-SDV model outperforms the rest competing models except at forecast horizon $h = 4$. The p-values of the DM test implies that most of the best forecasting models provide significantly more accurate forecasts than the benchmark at 10% level.

For PCE inflation measure, various Phillips curve models still provide more accurate inflation forecasts than the time series model. Among the four different Phillips curve models, the UC Phillips curve model beats almost all the rest alternatives in terms of forecasting accuracy. In four out of the eight comparisons, the UC Phillips curve model provides the smallest RMSEs amongst all the competing models. In the rest four comparisons, the best forecasting model is determined to be the UC-PC-SDV model. While the differences of the RMSEs between the UC Phillips curve model and UC-PC-SDV model are tiny. This may imply that the UC Phillips curve provides forecasts as accurate as the UC-PC-SDV model does, which is also consistent with the finding of the previous section. Moreover, in all cases, the forecasts of UC Phillips curve are significantly more accurate than the benchmark at 10% confidence level. We conclude that in general the UC Phillips model is the best forecasting model in accuracy for the PCE inflation measure.

We also conduct the Hansen et al. (2011) model confidence set (MCS) test to de-

	Post-1985 sample				Post-2008 sample			
Horizon:	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
				I: CORE-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Best time series model	0.730*	0.987*	0.750*	0.780*	0.669*	0.999	0.986	1.021
	(0.000)	(0.044)	(0.003)	(0.002)	(0.000)	(0.471)	(0.332)	(0.646)
Best PC	0.933*	0.983	0.923	0.837*	0.944	0.963	0.965	0.899*
	(0.039)	(0.244)	(0.196)	(0.073)	(0.306)	(0.134)	(0.119)	(0.053)
Best PC-SDV	1.097	1.238	0.730*	0.692*	0.889*	1.293	0.931	0.968
	(0.989)	(0.951)	(0.004)	(0.002)	(0.016)	(0.983)	(0.170)	(0.350)
Best UC-PC	0.939*	1.406	1.382	1.028	1.000	1.397	1.390	1.285
	(0.000)	(1.000)	(0.999)	(0.590)	(0.500)	(1.000)	(0.999)	(0.993)
Best UC-PC-SDV	0.999	1.427	1.425	1.056	1.138	1.506	1.456	1.383
	(0.489)	(1.000)	(1.000)	(0.675)	(0.991)	(0.998)	(0.993)	(0.999)
				II: ALL-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Best time series model	1.006	1.001	0.948	0.901*	0.936*	0.954*	0.999	0.913
	(0.751)	(0.521)	(0.106)	(0.010)	(0.059)	(0.003)	(0.467)	(0.234)
Best PC	0.974	0.937*	0.964	0.894*	0.993	0.995	0.977*	0.934*
	(0.220)	(0.007)	(0.191)	(0.042)	(0.270)	(0.137)	(0.024)	(0.031)
Best PC-SDV	1.079	1.09	0.934*	0.886*	0.875*	0.909*	0.962*	0.775*
	(0.974)	(0.975)	(0.019)	(0.006)	(0.076)	(0.011)	(0.073)	(0.038)
Best UC-PC	1.023	1.012	0.887*	0.899*	0.927	0.934*	1.013	0.817*
	(0.787)	(0.589)	(0.043)	(0.059)	(0.104)	(0.000)	(0.735)	(0.044)
Best UC-PC-SDV	1.014	1.018	0.935	0.880*	0.893*	0.873*	1.002	0.789**
	(0.730)	(0.642)	(0.161)	(0.019)	(0.100)	(0.052)	(0.521)	(0.019)
				III: PCE Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Best time series model	0.897*	0.920*	0.958	0.979	0.896	0.979	0.998	0.996
	(0.012)	(0.012)	(0.154)	(0.253)	(0.141)	(0.201)	(0.265)	(0.476)
Best PC	0.887*	0.990	0.975*	0.928	0.907*	0.962	0.953*	0.846*
	(0.011)	(0.346)	(0.054)	(0.100)	(0.083)	(0.201)	(0.069)	(0.000)
Best PC-SDV	0.895*	0.963	0.999	1.028	0.833*	0.946	0.943	0.920*
	(0.009)	(0.264)	(0.493)	(0.669)	(0.069)	(0.169)	(0.188)	(0.011)
Best UC-PC	0.785*	0.913*	0.924*	0.896*	0.785*	0.887**	0.935**	0.807*
	(0.000)	(0.016)	(0.013)	(0.011)	(0.090)	(0.024)	(0.039)	(0.001)
Best UC-PC-SDV	0.874*	0.945*	0.958	0.874*	0.775*	0.854*	0.861*	0.979
	(0.029)	(0.008)	(0.184)	(0.048)	(0.001)	(0.083)	(0.049)	(0.395)

Table 2.7: Forecasting comparison for the out-of-sample performance of the best models from different categories: time series models, the Phillips curve models (PC), the Phillips curve model with score-driven variance (PC-SDV), the UC Phillips curve model (UC-PC) and the best UC Phillips curve model with score-driven variance (UC-PC-SDV). Out-of-sample forecasts are made for quarterly U.S. inflation. Each entry reports the ratio of the model RMSE to the RMSE of the benchmark (AR(AIC)) forecasts. The smallest RMSE ratios at each forecast horizons are highlighted. Entries in brackets denoted the DM-test p-values, which indicate the significance of a model's performance relative to the benchmark model, where * indicate rejection of the null at 10% level.

	Post-1985 sample				Post-2008 sample			
Horizon:	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
				I: CORE-CPI Index				
Benchmark	0.000	0.552*	0.000	0.000	0.057	0.762*	0.672*	0.830*
Best time series model	1.000*	0.852*	0.523*	0.003	1.000*	0.762*	0.794*	0.830*
Best PC	0.000	1.000*	0.001	0.003	0.091	1.000*	0.794*	1.000*
Best PC-SDV	0.000	0.007	1.000*	1.000*	0.090	0.199*	1.000*	0.830*
Best UC-PC	0.000	0.000	0.000	0.000	0.091	0.030	0.055	0.067
Best UC-PC-SDV	0.000	0.000	0.000	0.000	0.090	0.030	0.025	0.013
				II: ALL-CPI Index				
benchmark	0.781*	0.356*	0.123*	0.001	0.216*	0.118*	0.487*	0.113*
Best time series model	0.781*	0.372*	0.353*	0.735*	0.420*	0.587*	0.691*	0.413*
Best PC	1.000*	1.000*	0.353*	0.910*	0.136*	0.118*	0.691*	0.413*
Best PC-SDV	0.319*	0.045	0.353*	0.910*	1.000*	0.587*	1.000*	1.000*
Best UC-PC	0.781*	0.372*	1.000*	0.735*	0.338*	0.493*	0.653*	0.504*
Best UC-PC-SDV	0.781*	0.372*	0.184*	1.000*	0.216*	1.000*	0.691*	0.800*
				III: PCE Index				
benchmark	0.014	0.107*	0.157*	0.016	0.206*	0.182*	0.135*	0.000
Best time series model	0.109*	0.699*	0.550*	0.018	0.464*	0.182*	0.135*	0.035
Best PC	0.109*	0.217*	0.550*	0.146*	0.464*	0.182*	0.135*	0.591*
Best PC-SDV	0.109*	0.345*	0.550*	0.016	0.584*	0.182*	0.368*	0.055
Best UC-PC	1.000*	1.000*	1.000*	0.741*	0.877*	0.101*	0.554*	1.000*
Best UC-PC-SDV	0.109*	0.217*	0.550*	1.000*	1.000*	1.000*	1.000*	0.035

Table 2.8: reports the p-value of model confidence set test for the best models from different categories: time series models, the Phillips curve models (PC), the Phillips curve model with score driven variance (PC-SDV), the UC Phillips curve model (UC-PC) and the best UC Phillips curve model with score driven variance (UC-PC-SDV). Out-of-sample forecasts are made for quarterly U.S. inflation. Entries with an asterisk imply the corresponding models are selected by the MCS method under 10% confidence level.

termine the best model set among the best forecasting models. The results for model confidence set method are reported in Table (2.4). Entry with an asterisk means the corresponding model is included in the best model set, at 10% confidence level. When forecasting core-CPI inflation, the best time series model is included in the best model set in most cases. When forecasting all-CPI and PCE inflation, all the considered forecasting models are included in the best model set in most cases.

2.4 Conclusion

We have studied different univariate time series models and four different Phillips curve models in their ability to forecast U.S. inflation measures. The forecast evaluations have been based on three different measures of inflation: core-CPI, all-CPI and CPE

inflation. The forecasting precision is measured by the root mean squared error (RMSE). Furthermore, several tests have been conducted to show the significance of the forecasting performances, including the Diebold and Mariano (1995a) test and the model confidence set method. In the empirical study, we have first investigated whether the inclusion of an unobserved component for volatility in the classic Phillips curve model leads to more accurate forecasts of U.S. inflation. The results show that adding the time-varying components does not always lead to a better forecasting model. Then we have compared the forecasting performance of different Phillips curve models with univariate time series models to get an overall picture of the performances of the considered models. We found that the time series models provide consistent accurate forecasts at different horizons in the great moderation period when the forecasting target is core-CPI inflation. The UC Phillips curve model provides the most accurate forecasts for the PCE inflation measure. Although the UC Phillips curve model outperforms the classic Phillips curve, we can conclude that the Phillips curve is still a good benchmark model for the forecasting of U.S. inflation.

2.5 Extra Tables

Table 2.9: Forecast evaluation: Phillips curve model

	Post-1985 sample				Post-2008 sample			
Horizon:	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
				I: CORE-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	1.000	0.983	0.997	0.980*	1.023	1.024	1.029	1.028
	(0.497)	(0.244)	(0.338)	(0.064)	(0.899)	(0.716)	(0.948)	(0.585)
RCONS	1.016	1.026	0.925**	0.874**	1.008	0.963	0.965	1.002
	(0.983)	(0.818)	(0.016)	(0.036)	(0.693)	(0.134)	(0.119)	(0.506)
NFPR	0.932**	0.993	0.923	0.837*	0.944	0.995*	0.982***	0.932
	(0.039)	(0.439)	(0.196)	(0.073)	(0.306)	(0.073)	(0.005)	(0.108)
HSFR	1.008	1.013	0.996	0.994	1.010	1.020	1.015	1.018
	(0.885)	(0.849)	(0.357)	(0.196)	(0.750)	(0.782)	(0.670)	(0.638)
UNEMP	1.023	1.031	1.038	1.017	0.998	1.037	1.021	1.001
	(0.900)	(0.819)	(0.840)	(0.782)	(0.484)	(0.698)	(0.666)	(0.506)
YFGM3	0.951	1.041	1.009	0.992	0.999	0.998	1.007	0.899*
	(0.134)	(0.916)	(0.668)	(0.341)	(0.433)	(0.409)	(0.598)	(0.053)
IPI	0.972*	0.998	0.989	1.000	1.022	1.024	0.983*	0.940
	(0.051)	(0.461)	(0.306)	(0.508)	(0.726)	(0.873)	(0.074)	(0.168)
PMI	0.935	1.082	1.013	0.982	0.997	0.983*	1.015	0.927
	(0.101)	(0.971)	(0.704)	(0.146)	(0.326)	(0.076)	(0.773)	(0.276)
RDPI	1.054	1.041	1.007	1.001	0.947	0.997	0.983	0.986
	(0.889)	(0.952)	(0.820)	(0.521)	(0.187)	(0.319)	(0.151)	(0.258)
				II: ALL-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	0.988	0.993	0.972*	0.990	1.016	1.028	0.985	1.010
	(0.182)	(0.244)	(0.081)	(0.109)	(0.964)	(0.916)	(0.213)	(0.851)

RCONS	1.011	0.937***	0.988	0.951**	1.008	1.028	0.977**	1.032
	(0.810)	(0.007)	(0.266)	(0.027)	(0.918)	(0.724)	(0.024)	(0.724)
NFPR	1.000	0.991	0.964	0.894**	0.994	0.995	0.988	0.934**
	(0.491)	(0.342)	(0.191)	(0.042)	(0.203)	(0.137)	(0.205)	(0.031)
HSFR	1.003	0.994	1.002	1.005	1.026	1.027	1.006	1.062
	(0.787)	(0.289)	(1.000)	(0.750)	(0.982)	(0.951)	(0.624)	(0.898)
UNEMP	1.018	1.024	1.027	1.022	1.003	1.001	1.006	1.031
	(0.963)	(0.945)	(0.863)	(0.873)	(0.596)	(0.576)	(0.925)	(0.714)
YFGM3	1.001	1.008	0.988	1.004	1.009	1.001	0.985*	0.997
	(0.529)	(0.707)	(0.249)	(0.762)	(0.822)	(0.544)	(0.070)	(0.454)
IPI	0.974	0.987	0.987	0.985	1.113	1.024	1.004	1.018
	(0.220)	(0.267)	(0.320)	(0.122)	(0.831)	(0.801)	(0.818)	(0.847)
PMI	0.993	1.020	0.988	1.002	1.010	1.006	0.988	0.998
	(0.289)	(0.915)	(0.206)	(0.726)	(0.833)	(0.757)	(0.227)	(0.463)
RDPI	1.007	0.999	0.994	0.998	0.993	1.006	1.008	1.065
	(0.787)	(0.476)	(0.394)	(0.435)	(0.270)	(0.644)	(0.963)	(0.773)
				III: PCE Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	0.991**	0.990*	0.991**	1.001	0.977*	0.996	1.000	1.052
	(0.023)	(0.056)	(0.036)	(0.635)	(0.065)	(0.377)	(0.499)	(0.809)
RCONS	0.886**	0.994	0.975*	0.995	0.907*	0.962	0.959	0.951*
	(0.011)	(0.286)	(0.054)	(0.185)	(0.083)	(0.201)	(0.166)	(0.087)
NFPR	1.001	1.002	0.983	0.928	0.983**	0.977*	0.953*	0.846***
	(0.507)	(0.511)	(0.386)	(0.100)	(0.048)	(0.077)	(0.069)	(0.000)
HSFR	1.025	0.999	1.002	0.987	1.024	1.009	1.031	1.282
	(0.978)	(0.484)	(0.551)	(0.171)	(0.998)	(0.655)	(0.678)	(0.989)
UNEMP	0.986	0.990	0.983	0.966	1.006	1.004	1.020	1.274
	(0.273)	(0.346)	(0.275)	(0.102)	(0.833)	(0.596)	(0.740)	(0.903)
YFGM3	0.999	1.000	0.996	1.001	1.015	0.980	1.011	1.031
	(0.484)	(0.496)	(0.286)	(0.533)	(0.858)	(0.100)	(1.000)	(0.915)
IPI	0.987*	0.998	0.995	1.003	0.980	0.983	0.996	1.028
	(0.063)	(0.427)	(0.320)	(0.731)	(0.368)	(0.343)	(0.330)	(0.675)
PMI	0.995	0.992	0.997	0.995	0.999	0.982	1.003	1.050
	(0.374)	(0.266)	(0.270)	(0.239)	(0.403)	(0.119)	(0.854)	(0.895)
RDPI	1.002	1.011	1.009	1.003	1.005	1.024	1.018	0.992
	(0.583)	(0.839)	(0.973)	(0.711)	(0.912)	(0.946)	(0.990)	(0.355)

Table 2.9: We forecast the quarterly inflation of U.S. over 1985Q1-2015Q4 and 2008Q1-2015Q4. Entries are RMSEs relative to the recursively estimated AR(AIC) model (benchmark). Entries in brackets denoted the p-values indicate the significance of a model's performance relative to the benchmark model , where *,** and *** indicate rejection of the null at 10%,5%, and 1% level.The best models for every inflation measure at different forecast horizons are highlighted.

Table 2.10: Forecast evaluation: Phillips curve model (MCS)

		Post-1985 sample				Post-2008 sample			
Horizon:		h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
CORE-CPI	Benchmark	0.013	0.880*	0.035	0.001	0.351*	0.883*	0.258*	0.755*
	GDP	0.061	1.000*	0.084	0.008	0.274*	0.838*	0.526*	0.285*
	RCONS	0.032	0.861*	0.999*	0.623*	0.674*	0.956*	0.895*	0.433*
	NFPR	1.000*	0.880*	1.000*	1.000*	0.952*	0.956*	1.000*	0.935*

	HSFR	0.032	0.861*	0.011	0.012	0.674*	0.487*	0.184*	0.755*
	UNEMP	0.000	0.831*	0.084	0.012	0.964*	0.838*	0.734*	0.935*
	YFGM3	0.399*	0.200*	0.084	0.017	0.432*	0.825*	0.230*	0.767*
	IPI	0.611*	0.880*	0.305*	0.012	0.647*	0.487*	0.889*	0.935*
	PMI	0.966*	0.015	0.084	0.039	0.517*	0.825*	0.187*	0.767*
	RDPI	0.061	0.861*	0.004	0.002	1.000*	1.000*	0.497*	1.000*
ALL-CPI	Benchmark	0.618*	0.490*	0.204*	0.003	0.388*	0.687*	0.597*	0.056
	GDP	0.940*	0.490*	0.971*	0.005	0.535*	0.619*	0.554*	0.141*
	RCONS	0.617*	1.000*	0.888*	0.094	0.831*	1.000*	1.000*	0.144*
	NFPR	0.867*	0.490*	1.000*	1.000*	1.000*	0.687*	0.887*	1.000*
	HSFR	0.659*	0.490*	0.094	0.014	0.002	0.443*	0.287*	0.144*
	UNEMP	0.430*	0.408*	0.261*	0.014	0.831*	0.687*	0.554*	0.047
	YFGM3	0.867*	0.201*	0.861*	0.003	0.380*	0.687*	0.554*	0.056
	IPI	1.000*	0.490*	0.888*	0.052	0.388*	0.687*	0.887*	0.052
	PMI	0.940*	0.090	0.861*	0.002	0.272*	0.678*	0.597*	0.052
	RDPI	0.617*	0.490*	0.861*	0.003	0.831*	0.617*	0.554*	0.352*
PCE Index	Benchmark	0.046	0.550*	0.234*	0.009	0.212*	0.833*	0.253*	0.021
	GDP	0.225*	0.882*	0.821*	0.002	0.766*	0.833*	0.171*	0.021
	RCONS	1.000*	0.874*	1.000*	0.035	1.000*	0.833*	0.253*	0.075
	NFPR	0.225*	0.924*	0.981*	1.000*	0.766*	1.000*	1.000*	1.000*
	HSFR	0.070	0.817*	0.148*	0.310*	0.003	0.288*	0.037	0.021
	UNEMP	0.225*	0.924*	0.981*	0.310*	0.205*	0.585*	0.090	0.047
	YFGM3	0.072	0.120*	0.234*	0.026	0.327*	0.833*	0.006	0.041
	IPI	0.225*	0.817*	0.720*	0.035	0.111*	0.585*	0.253*	0.024
	PMI	0.076	0.817*	0.158*	0.021	0.721*	0.216*	0.253*	0.075
	RDPI	0.059	0.422*	0.068	0.035	0.170*	0.791*	0.253*	0.075

Table 2.10: We forecast the quarterly inflation of U.S. over 1985Q1-2015Q4 and over 2008Q1-2015Q4. Entries are p-value of the model confidence set test, where * indicate that the corresponding model is selected in the best model set at 10% level.

Table 2.11: Forecast evaluation: Phillips curve model with score-driven variance

	Post-1985 sample				Post-2008 sample			
Horizon:	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
				I: CORE-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	1.257	1.327	1.054	0.792***	0.977	1.558	1.000	1.017
	(0.984)	(0.963)	(0.764)	(0.000)	(0.311)	(0.995)	(0.499)	(0.632)
RCONS	1.097	1.298	0.857**	0.707***	0.911***	1.315	1.003	0.995
	(0.989)	(0.977)	(0.012)	(0.006)	(0.010)	(0.982)	(0.514)	(0.477)
NFPR	1.634	1.332	0.928*	0.807**	0.991	1.356	0.939	0.996
	(1.000)	(0.956)	(0.097)	(0.022)	(0.433)	(0.968)	(0.222)	(0.492)
HSFR	1.099	1.325	0.731***	1.022	0.946	1.356	0.970	1.004
	(0.980)	(0.969)	(0.004)	(0.575)	(0.114)	(0.993)	(0.322)	(0.527)
UNEMP	1.169	1.238	0.863***	0.787***	1.189	1.728	1.354	0.968
	(0.963)	(0.951)	(0.002)	(0.000)	(0.869)	(1.000)	(0.955)	(0.350)
YFGM3	1.223	1.284	1.052	0.692***	0.889**	1.299	0.967	1.003
	(0.996)	(0.958)	(0.698)	(0.002)	(0.016)	(0.983)	(0.254)	(0.516)
IPI	1.122	1.343	0.740***	0.841***	0.993	1.312	0.931	1.066
	(0.953)	(0.978)	(0.002)	(0.001)	(0.454)	(0.992)	(0.170)	(0.973)
PMI	1.283	1.289	0.881***	0.718***	0.892**	1.297	0.978	0.980
	(0.996)	(0.964)	(0.006)	(0.002)	(0.017)	(0.984)	(0.334)	(0.380)
RDPI	1.099	1.325	0.751***	0.791***	0.952	1.293	0.963	1.054
	(0.906)	(0.980)	(0.002)	(0.006)	(0.203)	(0.983)	(0.227)	(0.760)
				II: ALL-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	1.088	1.123	0.965*	0.887***	0.925*	0.924**	1.008	0.897
	(0.967)	(0.993)	(0.090)	(0.004)	(0.066)	(0.026)	(0.586)	(0.178)
RCONS	1.148	1.090	0.953*	0.900***	0.921*	0.909**	0.962*	0.894*
	(0.990)	(0.975)	(0.060)	(0.005)	(0.090)	(0.011)	(0.073)	(0.100)
NFPR	1.189	1.182	0.942**	0.907*	0.901*	0.925**	1.019	0.775**
	(1.000)	(0.987)	(0.035)	(0.070)	(0.081)	(0.025)	(0.687)	(0.038)
HSFR	1.127	1.143	0.969	0.920***	0.940*	0.929**	1.005	0.895*
	(0.976)	(0.996)	(0.141)	(0.009)	(0.055)	(0.027)	(0.558)	(0.085)
UNEMP	1.088	1.126	0.970	0.894***	0.916	0.926**	1.073	0.880
	(0.982)	(0.992)	(0.216)	(0.001)	(0.130)	(0.036)	(0.901)	(0.146)
YFGM3	1.086	1.132	0.950*	0.913***	0.914	0.935**	0.971	0.919
	(0.990)	(0.995)	(0.051)	(0.010)	(0.100)	(0.011)	(0.164)	(0.230)
IPI	1.099	1.114	1.025	0.900***	0.888*	0.929**	1.014	0.931
	(0.976)	(0.994)	(0.702)	(0.004)	(0.099)	(0.020)	(0.672)	(0.169)
PMI	1.079	1.151	0.940**	0.898***	0.906*	0.948***	1.043	0.908*
	(0.974)	(0.997)	(0.028)	(0.004)	(0.086)	(0.002)	(0.929)	(0.056)
RDPI	1.169	1.107	0.934**	0.885***	0.874*	0.948**	0.989	1.024
	(0.991)	(0.986)	(0.019)	(0.006)	(0.076)	(0.019)	(0.324)	(0.557)
				III: PCE Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	0.990	0.981	1.006	1.101	1.039	0.973	1.033	1.184
	(0.346)	(0.203)	(0.560)	(0.962)	(0.806)	(0.299)	(0.775)	(0.931)
RCONS	0.895***	1.041	1.035	1.111	0.833*	0.946	0.970	1.054
	(0.009)	(0.816)	(0.887)	(0.957)	(0.069)	(0.169)	(0.243)	(0.748)
NFPR	1.228	1.056	1.060	1.028	1.032	0.966	0.965*	0.920**
	(0.992)	(0.723)	(0.793)	(0.669)	(0.777)	(0.216)	(0.091)	(0.011)
HSFR	1.128	0.963	0.999	1.100	1.124	0.950	0.943	1.127
	(0.975)	(0.264)	(0.493)	(0.948)	(0.987)	(0.172)	(0.188)	(0.991)

UNEMP	1.047	1.102	1.047	1.077	1.228	0.980	1.035	1.287
	(0.879)	(0.953)	(0.965)	(0.923)	(0.907)	(0.299)	(0.813)	(0.992)
YFGM3	1.107	0.966	1.034	1.039	1.055	1.003	1.019	0.936
	(0.952)	(0.110)	(0.903)	(0.847)	(0.906)	(0.544)	(0.689)	(0.166)
IPI	0.998	1.039	1.010	1.096	1.064	0.975	0.982	1.145
	(0.471)	(0.785)	(0.608)	(0.996)	(0.883)	(0.314)	(0.154)	(0.966)
PMI	1.079	1.007	1.027	1.096	1.139	0.985	0.995	1.228
	(0.927)	(0.668)	(0.760)	(0.971)	(0.973)	(0.315)	(0.337)	(0.998)
RDPI	1.063	1.011	1.036	1.067	1.196	0.984	1.007	1.016
	(0.956)	(0.657)	(0.887)	(0.969)	(0.917)	(0.301)	(0.776)	(0.577)

Table 2.11: We forecast the quarterly inflation of U.S. over 1985Q1-2015Q4 and 2008Q1-2015Q4. Entries are RMSEs relative to the recursively estimated AR(AIC) model (benchmark). Entries in brackets denoted the p-values indicate the significance of a model's performance relative to the benchmark model , where *,** and *** indicate rejection of the null at 10%,5%, and 1% level.The best models for every inflation measure at different forecast horizons are highlighted.

Table 2.12: Forecast evaluation: Phillips curve model with score-driven variance(MCS)

		Post-1985 sample				Post-2008 sample			
Horizon:		h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
CORE-CPI	Benchmark	1.000*	1.000*	0.000	0.000	0.826*	1.000*	0.976*	1.000*
	GDP	0.006	0.010	0.002	0.044	0.350*	0.602*	0.820*	0.587*
	RCONS	0.016	0.012	0.269*	0.639*	0.904*	0.931*	0.846*	0.520*
	NFPR	0.000	0.012	0.038	0.008	0.826*	0.931*	0.820*	0.252*
	HSFR	0.016	0.010	1.000*	0.000	0.722*	0.931*	0.515*	0.587*
	UNEMP	0.016	0.012	0.134*	0.074	0.350*	0.138*	0.441*	0.791*
	YFGM3	0.001	0.012	0.018	1.000*	0.904*	0.931*	0.913*	0.587*
	IPI	0.016	0.010	0.716*	0.035	0.904*	0.900*	0.913*	0.587*
	PMI	0.001	0.012	0.250*	0.210*	1.000*	0.931*	0.233*	0.759*
	RDPI	0.032	0.012	0.716*	0.044	0.904*	0.931*	0.936*	0.759*
ALL-CPI	Benchmark	1.000*	1.000*	0.271*	0.001	0.020	0.145*	0.937*	0.148*
	GDP	0.282*	0.134*	0.616*	0.943*	0.309*	0.314*	0.937*	0.148*
	RCONS	0.281*	0.142*	0.763*	0.579*	0.134*	1.000*	0.937*	0.148*
	NFPR	0.009	0.134*	0.894*	0.943*	0.274*	0.176*	0.858*	1.000*
	HSFR	0.282*	0.065	0.127*	0.527*	0.702*	0.189*	0.937*	0.148*
	UNEMP	0.282*	0.134*	0.616*	0.654*	0.274*	0.263*	0.031	0.148*
	YFGM3	0.282*	0.134*	0.763*	0.369*	0.199*	0.314*	0.447*	0.148*
	IPI	0.282*	0.134*	0.612*	0.579*	0.274*	0.314*	0.484*	0.136*
	PMI	0.282*	0.073	0.894*	0.579*	0.274*	0.620*	0.000	0.148*
	RDPI	0.248*	0.142*	1.000*	0.943*	1.000*	0.314*	1.000*	0.148*
PCE	Benchmark	0.272*	0.311*	0.987*	1.000*	0.292*	0.988*	0.845*	0.308*
	GDP	0.272*	0.314*	0.975*	0.060	0.292*	0.173*	0.153*	0.024
	RCONS	1.000*	0.311*	0.586*	0.096	0.918*	0.988*	0.901*	0.001
	NFPR	0.009	0.248*	0.812*	0.525*	0.281*	0.840*	0.845*	1.000*
	HSFR	0.035	0.314*	1.000*	0.096	0.006	0.830*	0.817*	0.090
	UNEMP	0.161*	0.061	0.604*	0.096	0.281*	1.000*	0.151*	0.022
	YFGM3	0.064	0.314*	0.644*	0.279*	0.242*	0.027	0.901*	0.256*
	IPI	0.272*	0.314*	0.951*	0.005	0.281*	0.018	0.371*	0.024
	PMI	0.064	0.314*	0.812*	0.060	0.000	0.540*	0.206*	0.001

RDPI	0.163*	0.314*	0.812*	0.096	0.020	0.840*	0.206*	0.018

Table 2.12: We forecast the quarterly inflation of U.S. over 1985Q1-2015Q4 and over 2008Q1-2015Q4. Entries are p-value of the model confidence set test, where * indicate that the corresponding model is selected in the best model set at 10% level.

Table 2.13: Forecast evaluation: UC Phillips curve model

	Post-1985 sample				Post-2008 sample			
Horizon:	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
				I: CORE-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	0.946**	1.428	1.394	1.028	1.075	1.480	1.511	1.495
	(0.019)	(1.000)	(0.999)	(0.590)	(0.727)	(0.988)	(1.000)	(1.000)
RCONS	0.961	1.422	1.416	1.075	1.061	1.458	1.422	1.366
	(0.106)	(1.000)	(1.000)	(0.739)	(1.000)	(0.995)	(0.993)	(0.999)
NFPR	0.964***	1.435	1.403	1.029	1.058	1.436	1.390	1.317
	(0.000)	(1.000)	(1.000)	(0.593)	(0.750)	(1.000)	(0.999)	(0.983)
HSFR	0.939***	1.420	1.419	1.042	1.051	1.470	1.483	1.481
	(0.000)	(1.000)	(1.000)	(0.634)	(0.831)	(0.995)	(0.998)	(1.000)
UNEMP	0.948***	1.416	1.382	1.033	1.029	1.428	1.425	1.471
	(0.000)	(1.000)	(0.999)	(0.605)	(0.580)	(0.987)	(0.993)	(1.000)
YFGM3	0.959***	1.430	1.399	1.059	1.037	1.454	1.453	1.325
	(0.000)	(1.000)	(0.999)	(0.685)	(0.645)	(0.999)	(0.992)	(0.990)
IPI	0.951***	1.416	1.397	1.042	1.087	1.503	1.517	1.458
	(0.000)	(1.000)	(0.999)	(0.629)	(0.881)	(0.995)	(0.999)	(1.000)
PMI	0.950**	1.425	1.399	1.043	1.037	1.455	1.459	1.355
	(0.030)	(1.000)	(0.999)	(0.637)	(1.000)	(0.999)	(0.995)	(0.997)
RDPI	0.973	1.406	1.402	1.032	1.000	1.397	1.417	1.285
	(0.308)	(1.000)	(1.000)	(0.603)	(0.500)	(1.000)	(0.987)	(0.993)
				II: All-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	1.046	1.042	0.893**	0.908*	0.949**	0.960***	1.022	0.871*
	(0.896)	(0.800)	(0.047)	(0.088)	(0.033)	(0.000)	(0.855)	(0.088)
RCONS	1.073	1.012	0.895**	0.913*	0.937**	0.934***	1.016	0.857*
	(0.987)	(0.589)	(0.046)	(0.084)	(0.036)	(0.000)	(0.738)	(0.071)
NFPR	1.050	1.043	0.905*	0.907*	0.934*	0.940**	1.013	0.817**
	(0.928)	(0.799)	(0.065)	(0.078)	(0.054)	(0.016)	(0.735)	(0.044)
HSFR	1.023	1.030	0.895*	0.909*	0.960**	0.960***	1.034	0.894
	(0.787)	(0.731)	(0.051)	(0.083)	(0.047)	(0.001)	(0.912)	(0.108)
UNEMP	1.028	1.029	0.887**	0.905*	0.943**	0.941**	1.028	0.917
	(0.794)	(0.713)	(0.043)	(0.074)	(0.019)	(0.018)	(0.861)	(0.264)
YFGM3	1.042	1.038	0.898*	0.900*	0.945*	0.949***	1.029	0.856*
	(0.907)	(0.819)	(0.055)	(0.062)	(0.066)	(0.000)	(0.891)	(0.079)
IPI	1.047	1.040	0.907*	0.908*	0.958	0.960	1.031	0.870*
	(0.935)	(0.789)	(0.066)	(0.082)	(0.120)	(0.128)	(0.875)	(0.077)
PMI	1.034	1.043	0.898*	0.899*	0.947*	0.944***	1.024	0.849*
	(0.850)	(0.822)	(0.056)	(0.059)	(0.067)	(0.001)	(0.920)	(0.090)
RDPI	1.049	1.040	0.906*	0.908*	0.927	0.955	1.028	0.950
	(0.946)	(0.770)	(0.089)	(0.082)	(0.104)	(0.172)	(0.855)	(0.342)
				III: PCE Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	0.891***	0.920***	0.934**	0.926**	0.895	0.944	0.983	0.989
	(0.005)	(0.004)	(0.012)	(0.021)	(0.128)	(0.168)	(0.356)	(0.432)
RCONS	0.785***	0.919**	0.930**	0.926**	0.785*	0.887**	0.937	0.900*
	(0.000)	(0.017)	(0.026)	(0.015)	(0.090)	(0.024)	(0.112)	(0.053)
NFPR	0.953	0.958	0.975	0.921*	0.944	0.945	0.935**	0.807***
	(0.214)	(0.238)	(0.332)	(0.074)	(0.141)	(0.174)	(0.039)	(0.001)
HSFR	0.919*	0.918**	0.948**	0.914**	0.991	0.976	1.009	1.218
	(0.055)	(0.019)	(0.034)	(0.016)	(0.452)	(0.375)	(0.558)	(0.990)

UNEMP	0.892**	0.913**	0.923**	0.896**	0.975	0.994	1.015	1.191
	(0.016)	(0.016)	(0.013)	(0.011)	(0.292)	(0.443)	(0.814)	(0.876)
YFGM3	0.901**	0.931**	0.951*	0.927**	0.972	0.954	0.994	0.935
	(0.015)	(0.021)	(0.053)	(0.016)	(0.273)	(0.200)	(0.220)	(0.138)
IPI	0.893***	0.935**	0.956*	0.926**	0.888	0.935	0.984	0.976
	(0.008)	(0.012)	(0.055)	(0.024)	(0.194)	(0.238)	(0.175)	(0.337)
PMI	0.902**	0.930**	0.950**	0.926**	0.957	0.946	0.993	0.944
	(0.019)	(0.017)	(0.045)	(0.019)	(0.194)	(0.222)	(0.229)	(0.205)
RDPI	0.897**	0.930**	0.958*	0.930**	0.992	1.005	1.007	0.934
	(0.028)	(0.014)	(0.065)	(0.032)	(0.439)	(0.536)	(0.797)	(0.134)

Table 2.13: We forecast the quarterly inflation of U.S. over 1985Q1-2015Q4 and 2008Q1-2015Q4. Entries are RMSEs relative to the recursively estimated AR(AIC) model (benchmark). Entries in brackets denoted the p-values indicate the significance of a model's performance relative to the benchmark model , where *,** and *** indicate rejection of the null at 10%,5%, and 1% level.The best models for every inflation measure at different forecast horizons are highlighted.

Table 2.14: Forecast evaluation: UC Phillips curve model (MCS)

		Post-1985 sample				Post-2008 sample			
Horizon:		h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
CORE-CPI	Benchmark	0.600*	1.000*	1.000*	1.000*	0.892*	1.000*	1.000*	1.000*
	GDP	0.971*	0.000	0.000	0.992*	0.734*	0.392*	0.063	0.080
	RCONS	0.966*	0.000	0.000	0.574*	0.431*	0.392*	0.063	0.079
	NFPR	0.584*	0.000	0.000	0.992*	0.520*	0.389*	0.063	0.080
	HSFR	0.976*	0.000	0.000	0.920*	0.913*	0.389*	0.059	0.080
	UNEMP	0.973*	0.000	0.000	0.992*	1.000*	0.392*	0.062	0.063
	YFGM3	0.644*	0.000	0.000	0.448*	0.623*	0.333*	0.063	0.080
	IPI	0.971*	0.000	0.000	0.920*	0.431*	0.374*	0.017	0.080
	PMI	0.966*	0.000	0.000	0.702*	0.623*	0.131*	0.063	0.080
	RDPI	0.966*	0.000	0.000	0.992*	0.913*	0.362*	0.063	0.080
ALL-CPI	Benchmark	1.000*	1.000*	0.306*	0.303*	0.133*	0.122*	1.000*	0.190*
	GDP	0.736*	0.798*	0.720*	0.763*	0.825*	0.657*	0.452*	0.190*
	RCONS	0.435*	0.919*	0.720*	0.693*	1.000*	1.000*	0.917*	0.117*
	NFPR	0.733*	0.777*	0.490*	0.606*	0.949*	0.772*	0.917*	1.000*
	HSFR	0.876*	0.829*	0.720*	0.763*	0.021	0.314*	0.452*	0.190*
	UNEMP	0.876*	0.919*	1.000*	0.744*	0.993*	0.657*	0.059	0.037
	YFGM3	0.852*	0.894*	0.608*	0.763*	0.775*	0.772*	0.201*	0.066
	IPI	0.789*	0.746*	0.362*	0.763*	0.588*	0.657*	0.711*	0.080
	PMI	0.876*	0.786*	0.590*	1.000*	0.759*	0.772*	0.361*	0.066
	RDPI	0.499*	0.798*	0.637*	0.744*	0.900*	0.525*	0.417*	0.368*
PCE	Benchmark	0.034	0.293*	0.232*	0.041	0.067	0.663*	0.146*	0.044
	GDP	0.220*	0.925*	0.590*	0.316*	0.960*	0.443*	0.121*	0.044
	RCONS	1.000*	0.925*	0.635*	0.316*	1.000*	0.663*	0.459*	0.044
	NFPR	0.091	0.293*	0.206*	0.316*	0.960*	1.000*	1.000*	1.000*
	HSFR	0.053	0.891*	0.232*	0.272*	0.000	0.040	0.039	0.044
	UNEMP	0.220*	1.000*	1.000*	1.000*	0.960*	0.361*	0.121*	0.044
	YFGM3	0.220*	0.418*	0.232*	0.316*	0.007	0.302*	0.039	0.044
	IPI	0.220*	0.650*	0.232*	0.316*	0.260*	0.173*	0.105*	0.044
	PMI	0.220*	0.650*	0.232*	0.316*	0.006	0.283*	0.105*	0.044

RDPI	0.220*	0.650*	0.089	0.316*	0.342*	0.302*	0.105*	0.044

Table 2.14: We forecast the quarterly inflation of U.S. over 1985Q1-2015Q4 and over 2008Q1-2015Q4. Entries are p-value of the model confidence set test, where * indicate that the corresponding model is selected in the best model set at 10% level.

Table 2.15: Forecast evaluation: UC Phillips curve model with score-driven variances

	Post-1985 sample				Post-2008 sample			
Horizon:	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
				I: CORE-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	1.005	1.427	1.526	1.091	1.194	1.506	1.818	1.655
	(0.546)	(1.000)	(1.000)	(0.808)	(0.875)	(0.998)	(0.991)	(0.999)
RCONS	1.003	1.503	1.505	1.118	1.143	1.645	1.545	1.503
	(0.532)	(1.000)	(1.000)	(0.816)	(0.996)	(0.996)	(0.985)	(1.000)
NFPR	1.035	1.537	1.425	1.114	1.152	1.645	1.463	1.515
	(0.834)	(1.000)	(1.000)	(0.796)	(1.000)	(0.997)	(0.985)	(1.000)
HSFR	0.999	1.501	1.471	1.091	1.146	1.662	1.456	1.532
	(0.489)	(1.000)	(0.999)	(0.746)	(1.000)	(0.998)	(0.993)	(1.000)
UNEMP	1.028	1.521	1.512	1.109	1.244	1.675	1.585	1.571
	(0.992)	(1.000)	(1.000)	(0.786)	(0.994)	(0.995)	(0.996)	(1.000)
YFGM3	1.021	1.503	1.494	1.131	1.177	1.675	1.519	1.480
	(0.716)	(1.000)	(1.000)	(0.837)	(1.000)	(0.997)	(0.988)	(0.997)
IPI	2.513	1.662	1.480	1.056	1.338	1.603	1.633	1.534
	(0.857)	(1.000)	(0.999)	(0.675)	(0.996)	(0.959)	(0.973)	(1.000)
PMI	1.015	1.498	1.508	1.064	1.138	1.684	1.554	1.383
	(0.656)	(1.000)	(1.000)	(0.695)	(0.991)	(0.998)	(0.978)	(0.999)
RDPI	1.003	1.460	1.470	1.100	1.584	1.737	1.691	1.766
	(0.527)	(1.000)	(0.999)	(0.767)	(0.960)	(0.997)	(0.993)	(0.999)
				II: ALL-CPI Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	1.154	1.188	0.985	0.915*	0.969*	1.007	1.034	0.821*
	(0.986)	(0.979)	(0.427)	(0.088)	(0.095)	(0.571)	(0.733)	(0.053)
RCONS	1.090	1.072	0.982	0.980	0.988*	1.010	1.062	0.848*
	(0.940)	(0.839)	(0.420)	(0.405)	(0.051)	(0.670)	(0.790)	(0.097)
NFPR	1.085	1.140	0.935	0.972	0.922*	0.982**	1.002	0.816*
	(0.980)	(0.932)	(0.161)	(0.366)	(0.051)	(0.043)	(0.521)	(0.071)
HSFR	1.165	1.119	0.956	0.917*	1.018	0.997	1.048	0.832**
	(0.989)	(0.890)	(0.291)	(0.100)	(0.758)	(0.467)	(0.757)	(0.020)
UNEMP	1.128	1.018	0.981	0.969	1.118	0.873*	1.068	0.828
	(0.952)	(0.642)	(0.413)	(0.365)	(0.875)	(0.052)	(0.839)	(0.121)
YFGM3	1.014	1.070	0.980	0.879**	0.893*	0.917**	1.090	0.789**
	(0.730)	(0.936)	(0.406)	(0.019)	(0.100)	(0.024)	(0.810)	(0.019)
IPI	1.161	1.136	0.989	0.887**	1.002	0.979	1.065	0.875***
	(0.962)	(0.957)	(0.451)	(0.035)	(0.513)	(0.345)	(0.798)	(0.000)
PMI	1.114	1.145	0.963	0.986	0.962*	0.919**	1.135	0.932
	(0.936)	(0.945)	(0.331)	(0.438)	(0.095)	(0.037)	(0.839)	(0.189)
RDPI	1.188	1.042	0.948	0.975	0.993	0.949**	1.069	0.821*
	(0.993)	(0.778)	(0.257)	(0.387)	(0.372)	(0.013)	(0.807)	(0.094)
				III: PCE Index				
Benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GDP	0.907***	0.971	1.035	0.874**	0.893	0.947	1.182	1.074
	(0.001)	(0.204)	(0.647)	(0.048)	(0.172)	(0.233)	(0.691)	(0.753)
RCONS	0.874**	0.968	1.007	0.912**	0.775***	0.854*	0.905*	1.126
	(0.029)	(0.266)	(0.538)	(0.019)	(0.001)	(0.083)	(0.085)	(0.749)
NFPR	0.945	0.972	0.998	0.883*	0.884	0.939	0.861*	1.019
	(0.165)	(0.229)	(0.492)	(0.070)	(0.153)	(0.414)	(0.049)	(0.570)
HSFR	0.925**	0.955	1.029	0.936**	0.938	0.977	1.052	1.105
	(0.030)	(0.153)	(0.611)	(0.024)	(0.408)	(0.303)	(0.966)	(0.898)

UNEMP	0.925	1.011	1.010	0.905**	1.096	0.887	1.061	0.979
	(0.115)	(0.568)	(0.545)	(0.021)	(0.903)	(0.136)	(0.916)	(0.395)
YFGM3	0.934*	0.964	0.990	0.916***	0.950	1.257	0.941	1.031
	(0.090)	(0.204)	(0.424)	(0.007)	(0.198)	(0.876)	(0.234)	(0.600)
IPI	0.982	0.945***	0.964	0.944**	0.910	0.956	1.006	1.080
	(0.113)	(0.008)	(0.102)	(0.035)	(0.217)	(0.423)	(0.564)	(0.798)
PMI	0.898***	0.955*	0.958	0.962	0.898	1.099	1.064	0.983
	(0.004)	(0.090)	(0.184)	(0.120)	(0.152)	(0.905)	(0.946)	(0.424)
RDPI	0.926*	0.969	1.015	0.939**	1.003	1.020	0.989	0.982
	(0.063)	(0.205)	(0.597)	(0.046)	(0.545)	(0.923)	(0.321)	(0.401)

Table 2.15: We forecast the quarterly inflation of U.S. over 1985Q1-2015Q4 and 2008Q1-2015Q4. Entries are RMSEs relative to the recursively estimated AR(AIC) model (benchmark). Entries in brackets denoted the p-values indicate the significance of a model's performance relative to the benchmark model , where *,** and *** indicate rejection of the null at 10%,5%, and 1% level.The best models for every inflation measure at different forecast horizons are highlighted.

Table 2.16: Forecast evaluation: UC Phillips curve model with score-driven variance(MCS)

		Post-1985 sample				Post-2008 sample			
Horizon:		h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
CORE-CPI	Benchmark	0.815*	1.000*	1.000*	1.000*	0.955*	1.000*	1.000*	1.000*
	GDP	0.815*	0.000	0.002	0.749*	0.659*	0.257*	0.114*	0.117*
	RCONS	0.815*	0.000	0.002	0.054	0.527*	0.257*	0.114*	0.102*
	NFPR	0.143*	0.000	0.002	0.024	0.627*	0.257*	0.114*	0.117*
	HSFR	0.815*	0.000	0.002	0.728*	0.951*	0.254*	0.103*	0.084
	UNEMP	0.288*	0.000	0.002	0.054	1.000*	0.257*	0.114*	0.117*
	YFGM3	0.622*	0.000	0.002	0.013	0.307*	0.254*	0.114*	0.102*
	IPI	0.815*	0.000	0.002	0.814*	0.659*	0.257*	0.114*	0.117*
	PMI	0.623*	0.000	0.002	0.814*	0.969*	0.254*	0.114*	0.117*
	RDPI	0.815*	0.000	0.002	0.335*	0.685*	0.254*	0.114*	0.117*
ALL-CPI	Benchmark	1.000*	1.000*	0.821*	0.036	0.106*	0.378*	0.878*	0.290*
	GDP	0.080	0.170*	0.227*	0.272*	0.661*	0.619*	0.405*	0.290*
	RCONS	0.399*	0.724*	0.527*	0.110*	0.661*	0.694*	0.878*	0.795*
	NFPR	0.345*	0.416*	1.000*	0.155*	0.949*	0.694*	1.000*	0.906*
	HSFR	0.314*	0.684*	0.912*	0.272*	0.675*	0.045	0.797*	0.927*
	UNEMP	0.399*	0.879*	0.587*	0.267*	0.213*	1.000*	0.835*	0.685*
	YFGM3	0.662*	0.724*	0.501*	1.000*	0.775*	0.966*	0.558*	1.000*
	IPI	0.345*	0.088	0.501*	0.541*	0.090	0.966*	0.216*	0.747*
	PMI	0.399*	0.223*	0.912*	0.098	0.661*	0.966*	0.878*	0.927*
	RDPI	0.080	0.831*	0.912*	0.155*	0.567*	0.966*	0.848*	0.906*
PCE	Benchmark	0.018	0.484*	0.687*	0.117*	0.228*	0.118*	0.477*	0.616*
	GDP	0.913*	0.637*	0.363*	1.000*	0.264*	0.084	0.477*	0.446*
	RCONS	1.000*	0.987*	0.687*	0.850*	0.495*	1.000*	1.000*	0.430*
	NFPR	0.530*	0.949*	0.690*	0.950*	0.650*	0.042	0.595*	1.000*
	HSFR	0.676*	0.988*	0.687*	0.714*	0.253*	0.071	0.595*	0.484*
	UNEMP	0.869*	0.637*	0.690*	0.914*	0.662*	0.071	0.350*	0.386*
	YFGM3	0.665*	0.949*	0.514*	0.925*	1.000*	0.071	0.595*	0.616*
	IPI	0.270*	0.988*	0.819*	0.540*	0.232*	0.118*	0.506*	0.616*

PMI	0.913*	0.988*	1.000*	0.850*	0.662*	0.092	0.595*	0.616*
RDPI	0.764*	0.637*	0.363*	0.789*	0.310*	0.010	0.595*	0.386*

Table 2.16: We forecast the quarterly inflation of U.S. over 1985Q1-2015Q4 and over 2008Q1-2015Q4. Entries are p-value of the model confidence set test, where * indicate that the corresponding model is selected in the best model set at 10% level.

Chapter 3

Weighted Maximum Likelihood Estimation with Optimal Forecasting Accuracy: Forecasting During the Global Recession

3.1 Introduction

Linear autoregressive (AR) models have been the workhorse of time-series econometrics for several decades. The flexibility of AR models allows econometricians to describe relatively complex dynamics in a remarkably simple way. From short-run temporal dependence with seasonal dynamics to long-run persistency generated by stochastic trends, the AR model is often able to outperform larger and more complex models. Given the simplistic structure of this class of linear reduced form models, and the great complexity of the data generating process (DGP) that underlies large market economies, it is only natural however that AR models show various signs of model misspecification. Parameter instability is one way in which this misspecification usually manifests itself. Indeed, a significant body of literature has found that evidence for parameter instability is widely spread across economic and financial applications; see Stock and Watson (1996) Stock and Watson (2007), in macroeconomic prediction, Schinasi and Swamy (1989), and Wolff (1987) and Goyal and Welch (2003) in financial forecasting. Unfortunately, as pointed out by Inoue et al. (2014), parameter instability is widely recognized as a crucial issue that significantly hampers the forecasting performance of econometric models; see Rossi

(2013), Giacomini and Rossi (2009), Paye and Timmermann (2006), Koop and Potter (2004), Goyal and Welch (2003), Clements and Hendry (1998), Stock and Watson (1996). Several methods have been proposed in the literature to improve forecasting performance in the presence of parameter instability. Rolling window estimation is among the most popular with important applications in finance, (Goyal and Welch, 2003), macroeconomics (Swanson, 1998), and exchange rate forecasting (Molodtsova and Papell, 2009), among others. More flexible weighting schemes that do not focus only on a window of observations, In this chapter we propose a weighted maximum likelihood estimator (WMLE) that optimizes the forecasting performance of the misspecified Gaussian AR(p) model. Our WMLE gives different weights to the sample observations, and obtains parameter estimates that are optimal for forecasting. For example, the WMLE can be used to give higher weight to more recent observations compared to observations far in the past. This may be desirable as various political, institutional and technological change the economy, rendering past observations increasingly obsolete for forecasting today's economic variables. Similarly, the WMLE can be used to give higher weight to historical periods of greater relevance. We will see, for example, that IPI forecasts during the current economic recession can be improved by giving higher weight to past episodes of economic recession. In fact, at any given point in time, the WMLE weights can be adjusted to reflect the importance of past observations for improving forecast accuracy. For example, it is only natural that forecasts for inflation during an international oil crises can be improved by paying special attention to price dynamics during past oil crises. Just as forecasts for unemployment during a period of strong fiscal austerity can benefit from giving emphasizing observations that lie in past episodes of strong fiscal austerity. As we shall see, the WMLE estimator often leads to a time-varying parameter AR(p) model whose parameter estimates are optimal for forecasting in the given period of interest. we show that forecasts of US IPI during the latest global recession, can be significantly improved by defining weights that make past recession periods more informative, but also, downweight observations far in the past. The flexibility of our WMLE offers great generality to researchers and professional forecasters. In particular, in contains as a special case not only the classical MLE, but also many popular weighted estimation methods that can handle parameter instability such as rolling window estimators. In this chapter, we show how to obtain the optimal WMLE weights through a cross-validation procedure. This procedure automatically picks up the important features from past data and adjusts the weights to deliver optimal out-of-sample forecasting accuracy for the AR(p) model. We will show that our method of weight selection by cross-validation ensures that the WMLE converges to the classical MLE when the AR model is correctly specified. Hence, the WMLE retains the efficiency properties of the classical MLE when the parameters are time-invariant. In this case, the WMLE converges to the 'true' time-invariant parameter vector, which is optimal for forecasting. At the same time, we show that if the AR(p) model is misspecified, then the cross-validation leads to optimal weights, and the WMLE provides time-varying parameter estimates for the AR(p) which significantly improve the

forecasting accuracy of the classical MLE. In contrast to the bulk of the literature, our method for finding optimal weights is shown to be valid under very general conditions on the DGP that is not restricted to pre-specified breaks, trends, etc. We also deviate from a substantial part of the literature in founding our approach on the maximum likelihood estimation of AR(p) models. Indeed, while Giraitis et al. (2012) have already proposed a cross-validation method for optimizing forecasts, their method is not related to maximum likelihood estimation of parameters and they do not use AR(p) models to forecast. Instead, Giraitis et al. (2012) produce one-step-ahead predictions directly through a weighted sample average of past observations. In essence, our approach defines also defines such weights, but it does so implicitly through the structure of the AR(p) and its estimated parameters. By taking the AR(p) as a starting point, we take advantage of the ability of this model to describe relatively complex dynamic dependence patterns in the data, including short-run dynamics, seasonalities, stochastic trends, etc. As such, we start at an appreciable level of forecasting accuracy. The WMLE is then used to further optimize the forecasting performance of the model. Additional empirical reveal also that the advantages of the WMLE extend also to forecasting a number of other time-series.

This chapter is organized as follows. Section 3.2 introduces the basic concepts behind weighted maximum likelihood and optimality in forecasting. Section 3.3.3 compares the WMLE with time-varying parameter models. Sections 3.6 feature applications of the WMLE to real data where the optimal weighting function takes an intuitive albeit complex form that would be difficult to produce using a typical time-varying parameter model since high weight must be given to certain observations far in the past.

3.2 Basic Concepts

In order to easily grasp the basic concepts underlying the WMLE, we look first at a very simple example concerned with estimating a population mean.

Consider a sequence of random variables $X_1, ..., X_T$ obtained from an unknown data generating process (DGP). Suppose that we decide to work with a statistical model which assumes that these variables are independent and identically distributed with Gaussian distribution $X_t \sim N(\mu, 1)$. In this case, the MLE of the unknown parameter μ, based on the sample $X_{s:T} := \{X_t\}_{t=s}^{T}$, for some $1 \leq s < T$, is naturally given by

$$\hat{\mu}_{s:T} = \arg\max \sum_{t=s}^{T} (X_t - \mu)^2 \quad \Rightarrow \quad \hat{\mu}_{s:T} = \frac{1}{T-s+1} \sum_{t=s}^{T} X_t.$$

Since the statistical model assumes that the data is independent, the n-step ahead forecast $\hat{X}_{T+n}$ conditional on the observed sample $X_s, ..., X_T$ is

$$\hat{x}_{T+n} = \mathbb{E}(X_{T+n} | X_s, ..., X_T) = \hat{\mu}_{s:T}.$$

3. Weighted Maximum Likelihood Estimation with Optimal Forecasting Accuracy: Forecasting During the Global Recession

If the assumptions imposed by the statistical model is correct, then the model is well specified and the DGP consists of a sequence of iid $N(\mu_0, 1)$ random variables with some unknown μ_0. In this special case, it is optimal to use the complete available sample $X_1, ..., X_T$, and hence set $s = 1$. Furthermore, the MLE provides not only an unbiased forecast $\hat{X}_{T+n}$, but also, the most accurate forecast among all unbiased estimators. Indeed, we have

$$\text{Bias}(\hat{X}_{T+n}) = \mathbb{E}(\hat{X}_{T+n} - X_{T+n}) = \mathbb{E}\hat{\mu}_{s:T} - \mu_0 = \frac{1}{T-s+1}\sum_{t=s}^{T}\mu_0 - \mu_0 = 0 \ ,$$

and the variance attains the Cramer-Rao lower bound (CRLB)

$$\mathbb{V}\text{ar}(\hat{X}_{T+n}) = \mathbb{V}\text{ar}(\hat{\mu}_{s:T}) = \frac{1}{(T-s+1)^2}\sum_{t=s}^{T}\mathbb{V}\text{ar}(X_t) = \frac{1}{T-s+1} = \text{CRLB}.$$

As a result, the forecast mean squared error (MSE) is the smallest among all the unbiased forecasts and it becomes smaller by setting $s = 1$ and letting the sample size T increase

$$\text{MSE}(\hat{X}_{T+n}) = \Big(\text{Bias}(\hat{X}_{T+n})\Big)^2 + \mathbb{V}\text{ar}(\hat{X}_{T+n}) = \frac{1}{T-s+1} \ \to \ 0 \quad \text{as} \quad T-s+1 \to \infty.$$

Unfortunately, this good state of affairs is not to be found when the model is misspecified. Consider for example the case where the mean has a small trend component; i.e. suppose that the true unknown DGP takes the form $X_t \sim N(\beta t, 1)$, while the statistical model still assumes the iid data with $X_t \sim N(\mu, 1)$. In this case, the MLE no longer provides the most accurate forecasts, and most importantly, it is no longer optimal to make use of the entire sample of data $X_1, ..., X_T$. This interesting situation occurs because the MLE attempts to approximate the distribution of the entire sample $X_s, ..., X_T$ by minimizing the empirical Kullback-Leibler divergence w.r.t. the true measure. To do so, the MLE becomes centered at the value $\beta T/2$ when we use the complete sample ($s = 1$). However, since $\mathbb{E}(X_{T+1}) = \beta(T + 1)$, this is clearly not an interesting estimator when we are interested in forecasting accuracy rather than fitting the distribution of past observations. An MLE that takes into account only the last few sample values can greatly improve the forecast MSE. The optimal value of $1 \leq s < T$ finds a good compromise between the forecast bias (which decreases with s) and the forecast variance (which increases with s). For concreteness, note that the forecast bias is strictly decreasing in s since

$$\text{Bias}(\hat{X}_{T+n}) = \mathbb{E}\hat{\mu}_{s:T} - \mathbb{E}X_{T+n} = \frac{1}{T-s+1}\sum_{t=s}^{T}\beta t - \beta(T+n) = \beta n + \frac{\beta}{2}(T-s).$$

Minimizing the forecast bias would thus be obtained by setting $s = T$ and hence making use of only the last observation X_T in the sample. However, this strategy is not optimal

in terms of the forecast's MSE since the variance of the forecast produced by this MLE would be too large

$$\mathbb{V}\text{ar}(\hat{X}_{T+n}) = \mathbb{V}\text{ar}(\hat{\mu}_{s:T}) = \frac{1}{(T-s+1)^2}\sum_{t=s}^{T}\mathbb{V}\text{ar}(X_t) = \frac{1}{(T-s+1)}.$$

In order to minimize the MSE, it is instead better to make use of the information contained in past data values, even if this comes at the cost of a larger forecast bias. This trade-off is neatly found in the MSE expression

$$\text{MSE}(\hat{X}_{T+n}) = \Big(\text{Bias}(\hat{X}_{T+n})\Big)^2 + \mathbb{V}\text{ar}(\hat{X}_{T+n}) = \frac{1}{4}\beta^2(2n+T-s)^2 + (T-s+1)^{-1}.$$

Figure (3.1) below plots the 1-step-ahead forecast MSE as a function of s for the case $T = 100$ and $\beta = 0.02$. The figure shows that instead of using the entire sample $X_1, ..., X_{100}$, it is optimal to use only the last 16 observations $X_{84}, ..., X_{100}$.

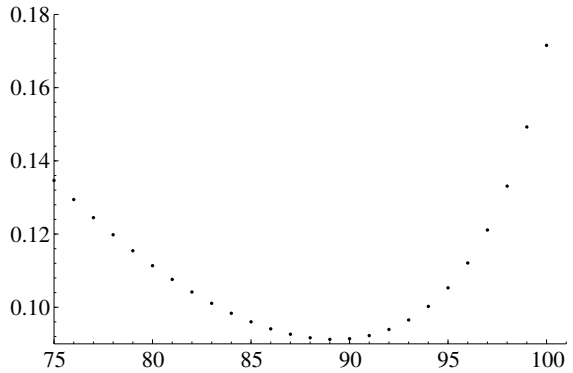

Figure 3.1: 1-step-ahead forecast MSE as function of s for $T = 100$ and $\beta = 0.02$.

The estimator $\hat{\mu}_{s:T}$ proposed above is essentially an MLE that gives zero weight to the likelihood contributions from data before s, and unit weight to the contributions of the remained observations. In essence, it is a WMLE of the form

$$\hat{\mu}_T = \arg\max \sum_{t=1}^{T} \ell_t(X_t;\mu)\cdot w_t$$

where $\ell_t(X_t;\mu)$ denotes the log likelihood contribution of the X_t observation and w_t is a weighting function

$$w_t = \begin{cases} 0 & 1 \le t < s \\ 1 & s \le t \le T \end{cases}.$$

This weighting function is however unnecessarily restrictive, and it can be easily improved upon, at least in terms of reducing the forecast MSE. Consider for example the case where

$w_t = \rho^{T-t}$ for some $0 < \rho \leq 1$. This weighting function gives unit weight to the last observation X_T and possibly decaying weight to past observations. In the special case where $\rho = 1$, we obtain the usual MLE. For general decay parameter ρ the WMLE takes the form

$$\hat{\mu}_T = \frac{\sum_{t=s}^{T} \rho^{T-t} X_t}{\sum_{t=s}^{T} \rho^{T-t}}$$

which boils down to the usual MLE when $\rho = 1$. It is also easy to show that

$$\text{Bias}(\hat{X}_{T+n}) = \mathbb{E}\frac{\sum_{t=s}^{T} \rho^{T-t}\beta t}{\sum_{t=s}^{T} \rho^{T-t}} - \beta(T+1) \quad \text{and} \quad \mathbb{V}\text{ar}(\hat{X}_{T+n}) = \frac{\sum_{t=s}^{T} \rho^{2(T-t)}}{\left(\sum_{t=s}^{T} \rho^{T-t}\right)^2}$$

so that

$$\text{MSE}(\hat{X}_{T+n}) = \frac{\beta^2 \left(\rho^{T+1}((\rho-1)T + \rho + s - s\rho - 2) + \rho^s\right)^2}{(\rho-1)^2 \left(\rho^s - \rho^{T+1}\right)^2} + \frac{(\rho-1)\left(\rho^{T+1} + \rho^s\right)}{(\rho+1)\left(\rho^{T+1} - \rho^s\right)}.$$

Figure (3.2) plots the forecast MSE of the WMLE as a function of the parameter ρ, as well as the optimal weight function. The plot shows that the WMLE with $\rho \approx 0.89$ considerably outperforms the standard MLE with uniform weight ($\rho = 1$).

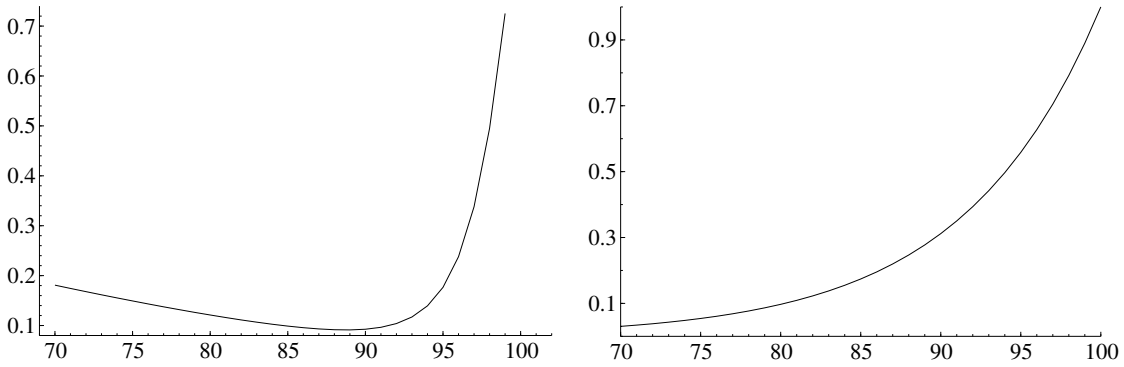

Figure 3.2: 1-step-ahead forecast MSE as function of ρ for $T = 100$ and $\beta = 0.02$ (left), and optimal weight function $w_t = 0.89^{T-t}$ (right).

These results are obviously not exclusive to the linear trending mean considered above. On the contrary, any time-variation in the true mean that is not appropriately captured by the statistical model will result in a non-constant optimal weighting function that improves the forecasting accuracy of the MLE. In fact, the WMLE opens the possibility for much more complex weighting functions than the simple exponential decay presented above. These weights can optimally adapt to breaks in the DGP, cyclical fluctuations and others.

3.3 WMLE for Autoregressive Models

We now turn our attention to the linear autoregressive (AR) model. This is arguably the most widely used class of models in time-series analysis and forecasting. The AR(p) model assumes that the time-series $\{X_t\}_{t\in\mathbb{Z}}$ is generated according to

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + \epsilon_t \quad , \quad t \in \mathbb{Z} \tag{3.1}$$

where $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of iid innovations with density $p_\epsilon(\boldsymbol{\lambda})$ indexed by the vector of parameters $\boldsymbol{\lambda}$. For simplicity, we collect all parameters in the vector $\boldsymbol{\theta} := (\boldsymbol{\alpha}, \boldsymbol{\lambda})$ where $\boldsymbol{\alpha} := (\alpha_0, ..., \alpha_p)$.

AR(p) models are widely used and appreciated for their simplicity. Even simple AR(1) models can often outperform larger and more complex models. Furthermore, the parameters of AR(p) models are easy and fast to estimate, and these models produce highly tractable autocorrelation functions, forecast functions, forecast confidence bounds and impulse response functions.

If the AR(p) model is well specified, then it is easy to show that the classical MLE produces optimal forecasts in Kullback-Leibler divergence. In the case of the Gaussian AR(p) the forecasts minimize also the mean squared error loss function. The fundamental reason behind this result is that the in-sample fit and out-of-sample forecasting performance are tightly 'linked' when the model is well specified.

Unfortunately, if the model is misspecified, then achieving a good in-sample fit over the entire data set, becomes a separate problem from that of achieving good out-of-sample forecasting accuracy. As we shall see, Proposition 3.3 in Section 3.4 shows that for every misspecified AR model, there exists a non-constant sequence of parameters that improves the approximation of the AR model to the true data-generating process as well as its forecasting accuracy. Furthermore, we will show that the WMLE can help us find such an appropriate sequence of parameters. In particular, we show that the WMLE outperforms the classical MLE in terms of forecasting accuracy.

3.3.1 The WMLE

Let $\mathbf{w} \in \mathbb{R}_+^T$ denote a weight vector with elements $(w_1, ..., w_T)$ that define the weight of the log-likelihood contribution at time $t = 1, ..., T$. For any given $\mathbf{w}$, the WMLE $\hat{\boldsymbol{\theta}}(\mathbf{w})$ for the vector $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}}(\mathbf{w}) := \arg\max_{\boldsymbol{\theta}\in\Theta} \frac{1}{T} \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}) \cdot w_t \tag{3.2}$$

where $\ell_i(\boldsymbol{\theta})$ denotes the logarithm of the conditional density of X_i given $X_{i-1}, X_{i-2}, ..., X_1$,

$$\ell_i(\boldsymbol{\theta}) := \log p_\epsilon\big(e_i(\boldsymbol{\alpha}); \boldsymbol{\lambda}\big)$$

where $e_t(\boldsymbol{\alpha})$ denotes the residual term

$$e_t(\boldsymbol{\alpha}) := X_t - \alpha_0 - \alpha_1 X_{t-1} - ... - \alpha_p X_{t-p} \ , \quad t = p+1, ..., T.$$

The fundamental difference between the WMLE and the classical MLE is thus the introduction of weights for the log likelihood of different observations. Observations with relatively large (small) weight will have a relatively larger (smaller) influence in the estimation of the parameter vector $\boldsymbol{\theta}$. As such, the WML parameter estimates will attempt to fit more accurately the dynamics of the time-series at certain periods of interest. Since the weight vector $\mathbf{w}$ determines the parameter estimates, it is useful to note that the WMLE is a map $\hat{\boldsymbol{\theta}} : \Omega \times \mathbb{R}_+^T \to \Theta$, where Ω is the event space of the underlying probability space of interest. Similarly, given the vector $\mathbf{w}$, the random estimator maps elements of Ω to the parameter space Θ, i.e. $\hat{\boldsymbol{\theta}}(\mathbf{w}) : \Omega \to \Theta$.

In order to analyze the out-of-sample forecasting performance of the AR(p) under the WMLE estimates, it is important to define a class of weight vectors that use information only until a certain point in time. In particular, we let $\mathbf{w}_k$, $T' \leq k \leq T$ denote a weight vector whose elements are uniformly set to zero for all $t > k$

$$\mathbf{w}_k = (w_1, ..., w_k, 0, \cdots, 0).$$

Note that all weight vectors have size T. However, $\mathbf{w}_k$ effectively defines a WMLE $\hat{\boldsymbol{\theta}}(\mathbf{w}_k)$ that only makes of the shorter sample $X_1, ..., X_k$. Finally, we let $\mathcal{W}$ denote the matrix of all weights $\mathbf{W}$ so that each $\mathbf{w}_k$ corresponds to a row of $\mathbf{W}$ Throughout, we let the rows of $\mathbf{W}$ be normalized to sum up to one.

In general, the WMLE estimates will differ from one time period to the next. For example, the parameter estimate $\hat{\boldsymbol{\theta}}(\mathbf{w}_k)$ obtained using the sub-sample $X_1, ..., X_k$, will typically be different from the parameter estimate $\hat{\boldsymbol{\theta}}(\mathbf{w}_{k+1})$ obtained using the sub-sample $\mathbf{X}_1, ..., X_{k+1}$. For this reason, the WMLE can potentially be used to construct a sequence of parameter estimates

$$\hat{\boldsymbol{\theta}}(\mathbf{w}_k), \ \hat{\boldsymbol{\theta}}(\mathbf{w}_{k+1}), \ \hat{\boldsymbol{\theta}}(\mathbf{w}_{k+2}), \ ...$$

that describes the parameter instability in the AR(p).

In order to better understand the WMLE, it is useful to note how different weight sequences give origin to a number of different well known estimators. Below we highlight the relation between the WMLE and other well-known estimators.

3.3.2 Special Classes of Weights

The famous recursive least squares filter originally proposed by Gauss in 1821 in his work *'Theoria combinationis observationum erroribus minimis obnoxiae'* and rediscovered by Plackett (1950), consists of a sequence of estimators obtained recursively over a window of increasing length. Recursive ML estimators can be obtained in the WML setting by

defining a matrix of weights $\mathcal{W}$ with rows $\mathbf{w}_k$ having elements that satisfy $\mathbf{w}_{k,1} = \mathbf{w}_{k,2} = \cdots = w_{k,k} = 1$ for every k.

The popular rolling window estimators with window size h used for improved forecasting performance in Swanson (1998), Goyal and Welch (2003), Molodtsova and Papell (2009) and Inoue et al. (2014), among others, are obtained as a special case of the WMLE by letting the matrix of weights $\mathcal{W}$ with rows $\mathbf{w}_k$ having elements that satisfy $\mathbf{w}_{k,t} = 1$ if $k - t \leq l$ and $\mathbf{w}_{k,t} = 0$ otherwise, for some window length l and every k.

As we shall see below, rather of pre-determined, the elements of the weight vectors $\mathbf{w}_k$ can be estimated to provide optimal forecast performance for the AR(p). For simplicity, computational efficiency, and most importantly, statistical efficiency, it will often be beneficial to parameterize the weighting function by a small parameter vector $\boldsymbol{\rho}$. As such, instead of estimating all the elements of any given weight vector $\mathbf{w}_k$ individually, we can instead estimate a small vector $\boldsymbol{\rho}$ that determines the individual weights in a parametric fashion. For example, Exponentially decaying weights can be obtained with a scalar parameter ρ by setting $\mathbf{w}_{k,t}(\rho) = \rho^{(k-t)}$. This class of weights normalizes the weight of the most recent observation $t = k$ to unity $\mathbf{w}_{k,k} = 1$, and allows for data far in the past to be given a lower weight the estimation of $\boldsymbol{\theta}$. This class contains the MLE as a special case ($\rho = 0$), as well as a wide range of weighting decays ($0 < \rho \geq 1$).

Another interesting class of weights is obtained by letting the elements of $\mathbf{w}_k(\boldsymbol{\rho})$ depend on lagged values of X_t and/or other variables of interest Z_t. For example, in Section 3.5, we show that forecasts of US IPI during the latest global recession, can be significantly improved by defining weights that make past recession periods more informative, but also, downweight observations far in the past. In particular, we use the NBER recession indicator Z_t and define the weights of the vector $\mathbf{w}_k(\boldsymbol{\rho}, Z_t)$ as follows

$$\mathbf{w}_{k,t}(\boldsymbol{\rho}, Z_t) = \rho_1^{(k-t)}(1 + (\rho_2 - 1) \cdot Z_t)$$

where $\boldsymbol{\rho} = (\rho_1, \rho_2)$ with $0 \leq \rho_1 \leq 1$, $\rho_2 \geq 1$ and Z_t is the indicator.

3.3.3 Estimation of Optimal Weights by Cross-Validation

The recursive and rolling window estimation techniques mentioned in Section 3.3.2, can provide important insight into parameter instability, be it in the form of breaks, trends, seasonality or random changes. However, those methods are not designed from the outset to optimize the forecasting performance of the AR model. Instead, any improvements in forecast accuracy are the result of an ad-hoc improvement in model specification. Indeed, rolling window estimation will provide improved forecasts if it turns out that the exclusive use of recent data is advantageous for forecasting. Similarly, time-varying parameter models, will deliver improved forecast accuracy over their static parameter counterparts, if the specific weights on past data introduced indirectly by the parameter updating equations happen to improve the forecasting ability of the model. In this section

we propose a cross-validation method for finding the matrix of weights $\mathbf{W}$ that optimizes the forecasting performance of our AR(p) model. There are various popular choices of criterion functions that reflect forecasting accuracy. The mean absolute forecast error (MAFE), the mean squared forecast error (MSFE), the out-of-sample log likelihood, etc. If the MSFE is chosen then we say that the matrix of weights $\mathcal{W}$ is the best WMLE in an n-step-ahead MSFE sense if

$$\hat{\mathbf{W}} = \arg\min_{\mathcal{W}} \frac{1}{T - T' - n} \sum_{k=T'}^{T-n} \Big(\hat{X}_{k+n}\big(\hat{\boldsymbol{\theta}}(\mathbf{w}_k)\big) - X_{k+n} \Big)^2 \tag{3.3}$$

where, as noted before, the WMLE $\hat{\boldsymbol{\theta}}$ maps events $\omega \in \Omega$ and weight vectors $\mathbf{w} \in \mathbb{R}_+^T$ to the parameter space, i.e. $\hat{\boldsymbol{\theta}} : \Omega \times \mathbb{R}_+^T \to \Theta$, and T' defines the sample point from which the forecasting accuracy begins to be measured. A large T' gives us more data to estimate the parameter vector $\boldsymbol{\theta}$ by WML, but a small number of observations to evaluated the forecasting accuracy of the model and optimize the weights. On the contrary, a small T' increases the uncertainty in the estimation of $\boldsymbol{\theta}$ but gives us a larger sample to determine the optimal weights. In Section 3.4 we will show how to determine the optimal T'. When the matrix $\mathbf{W}$ is parameterized by a vector $\boldsymbol{\rho}$, then the optimal weights are defined as $\mathbf{W}(\hat{\boldsymbol{\rho}})$ where

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} \frac{1}{T - T' - n} \sum_{k=T'}^{T-n} \Big(\hat{X}_{k+n}\big(\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\boldsymbol{\rho}))\big) - X_{k+n} \Big)^2. \tag{3.4}$$

At this point, it is important to note that the optimization of the weights in either (3.3) or (3.4) relies on the relation between weight vectors $\mathbf{w}_k(\hat{\boldsymbol{\rho}})$ and parameter estimates $\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\hat{\boldsymbol{\rho}}))$. The optimization would be trivially simple if, given a sample of data, the mapping $\hat{\boldsymbol{\theta}} : \mathbb{R}_+^T \to \Theta$ from weight vectors in $\mathbb{R}_+^T$ to point estimates in Θ were known analytically. In general however, this map is unknown analytically. As a result we much approach the optimization numerically.

Below we detail the recursive algorithm designed to estimate the optimal weights. We start with a uniform unit weighting function $w_t(\boldsymbol{\rho}) = 1 \ \forall \ t$ and optimize the likelihood to obtain the standard MLE. Then we optimize the weights using (3.4) Conditional on the new $\hat{\rho}$, we can obtained a new WMLE. We iterate until convergence. Below $Q_n(\boldsymbol{\rho})$ denotes the n-step-ahead forecasting performance criterion chosen to optimize the weights, of which the MSFE in (3.4) is a special case.

Algorithm 1.

---

1 Set $\boldsymbol{\rho}_1$ such that $\mathbf{w}_{k,t}(\boldsymbol{\rho}_1) = 1 \ \forall \ t \leq k$, and
obtain the ML estimates $\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\boldsymbol{\rho}_1))$, $k = T', ..., T - n$.

2 Given the ML estimates $\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\hat{\boldsymbol{\rho}}_1))$, $k = T', ..., T-n$, obtain the best weights $\hat{\boldsymbol{\rho}}_2$ by minimizing $Q_n(\boldsymbol{\rho})$.

3 For $j \geq 2$

If $Q_n(\hat{\boldsymbol{\rho}}_{j+1}) < Q_n(\hat{\boldsymbol{\rho}}_j)$:

3.1. Given $\boldsymbol{\rho}_j$ obtain $\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\boldsymbol{\rho}_1))$, $k = T', ..., T-n$.

3.2. Given the ML estimates $\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\hat{\boldsymbol{\rho}}_1))$, $k = T', ..., T-n$, obtain the best weights $\hat{\boldsymbol{\rho}}_2$ by minimizing $Q_n(\boldsymbol{\rho})$.

3.3. Repeat step 3 with $j = j+1$.

If $Q_n(\hat{\boldsymbol{\rho}}_{j+1}) \geq Q_n(\hat{\boldsymbol{\rho}}_j)$:

3.4 Collect $(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\rho}}_j), \hat{\boldsymbol{\rho}}_j)$ and stop iterating.

This simple steepest-ascent algorithm has revealed itself to be fast and stable in both Monte Carlo simulations and applications. Since the algorithm is initiated at the classical ML estimates, the WMLE will feature non-uniform weights and differ from the MLE only when there is space for improvement over the MLE.

3.4 Theoretical Foundations for the Weighted MLE

The cross-validation optimization described in the previous section is intuitively appealing as it leads to weights that improve not the in-sample fit of the AR(p) model but instead, its out-of-sample forecasting performance. Below we provide theoretical foundations for this procedure.

First, we analyze the WMLE as a generalization of the classical MLE. In particular, we show that if the AR(p) model is well specified, then the WMLE is asymptotically equivalent to the MLE, and hence uncovers the true parameter vector and minimizes forecast errors. On the other hand, we also show that, if the model is misspecified, then there exist non-uniform weights that improve upon the MLE parameter estimates in terms of forecasting performance. Furthermore, we show that our algorithm for finding optimal weights delivers that WMLE outperforms the MLE under very general conditions.

Second, we give conditions under which our cross-validation procedure delivers a WMLE that provide optimal forecasting performance. We show that these results apply to a wide range of forecasting performance criteria.

Third, we implement a Diebold-Mariano test that can be used to infer whether the improvements in forecasting accuracy from a change in weights are statistically significant or not. We analyze the validity of the asymptotic distribution of the statistic and study its finite sample size and power in a Monte Carlo exercise.

As we have seen before, the WMLE can be used to describe instability in the parameters of AR(p) models. Indeed, we noted that the WMLE contains recursive estimators as well as rolling window estimators as special cases. Proposition 3.4, shows that many

DGPs can be written in the form of a time-varying parameter AR(p) model with Gaussian innovations. For simplicity, we restrict our attention to a DGP with stochastic contracting dynamics. This allows us to apply laws of large numbers and central limit theorems.

The contracting dynamics are by no means necessary for the theory that follows, but they are sufficient. As such, these results can be extended to a host of other settings that allow for heterogeneous dynamics, deterministic components, etc. Apart from the general contracting behavior, Proposition A.1 still allows for a DGP with very general dynamics. Indeed, X_t can depend nonlinearly on its past, as well as on a potentially very large vector V_t of variables that may include not only innovations and random breaks, but also a wide range of exogenous strictly stationary and ergodic (SE) variables with complex dynamics and temporal dependence patterns.

The SE nature and bounded moments of the process $\{X_t\}_{t\in\mathbb{Z}}$ are obtained using Theorem 3.1 of Bougerol (1993) and Theorem A10.1 of Blasques et al. (2014b). The SE nature of the time-varying AR(p) parameters is a direct consequence of Krengel's Theorem. Below, we let $\mathbb{C}^1(\mathcal{X}\times\mathcal{V})$ denote the space of real-valued continuously differentiable functions defined on the set $\mathcal{X}\times\mathcal{V}$. Similarly, $\mathbb{L}(\mathcal{X})$ denotes the space of Lipschitz continuous functions defined on $\mathcal{X}$, and $\mathbb{L}_V(\mathcal{X}) := \{\mathbb{L}_v(\mathcal{X}), v\in\mathcal{V}\}$ denotes the class of functions that are Lipschitz on $\mathcal{X}$ uniformly over $v\in V$.

Proposition 3.1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ generated according to*

$$X_t = \phi(X_{t-1}, V_t) \quad , \quad t\in\mathbb{Z} \tag{3.5}$$

where
(i) $\{V_t\}_{t\in\mathbb{Z}}$ is an SE n_V-variate stochastic sequence
(ii) $\phi\in\mathbb{C}^1(\mathcal{X}\times\mathcal{V})$ and $\phi\in\mathbb{L}_V(\mathcal{X})$;
(iii) $\mathbb{E}|\phi(x,V_t)|^4<\infty$ for some $x\in\mathcal{X}$; and
(iv) $\mathbb{E}\sup_{x\in\mathcal{X}}|\phi'_x(x,Z_t)|^4<1$.
Then the following time-varying AR(p) representation holds

$$X_t = \alpha_{0,t} + \alpha_{1,t}X_{t-1} + ... + \alpha_{p,t}X_{t-p} + \epsilon_t \quad , \quad \epsilon_t\sim N(0,\sigma^2_\epsilon) \quad , \quad t\in\mathbb{Z} \tag{3.6}$$

where $\{\alpha_{i,t}\}_{t\in\mathbb{Z}}$ is SE for every $i=1,...,p$, and furthermore $\{X_t\}_{t\in\mathbb{Z}}$ is also SE and has two bounded moments $\mathbb{E}|X_t|^4<\infty$.

One distinct feature of the WMLE estimator is the fact that it reduces to the MLE when the weights are unnecessary or undesirable. Proposition 3.2 shows precisely the weights converge in probability to unity when the model is well specified. Specifically, the cross-validation method that we propose for estimating the weights ensures that the weighted likelihood function converges in probability to the classical likelihood function as the size of the estimation sample $S = T'-p$ and cross-validation sample $H := T-T'-n$ diverge to infinity sequentially. The Monte Carlo exercise in Section 3.5, confirms that the weights remain close to unity even in finite samples of empirically relevant size.

Proposition 3.2. *Let $\mathcal{W}$ be compact and suppose the conditions of Proposition 3.1 hold with $V_t = \epsilon_t \ \forall \ t \in \mathbb{Z}$ and*

$$\phi(X_{t-1}, \epsilon_t) = \alpha_0 + \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + \epsilon_t \ \forall \ t \in \mathbb{Z}.$$

Then the MSFE criterion in (3.3) ensures that $\hat{\mathbf{w}}_{k,t} \xrightarrow{p} 1 \ \forall \ (k,t)$, as $S \to \infty$ and $H \to \infty$ sequentially, for any given forecasting horizon $n \geq 1$ and lag order $p \geq 1$.

By application of Berge's Maximum Theorem , we obtain as a corollary that the WMLE converges in probability to the MLE as the cross-validation sample H diverges to infinity. Naturally, when both the cross-validation sample H and the estimation sample S diverge to infinity, then the WMLE converges to the true parameter $\boldsymbol{\theta}_0 \in \Theta$, just as the MLE does. The Monte Carlo exercise in Section 3.5, reveals that the WMLE performs well also in finite samples.

Corollary 3.1. *Let the conditions of Proposition 3.2 hold. Then $\|\hat{\boldsymbol{\theta}}(\mathbf{w_k}) - \hat{\boldsymbol{\theta}}(1)\| \xrightarrow{p} 0$ as $H \to \infty$ and $\hat{\boldsymbol{\theta}}(\mathbf{w}_k) \xrightarrow{p} \boldsymbol{\theta}_0$ as $S \to \infty$ for $k = T', ..., T$ and any given $n \geq 1$.*

Under incorrect model specification, recursive or rolling-window estimators, can often improve upon full-sample estimators by allowing for time-varying parameters that better capture the dynamics of the data at any given period of time. Similarly, the WMLE will be able to improve the forecasting performance of the AR(p) by allowing for time-varying parameters that can improve the out-of-sample performance of the model. The existence of such a sequence of parameters is another simple, albeit important and general, consequence of Proposition 3.1. Below we let $\text{MSFE}_n(\boldsymbol{\theta})$ denote the n-step ahead mean squared error achieved by the AR(p) model under some parameter vector $\boldsymbol{\theta} \in \Theta$,

$$\text{MSFE}_n(\boldsymbol{\theta}) = \mathbb{E}_t\Big(X_{t+n} - \hat{X}_{t+n}(\boldsymbol{\theta})\Big).$$

Corollary 3.2. *Let the conditions of Proposition 3.1 hold, and suppose that*

$$\phi(X_{t-1}, V_t) \neq \alpha_0 + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \epsilon_t \quad , \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \ ,$$

for every $\boldsymbol{\theta} \in \Theta$ and some $t \in \mathbb{Z}$. Then there exists a non-constant sequence $\{\boldsymbol{\theta}_t\}_{t \in \mathbb{Z}}$ of points in Θ such that $\text{MSFE}_n(\boldsymbol{\theta}_t) < \text{MSFE}_n(\boldsymbol{\theta})$ for any given $\boldsymbol{\theta} \in \Theta$ and $n \geq 1$.

Corollary 3.2 highlights that time-varying parameters can improve the forecasting of the AR(p) when the model is a simplistic representation of the data. Corollary 3.1 revealed that the WMLE will only deliver time-varying parameters in large samples, when the model is well specified.

Proposition 3.3 now focuses on the properties of the WMLE algorithm proposed in the previous section. First, it highlights that the algorithm is designed to ensure that the WMLE outperforms (or is at least as good as) the MLE in terms of the forecasting

accuracy of the AR(p) model in the cross-validation sample. Furthermore, Proposition 3.3 shows that under appropriate regularity conditions, the WMLE algorithm will actually uncover the weights that optimize the forecasting performance of the AR(p) model in the cross-validation sample. Once again, we let $\hat{\boldsymbol{\rho}}_j$ denote the j-th iteration weights and $Q_n(\boldsymbol{\rho})$ denote the mean squared error in the cross-validation sample obtained under $\boldsymbol{\rho}$

$$Q_n(\boldsymbol{\rho}) := \frac{1}{T - T' - n} \sum_{k=T'}^{T-n} \Big(\hat{X}_{k+n}\big(\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\boldsymbol{\rho}))\big) - X_{k+n} \Big)^2.$$

The Monte Carlo exercise in Section 3.5, as well as the application to US IPI data in Section 3.6, reveals that the WMLE is indeed capable of significantly improving the forecasting performance of the AR(p) model.

Proposition 3.3. *For any given realized sample $\{x_t\}_{t=1}^T$, Algorithm 1 ensures that*

$$Q_n(\boldsymbol{\rho}_{j+1}) \leq Q_n(\boldsymbol{\rho}_j) \quad \forall\ j \geq 1$$

and hence the WMLE outperforms the MLE under the Q_n criterion.
If furthermore it holds that

$$\sup_T \sup_{\boldsymbol{\rho}} \left| \partial \hat{\boldsymbol{\theta}}(\boldsymbol{\rho}) / \partial \boldsymbol{\rho} \right| < 1 \quad \textit{and} \quad \sup_T \sup_{\boldsymbol{\theta}} \left| \partial \hat{\boldsymbol{\rho}}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \right| < 1.$$

Then $\boldsymbol{\rho}_j \to \boldsymbol{\rho}^*$ and $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}^*$ as $j \to \infty$, for any given $n \geq 1$.

The two main conditions of Proposition 3.3 ensure the contraction of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\rho}}$ as maps $\hat{\boldsymbol{\theta}} : \boldsymbol{\rho} \mapsto \boldsymbol{\theta}$ and $\hat{\boldsymbol{\rho}} : \boldsymbol{\theta} \mapsto \boldsymbol{\rho}$. Since these maps are not known analytically, the contracting behavior can only be verified numerically. This can be achieved by optimizing the derivatives stated above, and ensuring that there maximum is less than one.

The result established in Proposition 3.3 is important, but it ensures only that that the WMLE improves the finite sample FMSE. In other words, the algorithm discussed in Section 3.3.3 delivers weights that optimize the forecasting performance within the cross-validation sample. However, due to sampling error it is impossible to ensure that the true forecasting performance has improved from the MLE to some WMLE with weighting matrix $\mathbf{W}$. As we shall see, a simple Diebold-Mariano (DM) test statistic that can help us asses whether the improvement in forecasting performance is statistically significant or not. Proposition 3.4 highlights the validity of the asymptotic distribution derived byDiebold and Mariano (1995b) under the assumptions of Proposition 3.1 when the null hypothesis compares the MLE against an alternative WMLE.

Below we let MSFE($\mathbf{W}$) denote the MSFE achieved by the AR(p) model under the WMLE with weight matrix $\mathbf{W}$, and let $\mathbf{W}^*$ denote the best possible WMLE weight matrix for the AR(p) model

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} MSFE(\mathbf{W}).$$

Furthermore, we let MSFE(1) denote the MSFE achieved by the MLE. Under correct model specification we naturally have that MSFE(1)=MSFE($\mathbf{W}^*$).

Proposition 3.4 states a DM test with a null hypothesis of correct specification where both the MLE and WMLE provide equal forecasting accuracy

$$\mathrm{H}_0 : \mathrm{MSFE}(\mathbf{W}^*) = \mathrm{MSFE}(\mathbf{W}) = \mathrm{MSFE}(1)$$

against an alternative of incorrect specification where the WMLE provides improved forecasting accuracy over the MLE

$$\mathrm{H}_1 : \mathrm{MSFE}(\mathbf{W}) < \mathrm{MSFE}(1).$$

As noted by Diebold (2013), the question of whether the DM assumptions hold in practice is in any case an empirical issue for which there exist tests that one may wish to employ. The Monte Carlo exercise in Section 3.5 reveals that the finite sample distribution is well approximated by the standard normal asymptotic distribution. Below we let $\bar{d}_k(\mathbf{w}_k)$ and $\mathrm{Ste}(d_k(\mathbf{w}_k)$ denote the sample average and standard error of the MSFE difference of the WMLE w.r.t. the MLE at period k,

$$\bar{d}_k(\mathbf{w}_k) := u_k(1)^2 - u_k(\mathbf{w}_k)^2 \quad \text{where} \quad u_k(\mathbf{w}_k) := \big(\hat{X}_{k+n}\big(\hat{\boldsymbol{\theta}}(\mathbf{w}_k)\big) - X_{k+n}\big)^2.$$

Proposition 3.4. *Let the conditions of Proposition 3.4 hold. Then*

$$\bar{d}_k(\mathbf{w}_k, \mathbf{w}_k')/\mathrm{Ste}(d_k(\mathbf{w}_k, \mathbf{w}_k') \xrightarrow{d} N(0,1) \quad \textit{as} \quad T \to \infty$$

for any given pair (T', n) *under the null hypothesis* $\mathrm{H}_0 : \mathrm{MSFE}(\mathbf{W}^*) = \mathrm{MSFE}(\mathbf{W}) = \mathrm{MSFE}(1)$*, and*

$$\bar{d}_k(\mathbf{w}_k)/\mathrm{Ste}(d_k(\mathbf{w}_k) \to \infty \quad \textit{as} \quad T \to \infty$$

under the alternative hypothesis $\mathrm{H}_1 : \mathrm{MSFE}(\mathbf{W}) < \mathrm{MSFE}(1)$*.*

We note that the DM test is the natural tool for comparing the forecasts produced under two WMLE estimates. As noted in Giacomini and White (2006a) in recent years, the bulk of the literature has moved from *testing forecasts* to *testing models*, evaluated at their pseudo-true parameters (see e.g. West (1996a), and Clark and McCracken (2001a)). Clearly, here we are not interested in testing different models, as all the forecasts come from the AR(p). Instead, we are interested at testing the forecasts obtained under the different parameter estimates produced by the WMLE and MLE.

3.5 Simulation Experiment

In this section, we investigate the finite sample performance of the WML estimation in the context of AR(1) model using Monte Carlo simulation. We consider four different data

generating processes(DGP) for the time series y_t. Among these four different AR(1) process, the autoregressive coefficient could be time-invariant, time-varying, regime switching or subject to a structure break. We aim to investigate that in which cases the optimal weighting function can improve the forecasting accuracy of the MLE. We concentrate on 1-step ahead forecasts based on rolling-window method with window length k. Although we consider four different data generating processes(DGPs) in the simulation experiment, the WML parameters are estimated based on the ordinary AR(1) model:

$$y_t = \alpha + \beta y_{t-1} + \epsilon_t, \tag{3.7}$$

where α, β are coefficients and ϵ_ts are i.i.d normally distributed with mean zero and variance σ_ϵ^2.

We consider two kinds of weighting functions which we apply in Monte Carlo experiment: the exponential weighting function ($w_t = \rho_1^{k-t}$, where $\rho_\in[0,1]$) and the binary weighting function with decay ($w_t = \rho_1^{(k-t)}(1+(\rho_2-1)\cdot Z_t)$, where Z_t is the predetermined indicator for recession period). The exponential weighting function is applied in the first three experiments; while the binary weighting function with decay is applied in the last experiment.

3.5.1 Experiment 1: Time-Invariant AR(1) model

In the first experiment, the data are generated by an AR(1) model with time-invariant parameters and the WML parameter ,ρ_1, is estimated based on the Equation (3.7). The data generation process (DGP) is specified by:

$$\text{DGP: } y_t = \alpha + \beta y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{3.8}$$

where $\alpha = 0.13$, $\beta = 0.5$ and ϵ_ts are i.i.d distributed with variance $\sigma_\epsilon^2 = 0.5$. The exponential weighting function is considered when applying WMLE algorithm and we use Equation (3.7) with the estimates, $\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\hat{\boldsymbol{\rho}}))$, from the WMLE to calculate the forecasts.

In Experiment 1, we generate a time-series by the AR(1) model. The rolling-window length, k, is selected to be 120 and forecasts are made for time period $t = 701 \ldots 760$. The generated data are considered as monthly data. For each simulated data set, we compute the WML estimates and use these estimates to calculate the 1-step-ahead forecasts. Since the model is accurately specified we expect that the estimated WML parameter, ρ_1, is close to 1, which means the original AR(1) model can already provide accurate forecasts.

The left panel of Figure (3.3) shows the simulation density of WML parameter, ρ_1, in Experiment 1. The simulated parameters peak at 1. To be precise, the simulated parameter mean is 0.9953 and the median is 1. In the set up, we restrict on our weighting parameter between 0 and 1, so in this case we shall put more attention on the median of the simulation results rather than the mean. The feature of the simulated distribution indicates that the WML methodology can hardly improve forecasting accuracy. Such

finding is consistent with Proposition 3.2 in Section 3.4 which shows the weights converge in probability to unity when the model is well specified. The right panel of Figure (3.3) shows the respective average sample weight and its 95% confidence bounds.

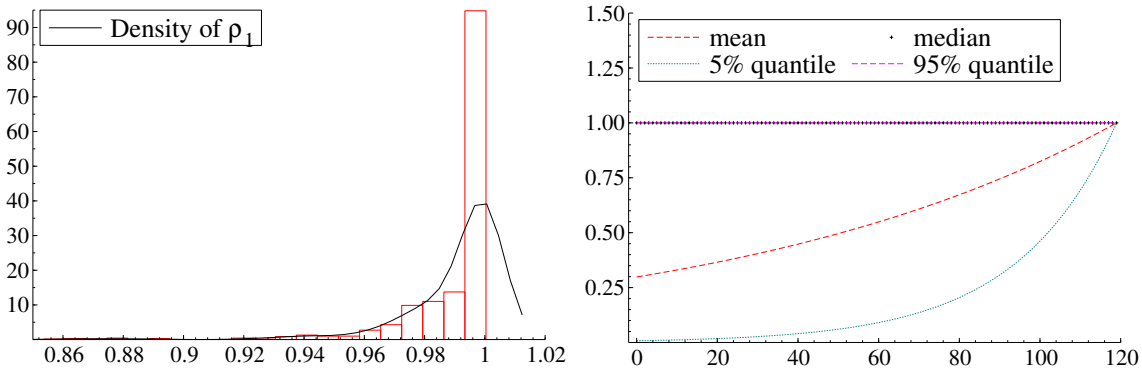

Figure 3.3: Simulation results for Experiment 1. The data are generated by an AR(1) model and the WML parameter, ρ_1, is estimated based on Equation (3.7). Exponential weighting function is applied. The left panel presents the simulation density of WML parameter, ρ_1, over 1,000 simulations. The rights panel presents the average sample weights and its 90% confidence bound.

3.5.2 Experiment 2: Time-varying AR(1) model

In the second experiment, the time series are generated by an AR(1) model with time-varying coefficient, β, and the WML parameter, ρ_1, is estimated based on Equation (3.7). The data generation process (DGP) is specified by:

$$\text{DGP: } y_t = \alpha + \beta_t y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{3.9}$$

$$\beta_t = 0.5 + 0.5\sin(2\pi/B), \tag{3.10}$$

where $\alpha = 0.13$ and ϵ_ts are i.i.d distributed with variance $\sigma_\epsilon^2 = 0.5$. The coefficient, β is varying between 0 and 1 with respect to business cycle length $B = 72$. The exponential weighting function is considered when applying WMLE algorithm and we use Equation (3.7) with the estimates, $\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\hat{\boldsymbol{\rho}}))$, from the WMLE to calculate the forecasts.

In Experiment 2, we generate a time-series with size $T = 760$. The generated data are considered as monthly data. We consider a DGP which includes a six-year business cycle, which is the average business cycle length of U.S.. The rolling-window size is set to be 60 which is a bit shorter than the business cycle. An illustration of such DGP is shown in Figure (3.4). Forecasts are made for time period $t = 701 \ldots 760$. In this experiment, the

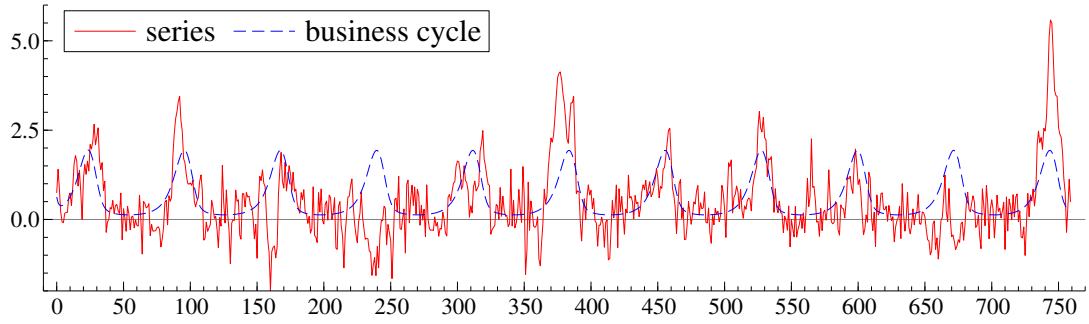

Figure 3.4: Illustration of DGP for Experiment 2.

time series implies an instability of the coefficients in the model and the original AR(1) model is misspecified, thus the WML parameter ρ_1 is expected to be smaller than 1. This means recent observations are more relevant to the forecasts in the future.

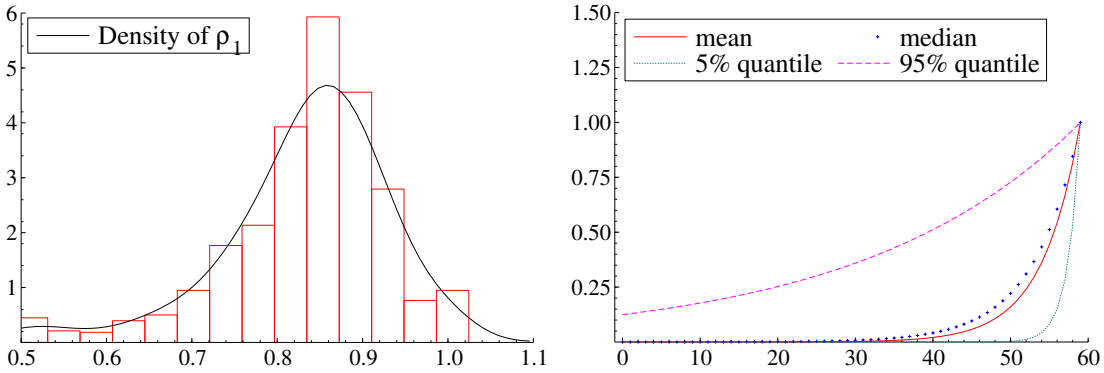

Figure 3.5: Simulation results of WML method over 1,000 simulations. The data are generated by an AR(1) model with time-varying coefficient, β, and the WML parameter, ρ_1, is estimated based on Equation (3.7). Exponential weighting function is applied. The left panel presents the simulated density of WML parameter, ρ_1. The right panel shows the average sample weights and its 90% confidence bound.

The left panel in Figure (3.5) presents the simulated density of WML parameter, ρ_1. The right panel shows the average sample weights and its 90% confidence bound. The right panel of Figure (3.5) shows that both the mean and the median of the simulated WML parameter, ρ_1, is smaller than 1. The general picture of Figure (3.5) is that when the time series contains certain time-varying components in it and the considered

forecasting model is misspecified, the MLE method can improve forecasting accuracy by putting more weights to the recent observation. Such finding is also consistent with Proposition 3.3 in Section 3.4.

3.5.3 Experiment 3: AR(1) Model with a Structure Break

In Experiment 3, the data are generated by an AR(1) model with a structure break in the coefficient, β, and the WML parameter, ρ_1, is estimated based on Equation (3.7). The data generation process (DGP) is specified by:

$$y_t = \alpha + \beta_t y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{3.11}$$

$$\beta_t = 0.2 + 0.7 I_t, \tag{3.12}$$

where $\alpha = 0.13$ and ϵ_ts are i.i.d distributed with variance $\sigma_\epsilon^2 = 0.5$. The indicator I_t is set to $I_t = 0$ for $t < 420$ and $I_t = 1$ for $t \geq 420$. The exponential weighting function is considered when applying WMLE algorithm and we use Equation (3.7) with the estimates, $\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\hat{\boldsymbol{\rho}}))$, from the WMLE to calculate the forecasts.

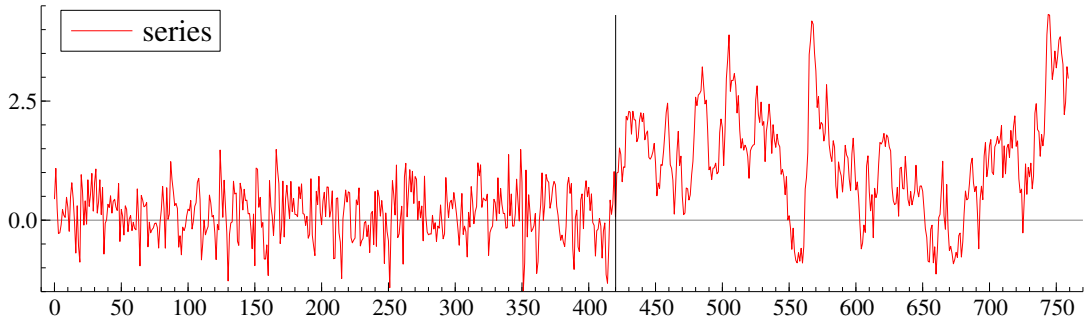

Figure 3.6: Illustration of DGP for Experiment 3.

In Experiment 3, we generate a time-series with size $T = 760$. The rolling-window length, τ, is selected to be 120. The break-point is set to $t = 420$. We evaluate the forecasting performance in three different forecasting periods . The first forecasting period is the period before the break-point (BB) starting from $t = 360$. The second one is the period right after the break-point (JAB) starting from $t = 444$ and the last one is the period long after (LAB) the break-point starting from $t = 492$. For all cases, we evaluate the forecasting performances of 60 observations. The generated data are considered as monthly data. This means, the length of forecasting period is 5 years. Figure (3.6) presents a realization of the data from Experiment 3.

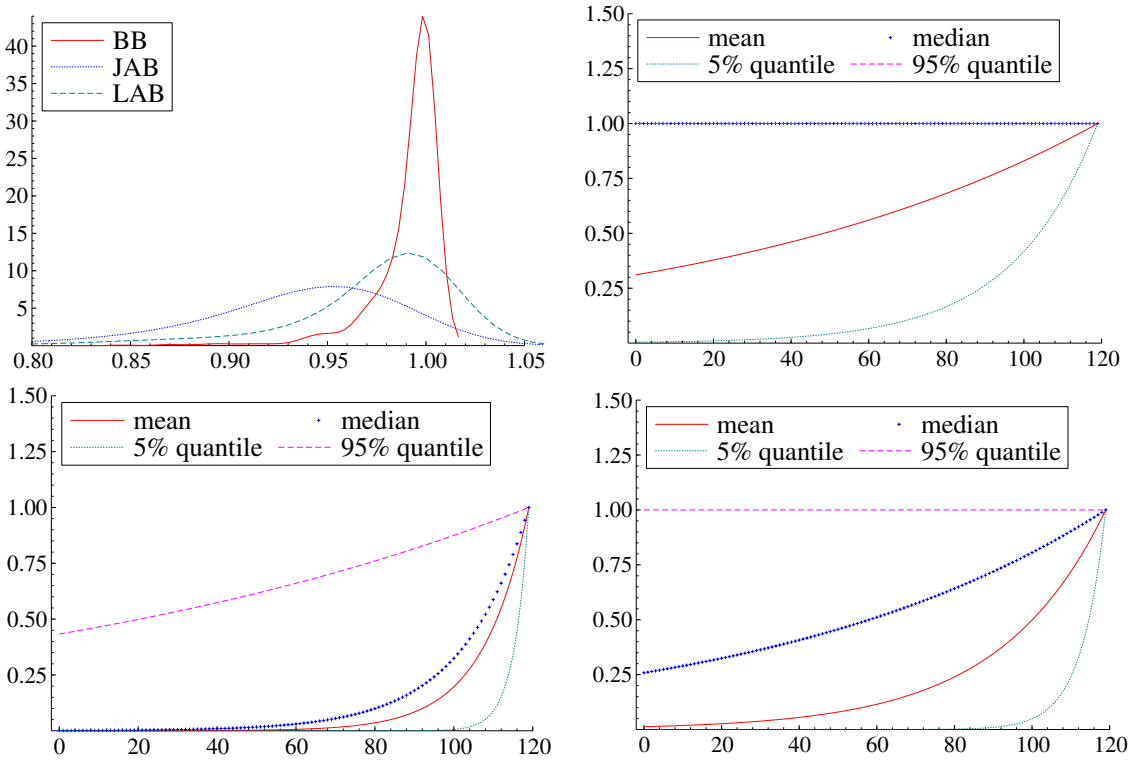

Figure 3.7: Simulation densities of WML parameter, ρ_1, over 1,000 simulations. The data are generated by an AR(1) model with a structure break in the coefficient, β, and the WML parameter, ρ_1, is estimated based on Equation (3.7). Exponential weighting function is applied. The upper-left panel presents the simulated densities for three forecasting periods: the period before the breaking point (BB), the period just after the breaking point (JAB) and the period long after the breaking point (LAB). The upper-right panel presents the average sample weights and its 90% bound for forecasting period before the breaking point. The bottom-left panel presents the average sample weights and its 90% bound for forecasting period right after the breaking point. The bottom-right panel presents the average sample weights and its 90% bound for forecasting period long after the breaking point.

When forecasts are calculated before the break point, the time series is generated by an ordinary AR(1) model, thus the weighted ML method converges to the classical ML method, the simulated WML parameter is expected to be 1. When forecasts are calculated just after the break point, the data before break point are less relevant for calculating future forecasts, thus the WML parameter, ρ_1, will significantly smaller than 1. As the forecasting point is getting far away from the break point, the WML parameter, ρ_1 moves generally back to 1 again.

Figure (3.7) presents the density of WML parameter, ρ_1, in Experiment 3. The WML method can hardly improve forecasting accuracy for the forecasting period before the breaking point because the forecasting model is well specified just as the finding of Experiment 1. For foresting period right after the structure break the simulated mean of WML parameter is 0.93485, which indicates more weight should be put on recent observation in order to provider better forecasts. Finally, for the period long after structure break, fewer observations before the breaking point are included in the estimation window and the WML parameter tends to peak at 1 again.

3.5.4 Experiment 4: regime-switching AR(1) model

In Experiment 4, the data are generated by an AR(1) model with a two-state regime-switching coefficient, $\beta.$, and the WML parameters are estimated based on Equation (3.7). The data generation process (DGP) is specified by:

$$y_t = \alpha + \beta_t y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{3.13}$$

$$\beta_t = 0.9 - 0.7 I_t, \tag{3.14}$$

where $\alpha = 0.13$ and ϵ_ts are i.i.d distributed with variance $\sigma_\epsilon^2 = 0.5$. I_t is predetermined recession indicator. A value of 1 indicates a recessionary period; while a value of 0 indicates an expansionary period. Model is estimated using the binary weighting function with decay.

In Experiment 4, we generate a time-series with size $T = 720$. The rolling-window length, τ, is selected to be 120. Forecasts are made for time period $t = 709 \ldots 720$. The generated data are considered as monthly data. The time series we generated follows a 6-year business cycle, where a 61-month expansion is followed by an 11-month recession. Notice that the forecasting period we considered is the last recession period in the simulated series. An illustration of the DGP is shown in Figure (3.8).

The time series is estimated for (3.7) using WMLE algorithm with binary weighting function and we use the estimates, $\hat{\boldsymbol{\theta}}(\mathbf{w}_k(\hat{\boldsymbol{\rho}}))$, from the WMLE to calculate the forecasts. Simulated densities of weighting parameter using binary weighting function with decay are present in Figure (3.9). The simulated binary parameter ρ_2 tends to have value larger than 1, which indicates that when calculate forecasting in recession period, more attention should be paid to the past recessions as discussed in the previous sections. Moreover,

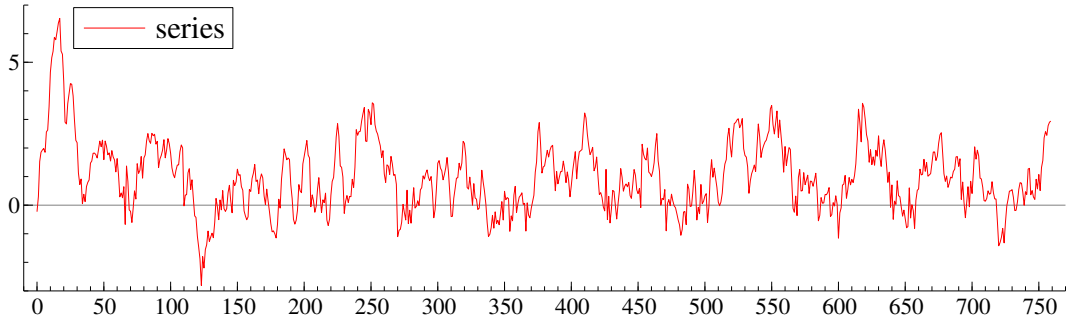

Figure 3.8: Illustration of DGP for Experiment 4.

parameter ρ_1 is peak near one. To be precise, the simulated parameter mean is 0.9781 and the median is 1. This result indicates that by using binary weighting function we do not observe strong evidence of decaying weights in the past observations.

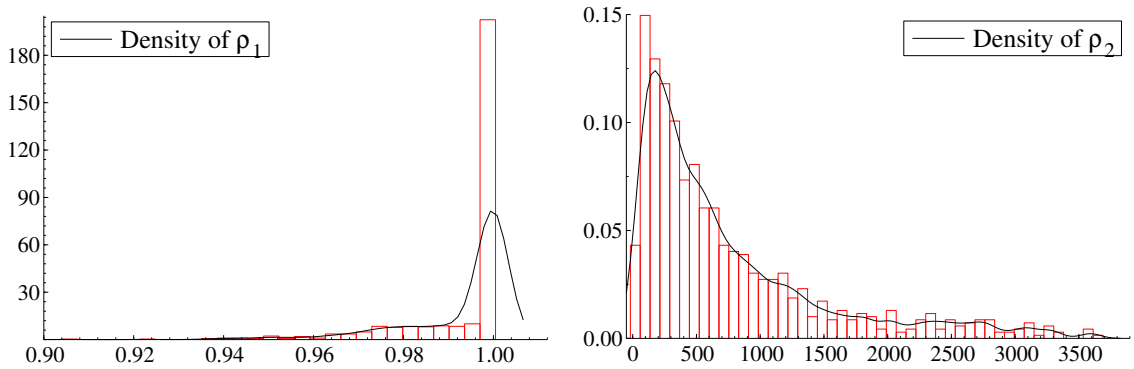

Figure 3.9: Simulation results of WML method using binary weighting function, over 1,000 simulations. The data are generated by an AR(1) model with a two-state regime switching coefficient, β. The WML parameters are estimated based on Equation (3.7). The upper-left panel presents the simulated density of parameter ρ_1. The upper-right panel presents the simulated density of parameter ρ_2. The bottom panel presents the median sample weights and its 90% bound for selected forecasting period.

3.6 Empirical Applications

In this section we study the performance of our WML estimator in empirically relevant applications. In section 3.6.1, we focus on forecasting the growth rate of the Industrial

Production Index (IPI) during the last global recession of 2008. We find that the WMLE can deliver significantly better forecasting performance than the classical MLE. This result is achieved by increasing the weight of observations coming from past recession periods. Moreover, we find that the improved out-of-sample forecasting accuracy delivered by the MLE is not driven by a single, or just a few, observations. Instead, the improved forecasting performance is present in general throughout the entire validation sample. In Sections 3.6.2 and 3.6.3 we show that the improved forecasting performance delivered by the WMLE is not restricted to forecasting the IPI during a recession period. In particular, Section 3.6.2 shows that the WMLE outperforms the MLE during expansion periods as well. Section 3.6.2 also illustrates that it is often important to allow for more complex weighting functions that allow for both ρ_1 and ρ_2 to be different from unity. Finally, Section 3.6.3 shows that the WMLE outperforms the MLE in other data sets as well. In effect, it reveals that the WML estimates can, in some cases, improve remarkably the forecasting performance of the model.

3.6.1 Forecasting IPI During the Global Recession

In this section we study whether the WML method can help to forecast the U.S. IPI. We focus on the IPI growth rate as it is a core indicator for the economy. We use the monthly data of the U.S. IPI. The sample period covers the period ranging from January 1950 until Dec 2009. We evaluate the forecasting accuracy and compare the results, in terms of root mean squared error (RMSE) and mean absolute error (MAE). We also apply the Diebold-Mariano (DM) test for comparing the predictive accuracy of the forecasting method.

In applications, we consider two windows of 25 years and 50 years respectively for the estimation of the parameters, and evaluate the forecasting accuracy on monthly ($h = 1$) and quarterly ($h = 4$) forecast horizons. We consider a weighting function which allows for a combination of binary weights and exponential decay. The binary weights can be used to emphasize the weight of past recessions in the estimation of parameters. The intuition behind this weighting scheme is that observations of the IPI coming from past recession periods may be more informative about the dynamics of the IPI during the global recession. The exponential weights can be used to weigh down observations far in the past. The intuition that underlies the exponential decay is straightforward: as the economy changes, observations far in the past are likely to be less informative about the current dynamics of the IPI than recent ones. Specifically, each weight w_t takes the form

$$w_t = \rho_1^{(k-t)}(1 + (\rho_2 - 1) \cdot Z_t)$$

where $Z_t = 1$ is the NBER recession indicator. Setting $\rho_1 = \rho_2 = 1$ delivers the classical unweighted MLE. Naturally, when $\rho_1 < 1$, past observations receive less weight than recent ones. Similarly, when $\rho_2 > 1$, observations from recession periods receive greater

estimation weight. Since ρ_1 and ρ_2 are estimated, the data will in some sense let us know how strong these effects should be.

For simplicity, we first set $\rho_2 = 1$ and apply the WML estimator with a binary weighting function to an AR(1) model and an AR(p) model where the number of lags p is selected by means of the Akaike's information criterion (AIC). We focus on out-of-sample forecasting of the monthly U.S IPI growth rate during the NBER global recession period which spans from January 2008 to June 2009. We emphasize that the sample where the forecasting accuracy was evaluated was not used for estimating the AR(1) parameters or the likelihood weights. Hence, the ML estimator can potentially outperform the WML estimator in this separate validation sample. Any improvement in forecasting accuracy should only to be expected if indeed the WMLE presents a real estimation advantage of the MLE.

The results are presented in Table (3.1). In this table we use the recession period of the year 2000 as our cross validation period to determine WML estimates that are used in the global recession of 2008. The column labeled Ratio shows the root mean-squared-forecast-error (RMSE) or mean-absolute-error (MAE) using WML method of the AR model relative to RMSE or MAE using normal ML method. The column labelled ρ_2 shows the WML parameter estimate for the binary weighting function. Entries in parentheses show the p-value of the DM test.

The RMSE and MAE ratios in Table (3.1). The WMLE performs better than the MLE when the ratios are smaller than 1. These entries are highlighted in gray. Furthermore, we indicate with *, **, and ***, the cases where the forecasting improvement is statistically significant at the 90%, 95%, and 99% confidence levels respectively. Clearly, Table (3.1) reveals that the WMLE strongly outperforms the MLE in the great majority of cases. The improvements of up to 31% in the forecasting accuracy of the AR(AIC) are quite remarkable. The p-values shown in brackets reveal that many of these improvements are also statistically significant at standard confidence levels. Most importantly, the parameter estimates of ρ_2 indicate that the improved performance comes precisely from giving greater weight to past recession periods. It is important to highlight that the improved forecasting performance is achieved not only for the simple AR(1) model, which is easily misspecified, but also, for the AR(AIC) which is flexible and includes several lags.

Curiously, the more reasonable estimates of ρ_2 lead to better results than the most extremes ones. For example, at a yearly forecasting horizon ($h = 4$), with an AR(AIC) model estimated on a rolling window of 50 years, the estimate of ρ_2 is such that an observation coming from a recession period receives approximately 9 times more weight than an observation coming from an expansion period. This means that one year of recession in the 2000, is given the same weight as the expansion in the previous 10 years. Most importantly, this relative weight leads to significant improvements in the RMSE as highlighted by the ratio of 0.778. In contrast, the estimate of 137.5 obtained for ρ_2 for the monthly forecast ($h = 1$) with a 25 year window, means essentially that only the

	RMSE				MAE			
	h=1		h=4		h=1		h=4	
	Ratio	ρ_2	Ratio	ρ_2	Ratio	ρ_2	Ratio	ρ_2
AR(1)								
25-year	1.0908	35.5	0.8297*	27.7	0.8310	31.5	0.7919*	27.1
	(0.3446)		(0.0679)		(0.1564)		(0.0500)	
50-year	0.9102**	15.7	0.7430**	12.3	0.8007***	13.8	0.5975***	11.1
	(0.0262)		(0.0241)		(0.0011)		(0.0000)	
AR(AIC)								
25-year	1.4909	137.5	0.9189	11.0	1.4039	228.5	0.9450	8.8
	(0.1460)		(0.2706)		(0.1785)		(0.3918)	
50-year	0.9094*	15.9	0.7782*	9.7	0.8427**	25.7	0.6955***	11.5
	(0.0667)		(0.0523)		(0.0153)		(0.0021)	

Table 3.1: WMLE forecasting results for monthly U.S IPI growth rate. The column labeled Ratio shows the RMSE or MAE using WML method of the AR model relative to RMSE or MAE using normal ML method. Entries in parentheses show the p-value of DM test. The forecasts were computed over the sample period 2008Jan-2009June.

recession periods matter. This extreme estimate of ρ_2, which produces WML estimates that essentially ignores data coming from expansion periods, turns out to behave poorly in the out-of-sample forecasting evaluation compared to the classical ML estimator.

The results for the weighting function that combines binary weights and exponential decay (i.e. for the case where both ρ_1 and ρ_2 are estimated) are made available in the Section 3.9 that accompanies this article. That table shows that, in this case, letting $\rho_1 \neq 1$ leads only to marginal improvements.

Figure (3.10) below plots the accumulated forecasting RMSE obtained from the ML and WML estimators. This figure shows that the improved performance of the WMLE is not driven by a single observation. On the contrary, it reveals that the WMLE achieves better forecasting accuracy by outperforming the MLE over the entire forecasting evaluation sample.

3.6.2 Using WMLE During the Expansion Period

In this section, we show that the improved forecasting performance achieved by the WMLE for the IPI growth rate is not restricted to the global financial recession period. On the contrary, the WMLE provides significant improvements in forecasting accuracy

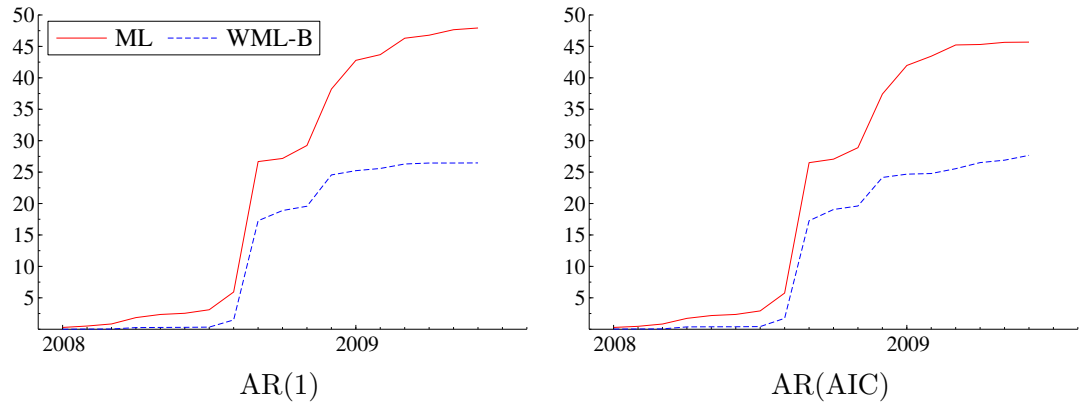

Figure 3.10: Accumulated root mean squared forecasting error from ML and WML method by using AR(1) and AR(AIC) models with a 50-year rolling window. The forecasts are made at forecasting horizon $h = 4$.

also during expansion periods. This is well illustrated in Table (3.2) below which takes the NBER expansion period spanning from January 2001 to December 2007 as the evaluation sample.

For simplicity, we focus only on the RMSE and a window size of 25 years. A table with the full results that include the MAE ratios and a window size of 50 years is available in the 3.9. The *Ratio* column of Table (3.2) shows that the WMLE delivers better out-of-sample forecasting RMSE obtained in the evaluation sample than the MLE. In two of these cases, the improvements seem to be marginally statistically significant at a 90% confidence level.

Note that, Z_t is now an *expansion indicator*. This means that the estimates of ρ_2 now reveal the binary weight associated to observations coming from expansion periods.

Table (3.2) reports improvements in forecasting accuracy of up to 13%. The magnitude of these improvements is not as expressive as in the recession case. In any case, we find statistically significant improvements as judged by the DM test. This number is especially relevant as it relates to the forecasts of the AR(AIC) model at a yearly horizon. Note also that it is quite natural for the improvements in forecasting accuracy to be smaller in the expansion case than in the recession case. This happens because the number of observations originating from expansion periods far outnumbers the observations coming from recession periods. As a result, the total weight that expansion periods have in the likelihood criterion is much larger, and this results in ML parameter estimates that are already quite adequate from expansion periods. As such, there is a smaller margin for improvement for the WMLE.

	h=1			h=4		
	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2
AR(1)	0.9708	0.977	53.6	0.9413*	1.000	2.2
	(0.2748)			(0.0986)		
AR(AIC)	0.9126*	0.994	79.2	0.8709	0.998	8.8
	(0.0983)			(0.1609)		

Table 3.2: Rolling-window forecasting results for monthly U.S IPI growth rate. The forecasts were computed over the expansion period 2001Jan-2007Dec

3.6.3 Using WMLE on Other Data Sets

We end this section with two additional applications of our WMLE. Our objective is to highlight that the WMLE can be used with different data sets and that it can deliver some remarkable improvements in forecasting accuracy in certain settings.

Table (3.3) reports the RMSE and MAE ratios obtained from an out-of-sample forecasting of both the *U.S. unemployment rate* and the *U.S. Total Non-Farm Payrolls* during the global recession of 2008. Due to size constraints, we do not report the estimates of ρ_1 and ρ_2. A full table with parameter estimates can be found in the Section 3.9.

The upper panel of Table (3.3), which refers to the *U.S. unemployment rate*, reveals that the WMLE is capable of delivering significant reductions in the out-of-sample forecasting RMSE and MAE. In particular, we report reductions of up to 34% in the monthly forecast RMSE of the AR(1) model, and up to 23% in the MAE forecast of the AR(AIC) at a yearly horizon.

The lower panel of Table (3.3) provides similar results for the forecasting of Total Non-Farm Payrolls in the US during the global recession. For the AR(1) model, The lower panel of Table (3.3) shows remarkable reductions of up to 48% in the out-of-sample forecasting RMSE, and reductions of more than 50% in the MAE at both monthly and yearly forecasting horizons. Most importantly, for the AR(AIC), we find reductions of more than 25% in the RMSE and over 40% in the MAE. Most of these reductions are statistically significant at any reasonable confidence level.

3.7 Conclusion

In this chapter, we introduced a new estimator that weights different observations in the likelihood function in order to deliver optimal forecasting accuracy for linear autoregres-

	RMSE Ratio		MAE Ratio	
	h=1	h=4	h=1	h=4
	U.S. Unemployment Rate			
AR(1)	0.6814***	0.7746*	0.7162**	0.7421*
	(0.0023)	(0.0788)	(0.0266)	(0.0586)
AR(AIC)	0.8710	0.8480	0.8235*	0.7780
	(0.1136)	(0.2922)	(0.0915)	(0.1567)
	U.S Total Non-farm Payrolls			
AR(1)	0.5212***	0.5455**	0.4401***	0.4582*
	(0.0000)	(0.0155)	(0.0000)	(0.0003)
AR(AIC)	0.6700***	0.6369	0.6087***	0.5659
	(0.0060)	(0.1715)	(0.0009)	(0.1538)

Table 3.3: Rolling-window forecasting results for both monthly U.S Unemployment Rate and Total Non-farm Payrolls. Entries Ratio show the RMSE or MAE using WML method of the AR model relative to RMSE or MAE using normal ML method. The forecasts were computed over the sample period 2008Jan-2009June

sive models. We showed how to estimate the optimal weights using a cross-validation technique. An application of the WMLE to the US IPI revealed that the forecasting accuracy during the latest global recession can be significantly improved by increasing the weights of observations corresponding to past recession periods. This highly intuitive empirical finding showed that we should look more carefully at past recessions when we wish to produce accurate forecast during the recent global financial crisis. Finally, we have shown that the advantages of the WMLE extend also to forecasting a number of other time-series. In effect, the WMLE seems to deliver dramatic improvements in forecasting accuracy in some applications that we studied.

3.8 Proofs of Theorems and Propositions

Proof of Proposition 3.1. Theorem 3.1 in Bougerol (1993) implies that the sequence $\{X_t\}_{t\in\mathbb{N}}$ initialized at $X_1 = x$ and generated according to (3.5) for every $t \in \mathbb{N}$, converges exponentially almost surely (e.a.s) to an SE limit sequence $\{X_t\}_{t\in\mathbb{Z}}$, initialized in the infi-

nite past, as long as $\{V_t\}_{t\in\mathbb{Z}}$ is an SE n_V-variate stochastic sequence, $\phi \in \mathbb{C}^1(\mathcal{X}, \mathcal{V})$, $\mathbb{E}\log^+|\phi(x, V_t)|^2 < \infty$ and $\mathbb{E}\log\sup_{x\in\mathcal{X}}|\phi'_x(x, Z_t)|^2 < 0$. The first two conditions are directly given by *(i)* and *(ii)*. The two remaining conditions are implied by *(iii)* and *(iv)*. The two bounded moments of $\{X_t\}_{t\in\mathbb{Z}}$ are ensured by conditions *(i)-(iv)* (Blasques et al. (2014b), Proposition SA.1).

The AR(p) representation follows trivially by re-writing the X_t as follows

$$\begin{aligned} X_t &= \phi(X_{t-1}, V_t) - \epsilon_t + \epsilon_t \\ &= \frac{\phi(X_{t-1}, V_t) - \epsilon_t}{\psi_0 + \Psi(L)X_t}(\psi_0 + \Psi(L)X_t) + \epsilon_t \end{aligned}$$

where $\Psi(L)$ denotes the lag polynomial $\boldsymbol{\alpha}(L) = \psi_1 L + \cdots + \psi_p(L^p)$, and finally defining

$$\alpha_{i,t} := \frac{\big(\phi(X_{t-1}, V_t) - \epsilon_t\big)}{\Psi(L)X_t}\psi_i \ , \qquad \text{for } i = 0, 1, ..., p.$$

The SE nature of $\alpha_{i,t}$, $i = 0, ..., p$, follows directly from Krengel's Theorem (Akcoglu (1979)) since every $\alpha_{i,t}$ is a measurable function of SE variables. □

Proof of Proposition 3.2. Under the conditions of Proposition A.1, we have that $\{X_t\}_{t\in\mathbb{Z}}$ is weakly stationary. Since furthermore, the Gaussian AR(p) model is well specified, it follows immediately that the Gaussian MLE converges to the true parameter as $S \to \infty$, i.e. $\hat{\boldsymbol{\theta}}(1) \xrightarrow{p} \boldsymbol{\theta}_0^*(1) = \boldsymbol{\theta}_0$; see e.g. Brockwell and Davies (1986).

Application of a continuous mapping theorem as $S \to \infty$ implies that

$$u_t(\hat{\boldsymbol{\theta}}(1)^2 \equiv \hat{X}_{k+n}\big(\hat{\boldsymbol{\theta}}(1)\big) - X_{k+n} \xrightarrow{p} u_t(\boldsymbol{\theta}_0^*(1))^2 \equiv \hat{X}_{k+n}\big(\boldsymbol{\theta}_0\big) - X_{k+n}.$$

As a result, the limit as $S \to \infty$ of the n-step-ahead MSFE forecast criterion based on H observed forecast errors under the true parameter $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_0^*(1)$ is given by

$$Q_H(\mathbf{W})) := \frac{1}{H-n}\sum_{t=1}^{H} u_t(\boldsymbol{\theta}_0^*(1))^2.$$

Since $u_t(\boldsymbol{\theta}_0) = \epsilon_t \ \forall \ t$, an application of the ergodic theorem then yields

$$Q_H(\mathbf{W}) \xrightarrow{p} \mathbb{E}u_t(\boldsymbol{\theta}_0^*(\mathbf{w}_k))^2 \quad \text{as} \quad H \to \infty.$$

Finally, note that Algorithm 1 is always initialized at a weight matrix $\mathbf{W}$ satisfying $\hat{\mathbf{w}}_{k,t} = 1 \ \forall \ (k, t)$. As a result, in the limit as $S \to \infty$ and $H \to \infty$, the probability that $\hat{\mathbf{w}}_{k,t} \neq 1$ for some (k, t) is given by

$$\mathbb{P}(\hat{\mathbf{w}}_{k,t} \neq 1) = \mathbb{P}\Big(\mathbb{E}u_t(\boldsymbol{\theta}_0^*(\mathbf{w}_k))^2 < \mathbb{E}u_t(\boldsymbol{\theta}_0^*(1))^2\Big) = 0$$

for any weight vector $\mathbf{w}_k$ with some element $w_{k,t} \neq 1$ at some pair (t, k). □

Proof of Proposition 3.3. The first claim follows trivially from the design of Algorithm 1. The second claim follows by noting that both the AR parameter vector $\hat{\boldsymbol{\theta}}$ and the weight parameter vector $\hat{\boldsymbol{\rho}}$ satisfy a recursive relation

$$\hat{\boldsymbol{\theta}}_{j+1} = \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\rho}}(\hat{\boldsymbol{\theta}}_j)) \quad \text{and} \quad \hat{\boldsymbol{\rho}}_{j+1} = \hat{\boldsymbol{\rho}}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\rho}}_j)) \quad \forall\ j \geq 1.$$

It is well known that the uniform unit bound on the derivative ensures the stability of the recursion towards a unique global fixed point for any initialization. □

Proof of Proposition 3.4. By Propositions A.1 and 3.2, the MLE can only perform as well as the WMLE if the model is well specified. As a result, the condition that $\text{MSFE}(\mathbf{W}^*) = \text{MSFE}(1)$ ensures that the model is well specified under the null. Furthermore, under correct specification, the WMLE can only satisfy $\text{MSFE}(\mathbf{W}) = \text{MSFE}(1)$ if it is constant.

Under the conditions of Proposition A.1, the data $\{X_t\}_{t\in\mathbb{Z}}$ is SE with four bounded moments. Since the MLE and WMLE are constant over k, it follows immediately that both $u_k(1)^2$ and $u_k(1)^2$ are weakly stationary. The DM assumptions are thus satisfied under H_0 since $\text{MSFE}(\mathbf{W}^*) = \text{MSFE}(\mathbf{W}) = \text{MSFE}(1)$ ensures also that $\mathbb{E}\bar{d}_k(\mathbf{w}_k) = 0$. □

3.9 Extra Tables

Table 3.4: AR Model Forecasting Results: Binary Weight with Decaying

	RMSE						MAE					
	h=1			h=4			h=1			h=4		
Window	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2
AR(1)												
25-year	1.0908	1.000	35.5	0.8297*	1.000	27.7	0.8310	1.000	31.5	0.7919*	1.000	27.1
	(0.3446)			(0.0679)			(0.1564)			(0.0500)		
50-year	0.9108**	1.000	15.7	0.7423**	1.000	12.3	0.8007***	1.000	13.8	0.5975***	1.000	11.1
	(0.0268)			(0.0244)			(0.0011)			(0.0000)		
AR(AIC)												
25-year	1.4691	0.999	119.0	0.9189	1.000	11.1	1.4131	1.000	250.0	0.9450	1.000	8.8
	(0.1481)			(0.2707)			(0.1754)			(0.3918)		
50-year	0.9293	0.998	15.9	0.7782*	1.000	9.7	0.8427**	1.000	25.7	0.6955***	1.000	11.5
	(0.1292)			(0.0523)			(0.0153)			(0.0021)		

Table 3.4: Rolling-window forecasting results for monthly U.S IPI growth rate. Entries Ratio show the mean-squared-forecast-error(MSE) or mean-absolute-error(MAE) using WML method of the AR model relative to MSE or MAE using normal ML method. columns labelled ρ_1 and ρ_2 show the optimal weighting parameter for combined weighting function. The forecasts were computed over the sample period 2008Jan-2009June

Table 3.5: AR Model Forecasting Results: Binary Weight with Decaying, Expansion Period

	RMSE						MAE					
	h=1			h=4			h=1			h=4		
window length	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2
AR(1)												
25-year	0.9708	0.977	53.6	0.9413*	1.000	2.2	0.9652	0.967	32.5	1.0000	1.000	1.0
	(0.2748)			(0.0986)			(0.2397)			(0.8536)		
50-year	0.9605	0.977	129.3	1.0000	1.000	1.0	0.9530	0.967	32.2	1.0000	1.000	1.0
	(0.2456)			()			(0.1943)			()		
AR(AIC)												
25-year	0.9126*	0.994	79.2	0.8709	0.998	8.8	0.9430	0.990	71.9	0.9154	0.993	12.6
	(0.0983)			(0.1609)			(0.1452)			(0.1632)		
50-year	0.9727	0.990	47.3	1.0155	0.999	1.0	0.9487	0.989	71.8	1.0686	0.994	1.0
	(0.2604)			(0.8852)			(0.1332)			(0.9716)		

Table 3.5: Rolling-window forecasting results for monthly U.S IPI growth rate. Entries Ratio show the mean-squared-forecast-error(MSE) or mean-absolute-error(MAE) using WML method of the AR model relative to MSE or MAE using normal ML method. columns labelled ρ_1 and ρ_2 show the optimal weighting parameter for combined weighting function. The forecasts were computed over the sample period 2001Jan-2007Dec

Table 3.6: AR Model Forecasting Results: U.S. Unemployment Rate

	RMSE						MAE					
	h=1			h=4			h=1			h=4		
Window	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2
AR(1)												
25-year	0.6626***	1.000	72.3	0.7449*	1.000	79.8	0.7095**	0.999	202.9	0.7596*	1.000	300.7
	(0.0009)			(0.0541)			(0.0153)			(0.0642)		
50-year	0.6814***	0.996	40.8	0.7746*	1.000	15.3	0.7162**	0.994	482.2	0.7421*	1.000	38.8
	(0.0023)			(0.0788)			(0.0266)			(0.0586)		
AR(AIC)												
25-year	0.8694	1.000	7.2	1.0291	0.946	1.0	0.9159	1.000	6.8	0.8398	0.952	1.0
	(0.1010)			(0.5258)			(0.2351)			(0.3481)		
50-year	0.8710	1.000	13.0	0.8480	0.998	14.3	0.8235*	1.000	39.7	0.7780	1.000	19.5
	(0.1136)			(0.2922)			(0.0915)			(0.1567)		

Table 3.6: Rolling-window forecasting results for monthly U.S Unemployment. Entries Ratio show the mean-squared-forecast-error(MSE) or MAE using WML method of the AR model relative to MSE or MAE using normal ML method. columns labelled ρ_1 and ρ_2 show the optimal weighting parameter for combined weighting function. The forecasts were computed over the sample period 2008Jan-2009June

Table 3.7: AR Model Forecasting Results: U.S. Total Non-farm Payrolls

	RMSE						MAE					
	h=1			h=4			h=1			h=4		
Window	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2
AR(1)												
25-year	0.5860***	1.000	66.5	0.7019	1.000	56.0	0.5027***	1.000	83.3	0.5943*	1.000	58.1
	(0.0010)			(0.1672)			(0.0002)			(0.0883)		
50-year	0.5212***	1.000	18.4	0.5455**	1.000	18.0	0.4401***	1.000	33.4	0.4582***	1.000	25.3
	(0.0000)			(0.0155)			(0.0000)			(0.0003)		
AR(AIC)												
25-year	0.7688**	0.990	7.4	0.8236	1.000	32.1	0.6393***	0.992	7.5	0.7253	0.997	16.4
	(0.0348)			(0.2893)			(0.0059)			(0.1954)		
50-year	0.6700***	1.000	8.7	0.6369	0.990	39.0	0.6087***	1.000	6.4	0.5659	0.989	23.2
	(0.0060)			(0.1715)			(0.0009)			(0.1538)		

Table 3.7: Rolling-window forecasting results for monthly U.S Total Non-farm Payrolls. Entries Ratio show the mean-squared-forecast-error(MSE) or MAE using WML method of the AR model relative to MSE or MAE using normal ML method. columns labelled ρ_1 and ρ_2 show the optimal weighting parameter for combined weighting function. The forecasts were computed over the sample period 2008Jan-2009June

Chapter 4

Weighted Maximum Likelihood for Dynamic Factor Analysis and Forecasting with Mixed Frequency Data

4.1 Introduction

The forecasting of macroeconomic and financial time series variables is of key importance for economic policy makers. Reliable forecasts are especially in high demand when the economic environment is uncertain as we have witnessed in the years during and after the financial crisis. Many different model-based approaches exist for this purpose, ranging from basic time series models to sophisticated structural dynamic macroeconomic models. The underlying idea of the dynamic factor model is to associate a relatively small set of factors to a high-dimensional panel of economic variables that includes the variables of interest and related variables. The dynamic factor model has become a popular tool for the forecasting of the variable of interest, amongst practitioners and econometricians. This is mainly due to their good forecast performance as shown in many studies.

The dynamic factor model can be viewed as a high-dimensional linear state space model. The estimation of the parameters in a dynamic factor model is a challenging task given the large number of parameters, mostly due to factor loading coefficients. A likelihood-based approach in which the Gaussian likelihood function is evaluated via the Kalman filter and is numerically maximized with respect to the parameter vector has been originally proposed by Engle and Watson (1981) for a model with one dynamic factor. Watson and Engle (1983) base their estimation procedure on an expectation-

maximization (EM) algorithm; see also Quah and Sargent (1993). More recently, feasible two-step approximate likelihood-based procedures are developed by Doz et al. (2011) and Bańbura and Modugno (2014). In Bräuning and Koopman (2014) and Jungbacker and Koopman (2015), specific data transformations are considered to facilitate the parameter estimation for high-dimensional dynamic factor models. In this study, we restrict ourselves to likelihood-based estimation procedures.

Weighted likelihood-based estimation

To address the notion that a single variable or a small selection of variables in a dynamic factor model is of key importance while all other variables can be regarded as instruments, we present a weighted likelihood-based estimation procedure for the purpose of providing a more accurate forecasting performance than obtained from a standard maximum likelihood procedure. Our proposed weighted maximum likelihood estimator gives simply more weight to the likelihood contribution from the variable of interest. As an example, for the nowcasting and forecasting of quarterly growth in gross domestic product, referred to as GDP growth, more weight can be given to the likelihood contribution from GDP growth in comparison to the contribution from the related variables that are included in the dynamic factor model.

The variable-specific weights introduced by our weighted ML estimator differ from other weighted ML estimators proposed in the literature. In most other cases, observation-specific weights in the likelihood function are considered. The local ML estimators studied in Tibshirani and Hastie (1987), Staniswalis (1989) and Eguchi and Copas (1998) assign a weight to each observation that depends on the distance to a given fixed point. The robust ML estimator of Markatou et al. (1997, 1998) Markatou et al. (1998) down-weights observations that are inconsistent with the postulated model. Similarly, Hu and Zidek (1997) devise a general principle of relevance that assigns different weights to different observations in an ML setting. In small samples, this type of estimator can provide important gains in the trade-off between bias and precision of the ML estimator. The large sample properties of these estimators are established in Wang et al. (2004) for given weights, and Wang and Zidek (2005) provide a method for estimating the weights based on cross-validation. In contrast we propose a weighted ML estimator that gives higher weight to a subset of a random vector, that is to an entire random scalar sequence within the multivariate stochastic process.

We discuss the asymptotic properties of our weighted maximum likelihood estimator and we show that the estimator is consistent and asymptotically normal. We also verify our new approach in a Monte Carlo study to investigate the effect of different choices for the weights in different scenarios. In an empirical study concerning the nowcasting and forecasting of U.S. GDP growth, we adopt the weighted likelihood function for the estimation of parameters in a mixed frequency dynamic factor model.

Mixed Frequency

In empirical studies, the dynamic factor model requires further modifications to handle mixed frequency data; Mariano and Murasawa (2003) has been the first to illustrate how a small-scale dynamic factor model for the U.S. economy can be adapted for mixed frequency data. Their model is formulated in state space form with a monthly time index. The monthly and quarterly variables are dependent on a common monthly dynamic factor and on idiosyncratic dynamic components. For the quarterly variable of interest, the Kalman filter can treat the missing observations that occur during the first two months in each quarter. More generally, any multivariate time series model can be formulated in terms of a high frequency time index and the periodically missing observations due to low frequency variables can be accounted for by the Kalman filter. Mittnik and Zadrozny (2005) report promising results based on this approach for the forecasting of German growth in GDP.

We consider an alternative approach based on ideas developed for periodic systems in the control engineering literature; see Bittanti and Colaneri (2000, 2009).Bittanti and Colaneri (2009) The main idea is to formulate the model with a low frequency time index and collect the observations for a high frequency variable in a vector. In the case of a quarterly time index and a monthly variable, the three consecutive monthly observations associated with a specific quarter are then stacked into a quarterly vector. Both monthly and quarterly dynamic processes can be formulated in a state space model with a quarterly time index. We discuss this solution for the mixed frequency dynamic factor model. The advantage of this approach is that it does not require the handling of missing observations and it can lead to computational efficiencies. A similar solution is considered by Marcellino et al. (2014) who propose a Bayesian regression model with stochastic volatility for producing current-quarter forecasts of GDP growth using many monthly economic variables. Such ideas are also explored for vector autoregressive systems by Chen et al. (2012), Ghysels (2012), Foroni et al. (2015) and Ghysels et al. (2015).

Empirical study

An important application of dynamic factor models is their use in the forecasting of quarterly GDP growth. A high-dimensional panel of macroeconomic variables is used to construct factors for the purpose of facilitating the forecasting of GDP growth. Empirical evidence is given by, amongst others, Stock and Watson (2002b) and Giannone et al. (2008) for the U.S., Marcellino et al. (2003) and Rünstler et al. (2009) for the euro area, and Schumacher and Breitung (2008) for Germany. In many of these and related studies, the problem of mixed frequency data arises since the variable of interest GDP growth is observed at a quarterly frequency while the other macroeconomic variables are observed at a monthly frequency. The treatment of mixed frequency data in a dynamic factor model is therefore a highly relevant issue in forecasting, nowcasting and backcasting GDP growth; see also the discussions in Bańbura et al. (2013).

In our empirical study for the U.S. economy, we consider three small- to medium-sized mixed frequency dynamic factor models with the purpose of forecasting quarterly U.S. GDP growth. The first model is a five-dimensional model similar to Mariano and Murasawa (2003), the second model is a fourteen-dimensional model similar to Bańbura et al. (2013) and the third model is a six-dimensional model similar to Aruoba et al. (2009). The first two models have only monthly related variables while the last model also includes a weekly related variable. For almost all cases, we present improvements in nowcasting and forecasting accuracy when parameters are estimated by the weighted maximum likelihood method.

Outline

The outline of this chapter is as follows. In Section 4.2 we present our weighted maximum likelihood approach that is introduced to increase the influence of the key variables in the estimation process for a joint multivariate dynamic model. Asymptotic properties of the resulting estimator are derived and we explore its small-sample properties in a Monte Carlo study. In Section 4.3 we show how mixed frequency dynamic factor models can be specified as observationally equivalent low frequency dynamic factor models. In many cases the low frequency formulations lead to computational gains. In Section 4.4 we present and explore the results of our empirical study concerning U.S. GDP growth. We compare the nowcasting and forecasting accuracies of our new approach for the three different dynamic factor models. We also establish the empirical relevance of the weighted estimation method of Section 4.2. Section 4.5 summarizes and concludes.

4.2 Weighted Maximum Likelihood: method and properties

We represent our high-dimensional panel of time series as the column vector z_t for which we have observations from $t = 1, \ldots, T$ where T is the overall time series length. We decompose z_t into variables of interest in y_t and related variables in x_t, we have $z_t = (y_t', x_t')'$ where a_t' is the transpose of column vector a_t. The dimension N_y of y_t is small and typically equal to one while the dimension N_x of x_t can be large. Hence the dimension N_z of z_t is also large since $N_z = N_y + N_x$. It is assumed that all time series variables in z_t have zero mean and are strictly stationary. The basic dynamic factor model for z_t can be represented by $z_t = \Lambda f_t + \varepsilon_t$ or

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \Lambda_y \\ \Lambda_x \end{bmatrix} f_t + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix}, \qquad f_t = \Phi f_{t-1} + \eta_t, \tag{4.1}$$

for $t = 1, \ldots, T$, where $\Lambda = [\Lambda_y', \Lambda_x']'$ is the factor loading matrix with dimension $N_z \times r$, Λ_i is the $N_i \times r$ factor loading sub-matrix Λ_i, for $i = y, x$ with A' being the transpose of matrix A, f_t is the $r \times 1$ vector with latent dynamic factors, $\varepsilon_t = (\varepsilon_{y,t}', \varepsilon_{x,t}')'$ is the observation disturbance vector, with $\varepsilon_{i,t}$ as the normally distributed $N_i \times 1$ disturbance

vector, for $i = x, y$, Φ is the autoregressive coefficient matrix, and η_t is the normally distributed factor disturbance vector. The dynamic factor f_t represents the common dynamic variations in the time series variables in z_t. The dynamic process for f_t is specified as a strictly stationary vector autoregressive process. Hence, matrix Φ is subject to the appropriate conditions for stationarity. Other stationary, linear dynamic processes can also be considered for f_t. For identification purposes we further assume that the factors are normalized, that is $\mathrm{E}(f_t) = 0$ and $\mathrm{Var}(f_t) = I_r$ with I_k being the $k \times k$ identity matrix for any positive integer k. In our treatment below, the initial factor f_1 is treated as a fixed value, that is $f_1 = f_1^*$. The disturbance vectors ε_t and η_t are assumed to be mutually and serially uncorrelated for all time periods. In particular, we have

$$\varepsilon_t \sim \mathrm{N}(0, \Sigma_\varepsilon), \qquad \eta_t \sim \mathrm{N}(0, \Sigma_\eta), \qquad \mathrm{Cov}(\varepsilon_t, \eta_s) = 0, \tag{4.2}$$

for $t, s = 1, \ldots, T$. To enforce the normalization of the factors f_t, the variance matrix Σ_η is restricted to be $\Sigma_\eta = I_r - \Phi\Phi'$. The remaining coefficient matrices Λ, Σ_ε and Φ are functions of the unknown parameter vector that we denote by ψ. This dynamic factor model is stylized for the purpose of presenting our developments below. However, our results are general for other multivariate dynamic specifications, including the mixed frequency dynamic factor model that we adopt in Section 4.3.

Different methods have been proposed for the estimation of the unknown parameter vector ψ; see the discussion in the introductory section. We restrict ourselves to those methods that are likelihood-based and aim at maximizing the loglikelihood function. The loglikelihood function relies on the joint logdensity $\log p(z; \psi) = \log p(y, x; \psi)$ where $p(\cdot)$ is the Gaussian density function, $z = (y', x')'$, $y = (y_1', \ldots, y_T')'$ and $x = (x_1', \ldots, x_T')'$. Since the dynamic factor model (4.1)–(4.2) can be represented as a stationary Gaussian linear state space model, the Kalman filter can be used to evaluate the loglikelihood function; we refer to Harvey (1989) and Durbin and Koopman (2012) for treatments of the Kalman filter. The maximum likelihood (ML) estimates are obtained by numerically maximizing the loglikelihood function with respect to ψ. It is a standard exercise of numerical optimization in which the Kalman filter evaluates the loglikelihood function whenever a different ψ is considered. However, the ML estimator is not necessarily the best estimator in the context of the dynamic factor model for the following two reasons: (i) the dynamic factor model only provides a parsimonious approximation to a high-dimensional complex data generation process of the variables in z_t; (ii) the dynamic factor model is typically used for the forecasting of the variables in y_t rather than the forecasting of all variables in z_t.

The dynamic factor model (4.1) – (4.2) provides a convenient framework for obtaining simple descriptions of potentially complex interactions between the economic variables. In particular, the common factors summarize partly the commonalities in the dynamic variations in the related variables x_t. Furthermore, the factors deliver a parsimonious description of the relationships between the variables of interest in y_t and the the related

variables in x_t. The dynamic factor model is mainly used to approximate the true and unknown data generation process. It is not intended to be an exact representation of the true underlying dynamics of the economy.

In the context of parameter estimation, we need to address the problem of model misspecification and the focus on a subset of variables only. Each of these issues are not necessarily sufficient to abandon the ML estimator, but taken together, they are. If we are only interested in a subset of the variables, but the model is correctly specified, then the ML estimator is still the best under the usual regularity conditions that make it consistent and efficient. In particular, by converging to the true parameter and attaining a minimum variance, the ML estimator provides the best possible parameter estimates for the purpose of forecasting the variable of interest. This is true even if the variable of interest happens to be only a subset of the observed variables. Similarly, if a model is misspecified but our interest lies in forecasting all observed variables, then there are still good reasons to employ the ML estimator. Under standard conditions, the ML estimator converges to a pseudo-true parameter that minimizes the Kullback-Leilber (KL) divergence between the true joint distribution of the data and the model-implied distribution. The KL divergence has well established information-theoretic optimal properties. Furthermore, under weak regularity conditions and depending on the distribution of the data, it is also easy to show that the pseudo-true parameter optimizes forecasting accuracy. However, we will argue that when we take the above points (i) and (ii) together, the ML estimator is no longer the best possible estimator available when interest lies in the forecasting of only a subset of the observed variables. These charasteristic features call for a novel estimation procedure with the aim to improve the forecasting accuracy of the variable of interest in the context of a misspecified model. We provide both theoretical and simulation-based evidence that a weighted ML estimator outperforms the classic ML estimator in forecasting the variable of interest. In Section 4.4, we show that the ability to outperform the ML estimator is also visible in an empirically relevant application to economic forecasting.

4.2.1 Weighted Maximum Likelihood Estimator

Consider the dynamic factor model (4.1) and (4.2) where both y_t and x_t can be treated as vectors. The loglikelihood function for the model is given by

$$\mathcal{L}_T(\psi, f_1^*) := \log p(y, x; \psi) = \log p(y|x; \psi) + \log p(x; \psi), \tag{4.3}$$

where the initial value of the factor $f_1 = f_1^*$ and the parameter vector ψ are both treated as fixed unknown values. It is a standard result that the joint density can be expressed as a conditional density multiplied by a marginal density. However, for our purposes the expression (4.3) is useful as it highlights the different roles of y and x: the variable y_t is our key variable for which we require accurate model-based forecasts while the variables represented by x_t are typically instrumental to improve the nowcasts and forecasts of y_t.

Under the assumption that y and x are jointly generated by the Gaussian dynamic factor model (4.1), we can apply the Kalman filter to evaluate the loglikelihood function via the prediction error decomposition; see Koopman and Durbin (2000) for a computationally efficient implementation in the context of a large-dimensional observation vector $z_t = (y_t', x_t')'$.

The maximum likelihood estimation of parameter vector ψ is based on applying a numerical quasi-Newton optimization method for the maximization of $\mathcal{L}_T(\psi, f_1^*)$, with respect to ψ. The numerical maximization is an iterative process. After its convergence, the maximum likelihood estimate of ψ is obtained. For each iteration in this process, various loglikelihood evaluations are required and they are carried out by the Kalman filter. In the context of our empirical study in Section 4.4, the treatment of the observations in z_t for the construction of the likelihood function is implied by the dynamic factor model. However, it is most likely that the dynamic factor model is misspecified as a model representation of the true data generation process for the variables represented in z_t. When our primary aim is to analyze y_t in particular, we may be less concerned with the misspecification of x_t, to some extent. We do not want to disregard the likelihood contribution of x_t completely since we still require an appropriate model representation of x_t given the dependence of y_t on x_t through the common dynamic factor f_t. To reflect the higher importance of y_t in comparison to x_t in the likelihood construction for the misspecified dynamic factor model, we propose to give different weights to the likelihood contributions of y_t and x_t explicitly. Hence we propose the weighted loglikelihood function

$$\mathcal{L}_T(\psi, w, f_1^*) = W \cdot \log p(y|x; \psi) + \log p(x; \psi), \tag{4.4}$$

for a fixed and predetermined weight $W \geq 1$ and with $w := W^{-1} \in [0, 1]$. The weight W is conveniently used in our Monte Carlo and empirical studies below while it is more appropriate to work with the inverse weight w in the asymptotic theory that is developed next. The construction of the weighted loglikelihood function does not need further modifications. The estimator of ψ that maximizes (4.4) is referred to as the weighted maximum likelihood (WML) estimator.

The novel WML estimator differs from other weighted ML estimators proposed in the literature. Our WML estimator is unique in introducing variable-specific weights rather than observation-specific weights in the likelihood function. For example, local ML estimators assign a weight to each observation that depends on the distance to a given fixed point; see Tibshirani and Hastie (1987), Staniswalis (1989) and Eguchi and Copas (1998). The robust ML estimator of Markatou et al. (1997, 1998)Markatou et al. (1998) are designed to reduce influence of outliers by down-weighting observations that are inconsistent with the postulated model. The general principle of relevance of Hu and Zidek (1997) assigns different weights to different observations in the likelihood function.

The motivation for the development of our WML estimator is also different. The WML estimator is designed to perform well when the model is misspecified and interest

lies in forecasting only a subset of the observed variables. For this reason we analyze the asymptotic properties of our WML estimator allowing for the possibility of model misspecification and focus on the approximation to an unknown data generation process.

4.2.2 Asymptotic Properties of the WML Estimator

The properties of the weighted maximum likelihood estimator are derived for any choice of weight $w := W^{-1} \in [0, 1]$. We show that, when the model is correctly specified, then the WML estimator $\hat{\psi}_T(w)$ is consistent and asymptotically normal for the true parameter vector $\psi_0 \in \Psi$. When the model is misspecified, we show that $\hat{\psi}_T(w)$ is consistent and asymptotically normal for a pseudo-true parameter $\psi_0^*(w) \in \Psi$ that minimizes a transformed Kullback–Leibler (KL) divergence between the true probability measure of the data and the measure implied by the model. We show that the transformed KL divergence takes the form of a pre-metric that gives more weight to fitting the conditional density of y_t when $0 < w < 1$. We use the term pre-metric to identify a map that satisfies all the axioms of metrics except for the *axiom of symmetry*, for example $d(x, y) = d(y, x)\ \forall\ (x, y)$. For the special case where $w = 1$, we obtain the classical pseudo-true parameter $\psi_0^*(1) \in \Psi$ of the ML estimator that minimizes the KL divergence. The proofs of the propositions and theorems presented in this section below are presented in the Online Appendix.

Proposition 4.1 below states standard conditions for the strict stationarity and ergodicity (SE) of the true processes $\{f_t\}_{t\in\mathbb{Z}}$, $\{x_t\}_{t\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ generated by the linear Gaussian model in (4.1) and (4.2), initialized in the infinite past. Below, we let $\|\cdot\|$ denote the Euclidean spectral norm, i.e. the p-norm that sets $p = 2$. The results extend naturally to other norms.

Proposition 4.1. *Let $\{x_t\}_{t\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ be generated according to (4.1) and (4.2) with*
(i) $\|\Phi\| < 1$ in (4.1) and $0 < \|\Sigma_\eta\| < \infty$ in (4.2);
(ii) $\|\Lambda_x\| < \infty$ in (4.1) and $0 < \|\Sigma_\varepsilon\| < \infty$ in (4.2);
(iii) $\|\Lambda_y\| < \infty$ in (4.1).
Then $\{x_t\}_{t\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ are SE sequences with bounded moments of any order; i.e. $\mathbb{E}|x_t|^r < \infty$ and $\mathbb{E}|y_t|^r < \infty\ \forall\ r > 0$.

Theorem 4.1 ensures the existence of the WML estimator as a random variable that takes values in the arg max set of the random likelihood function.

Theorem 4.1. (Existence) *For given $w \in [0, 1]$, let $(\Psi, \mathfrak{B}(\Psi))$ be a compact measurable space. Then there exists a.s. a measurable map $\hat{\psi}_T(w, \tilde{f}_1^*) : \Omega \to \Psi$ satisfying*

$$\hat{\psi}_T(w, \tilde{f}_1^*) \in \arg\max_{\psi\in\Psi} \mathcal{L}_T(\psi, w, \tilde{f}_1^*),$$

for all $T \in \mathbb{N}$ and every filter initialization $\tilde{f}_1^*$.

Theorem 4.2 establishes the strong consistency of the WML estimator of the true parameter vector $\psi_0 \in \Psi$ for any choice of weight $w \in (0, 1]$ for the likelihood. Since the true time-varying parameter $\{f_t\}_{t\in\mathbb{Z}}$ is unobserved, the estimation of the true parameter vector ψ_0 depends crucially on the use of a filter for this sequence. Below, we use tildes to distinguish the filter from the true parameter and hence let $\tilde{f}_t(\psi, \tilde{f}_1^*(\psi))$ denote the filter at time t, parameterized by ψ and initialized at point $\tilde{f}_1^*(\psi)$ at time $t = 1$. The asterisk highlights that $\tilde{f}_1^*(\psi)$ is simply an initialization for the filter. Note that the filter depends on the vector of static parameters $\psi \in \Psi$. Naturally, it is the static parameter ψ that defines the properties of the filtering sequence $\{\tilde{f}_t(\psi, \tilde{f}_1^*(\psi))\}_{t\in\mathbb{N}}$.

From an estimation perspective, we are interested in the filtered sequence as a function of the vector of parameters ψ. As such, we will adopt the stochastic recurrence approach of Straumann and Mikosch (2006) and concentrate on the random functions $\tilde{f}_t(\tilde{f}_1^*) = \tilde{f}_t(\cdot, \tilde{f}_1^*(\cdot))$ that take values in the separable Banach space $\mathbb{C}(\Psi, \|\cdot\|_\Psi)$ of continuous functions defined on Ψ, equipped with supremum norm $\|\cdot\|_\Psi$.

The consistency of the WMLE established in Theorem 4.2 is obtained under the assumption that the common factor model is correctly specified. Furthermore, Theorem 4.2 holds for any filter that identifies the parameter vector $\psi_0 \in \Psi$ and is asymptotically SE with bounded moments of second order. The identification of $\psi_0 \in \Psi$ is naturally ensured when the filter is invertible since then the limit filtering sequence $\{\tilde{f}_t\}_{t\in\mathbb{Z}}$ will then satisfy $\tilde{f}_t(\psi_0) = f_t \ \forall \ t \in \mathbb{Z}$. Furthermore, the exponential almost sure (e.a.s.) convergence of the filter $\{\tilde{f}_t(\tilde{f}_1^*)\}_{t\in\mathbb{N}}$ to a limit SE sequence $\{\tilde{f}_t\}_{t\in\mathbb{Z}}$ that does not depend on the initialization $\tilde{f}_1^*$ is ensured as long as standard contraction conditions hold. Both the identification condition and the e.a.s. convergence of the filter to an SE process with bounded second moment are standard and easy to establish in this linear Gaussian setting. For this reason, we do not repeat them here; see e.g. Mehra (1970) and Bougerol and Picard (1992)for such results on the classical Kalman filter, Bougerol (1993) and Straumann and Mikosch (2006) for extensions to other filters, and Blasques et al. (2014a) for identification, convergence results and bounded moments on a wide range of observation-driven filters. Theorem 4.2 thus assumes that ψ_0 maximizes the likelihood and assumes the convergence of the filtered sequence $\{\tilde{f}_t(\tilde{f}_1^*)\}_{t\in\mathbb{N}}$ initialized at $\tilde{f}_1^* \in \mathbb{C}(\Psi)$ to a unique limit SE sequence $\{\tilde{f}_t\}_{t\in\mathbb{Z}}$ with bounded second moment.

The uniform convergence of the filtering sequence to an SE limit let us to establish the uniform convergence of the weighted loglikelihood function by application of a uniform law of large numbers. We denote the limit weighted loglikelihood function by $\mathcal{L}_\infty(\psi, w)$. We only require the identification of ψ_0 in the usual ML setting ($w = 1$); that is identification w.r.t. the unweighted loglikelihood function $\mathcal{L}_T(\psi, 1)$. In the proof, we show that identification of ψ_0 in $\mathcal{L}_T(\psi, 1)$ implies identification of ψ_0 in $\mathcal{L}_T(\psi, w)$, for any $w \in (0, 1]$.

Theorem 4.2. (Consistency) *Let $\{x_t\}$ and $\{y_t\}$ be generated by the dynamic factor model defined in (4.1) and (4.2) under some $\psi_0 \in \Psi$, and suppose that the conditions of*

Propositions 4.1 and Theorem 4.1 hold. Suppose furthermore that

$$\mathcal{L}_\infty(\psi_0, 1) > \mathcal{L}_\infty(\psi, 1) \ \forall \ \psi \neq \psi_0$$

and there exists a unique SE sequence such that

$$\|\tilde{f}_t(\tilde{f}_1^*) - \tilde{f}_t\|_\Psi \overset{e.a.s.}{\to} 0 \ \forall \ \tilde{f}_1^* \quad as \quad t \to \infty \quad \quad with \quad \quad \mathbb{E}|\tilde{f}_t|^2 < \infty.$$

Then the WML estimator $\hat{\psi}_T(w, \tilde{f}_1^*)$ satisfies

$$\hat{\psi}_T(w, \tilde{f}_1^*) \overset{a.s.}{\to} \psi_0 \quad as \quad T \to \infty$$

for any choice of weight $w \in (0, 1]$ and any initialization $\tilde{f}_1^*$.

If the data $\{x_t\}$ and $\{y_t\}$ are obtained from an unknown data generating process but satisfy some regularity conditions, then we can still prove consistency of the WML estimator to pseudo-true parameter $\psi_0^*(w) \in \Psi$ that depends on the weight $w \in (0, 1]$.

The classical ML estimator converges to a limit pseudo-true parameter that minimizes the KL divergence between the true joint probability measure of the data and the measure implied by the model. Theorem 4.3 characterizes the limit pseudo-true parameter $\psi_0^*(w)$ as the minimizer of a transformed KL divergence for every given $w \in (0, 1]$. Similar to the KL divergence, this new transformed divergence is also a pre-metric on the space of probability measures. The transformed KL divergence is further shown to be a weighted average of two KL divergences that is bounded from above (for $w = 1$) by the KL divergence of the joint density of y_t and x_t, and bounded from below (for $w = 0$) by the conditional density of y_t given x_t. For $w \in (0, 1)$ the WML estimator converges to a pseudo-true parameter that gives more weight to the fit of the conditional model for y_t than the standard ML estimator.

Below we let p denote the true joint density of the vector $z_t := (y_t, x_t')'$, where x_t is the stacked vector of monthly variables x_t, and let $p(z_t) = p_1(y_t|x_t) \cdot p_2(x_t)$ so that p_1 denotes the true conditional density and y_t given x_t and p_2 the true marginal of x_t. Similarly, we let $q(\cdot; \psi)$ denote the joint density of z_t as defined by our parametric model under $\psi \in \Psi$, and let $q_1(\cdot; \psi)$ and $q_2(\cdot; \psi)$ be the counterparts of p_1 and p_2 for the parametric model density. Finally, given any two densities a and b, we let $\mathrm{KL}(a, b)$ denote the KL divergence between a and b.

Theorem 4.3. (Consistency) *Let $\{x_t\}$ and $\{y_t\}$ be SE and satisfy $\mathbb{E}|x_t|^2 < \infty$ and $\mathbb{E}|y_t|^2 < \infty$. Furthermore, let the conditions of Theorem 4.1 hold and suppose that*

$$\mathcal{L}_\infty(\psi_0^*(w), w) > \mathcal{L}_\infty(\psi, w) \ \forall \ \psi \neq \psi_0^*(w)$$

and there exists a unique SE sequence such that

$$\|\tilde{f}_t(\tilde{f}_1^*) - \tilde{f}_t\|_\Psi \overset{e.a.s.}{\to} 0 \ \ for \ every \ initialization \ \tilde{f}_1^* \ \ as \ t \to \infty \ \ with \ \ \mathbb{E}|\tilde{f}_t|^2 < \infty.$$

Then

$$\hat{\psi}_T(w, \tilde{f}_1^*) \xrightarrow{a.s.} \psi_0^*(w) \quad as\ T \to \infty$$

for any initialization $\tilde{f}_1^*$ *and any weight* $w \in (0, 1]$. *Furthermore, the pseudo-true parameter* $\psi_0^*(w)$ *minimizes a transformed KL divergence*

$$\mathrm{TKL}_w\big(q(\cdot;\psi), p\big) \;=\; \mathrm{KL}\big(q_1(\cdot;\psi), p_1\big) + w\mathrm{KL}\big(q_2(\cdot;\psi), p_2\big)$$

which is a pre-metric on the space of distributions satisfying for any $w \in (0, 1]$,

$$\mathrm{TKL}_1\big(q(\cdot;\psi), p\big) = \mathrm{KL}\big(q(\cdot;\psi), p\big) \quad , \quad \mathrm{TKL}_0\big(q(\cdot;\psi), p\big) = \mathrm{KL}\big(q_1(\cdot;\psi), p_1\big) \ ,$$

$$\mathrm{KL}\big(q_1(\cdot;\psi), p_1\big) \;\le\; \mathrm{TKL}_w\big(q(\cdot;\psi), p\big) \;\le\; \mathrm{KL}\big(q(\cdot;\psi), p\big) \ ,$$

$$\textit{and} \quad \mathrm{TKL}_w\big(q(\cdot;\psi), p\big) = 0 \quad \textit{if and only if} \quad \mathrm{KL}\big(q_1(\cdot;\psi), p_1\big) = 0.$$

Theorem 4.4 establishes the asymptotic normality of the WML estimator under the assumption that the dynamic factor model is correctly specified. Below we let $\mathcal{J}(\psi_0, w) := (\partial \mathbb{E}\ell_t(\psi_0, w)/\partial\psi) \times (\partial \ell_t(\psi_0, w)/\partial\psi)'$ denote the expected outer product of gradients and $\mathcal{I}(\psi_0, w) := \partial^2 \mathbb{E}\ell_t(\psi_0, w)/\partial\psi\partial\psi'$ be the Fisher information matrix. The asymptotic normality proof is written for filters whose derivative processes are asymptotically SE and have bounded moments; see Blasques et at. (2014) for a wide range of observation-driven filters satisfying such conditions. Below, $\{\tilde{df}_t(\tilde{df}_1^*)\}$ and $\{d\tilde{df}_t(d\tilde{df}_1^*)\}$ denote the first and second derivatives of the filter w.r.t. the parameter vector ψ, initialized at $\tilde{df}_1^*$ and $d\tilde{df}_1^*$, respectively. Their SE limits are denoted $\{\tilde{df}_t\}$ and $\{d\tilde{df}_t\}$. Note that asymptotic normality result holds for any weight $w \in (0, 1]$, but the asymptotic distribution of the WML estimator depends on the choice of weight w.

Theorem 4.4. (Asymptotic Normality) *Let the conditions of Theorem 4.2 hold and* ψ_0 *be a point in the interior of* Ψ. *Suppose furthermore that there exists a unique SE sequence* $\{\tilde{df}_t\}$ *such that*

$$\|\tilde{df}_t(\tilde{df}_1^*) - \tilde{df}_t\|_\Psi \xrightarrow{e.a.s.} 0 \ \forall\ \tilde{df}_1^* \quad as \quad t \to \infty \qquad with \qquad \mathbb{E}|\tilde{df}_t|^4 < \infty$$

and a unique SE sequence $\{d\tilde{df}_t\}$ *such that*

$$\|d\tilde{df}_t(d\tilde{df}_1^*) - d\tilde{df}_t\|_\Psi \xrightarrow{e.a.s.} 0 \ \forall\ d\tilde{df}_1^* \quad as \quad t \to \infty \qquad with \qquad \mathbb{E}|d\tilde{df}_t|^2 < \infty.$$

Then, for every $\tilde{f}_1^*$ *and every* $w \in (0, 1]$, *the ML estimator* $\hat{\psi}_T(\tilde{f}_1^*)$ *satisfies*

$$\sqrt{T}\big(\hat{\psi}_T(\tilde{f}_1^*, w) - \psi_0\big) \xrightarrow{d} N\Big(0 \ , \ \mathcal{I}^{-1}(\psi_0, w)\mathcal{J}(\psi_0, w)\mathcal{I}^{-1}(\psi_0, w)\Big) \quad as \quad T \to \infty.$$

Naturally, we can extend the asymptotic normality results to the misspecified dynamic factor model by centering the WML estimator at the pseudo-true parameter $\psi_0^*(w)$.

Theorem 4.5. (Asymptotic Normality) *Let the conditions of Theorem 4.3 hold and* $\psi_0^*(w)$ *be a point in the interior of* Ψ*. Suppose further that* $\{x_t\}$ *and* $\{y_t\}$ *are SE and satisfy* $\mathbb{E}|x_t|^4 < \infty$ *and* $\mathbb{E}|y_t|^4 < \infty$ *and there exists a unique SE sequence* $\{\tilde{df}_t\}$ *such that*

$$\|\tilde{df}_t(\tilde{df}_1^*) - \tilde{df}_t\|_\Psi \xrightarrow{e.a.s.} 0 \ \forall \ \tilde{df}_1^* \ \ as \ \ t \to \infty \quad with \quad \mathbb{E}|\tilde{df}_t|^4 < \infty$$

and a unique SE sequence $\{\tilde{ddf}_t\}$ *such that*

$$\|\tilde{ddf}_t(\tilde{ddf}_1^*) - \tilde{ddf}_t\|_\Psi \xrightarrow{e.a.s.} 0 \ \forall \ \tilde{ddf}_1^* \ \ as \ \ t \to \infty \quad with \quad \mathbb{E}|\tilde{ddf}_t|^2 < \infty.$$

Then, for every $\tilde{f}_1^*$ *and every* $w \in (0, 1]$*, the ML estimator* $\hat{\psi}_T(w, \tilde{f}_1^*)$ *satisfies*

$$\sqrt{T}\big(\hat{\psi}_T(\tilde{f}_1^*) - \psi_0^*(w)\big) \xrightarrow{d} N\Big(0 \ , \ \mathcal{I}^{-1}(\psi_0^*(w), w)\mathcal{J}(\psi_0^*(w), w)\mathcal{I}^{-1}(\psi_0^*(w), w)\Big) \ \ as \ \ T \to \infty.$$

4.2.3 Selecting Optimal Weights

In this section we follow Wang and Zidek (2005) in proposing a method for estimating optimal weights that is based on cross-validation. In particular, we will focus on obtaining weights that optimize the out-of-sample forecasting performance of the variable of interest. Furthermore, we propose the use of a Diebold-Mariano test that allows us to infer if the improvements in forecasting accuracy produced by different choices of weights are statistically significant; see Diebold and Mariano (1995c). We confirm the validity of the asymptotic distribution of the Diebold-Mariano test statistic under our set of assumptions.

For the purpose of estimating w by cross-validation, we will split the sample in two parts. The first part of the sample is used to estimate the model parameters, for any given choice of w. The second part of the sample is used to evaluate the out-of-sample forecast performance of the model and select the optimal weight w. Specifically, for some given w, we first estimate the parameter vector ψ using observations from period $t = 1$ to $t = T'$. The parameter estimate, denoted $\hat{\psi}_{1:T'}(w, \tilde{f}_1^*)$, is used to produce a one-step ahead prediction $\hat{y}_{T'+1}(\hat{\psi}_{1:T'}(w, \tilde{f}_2))$ for the related variable. Next, we obtain an estimate $\hat{\psi}_{2:T'+1}(w, \tilde{f}_2)$ using observations from period $t = 2$ to $t = T' + 1$ and produce another one-step ahead prediction $\hat{y}_{T'+2}(\hat{\psi}_{2:T'+1}(w, \tilde{f}_2))$. We repeat this procedure and obtain $H = T - T' - 1$ one-step ahead predictions using recursive samples, each based on the previous T' observations, as illustrated below,

$$\begin{array}{cccccccccccc}
y_1 & y_2 & y_3 & \cdots & & \cdots & y_{T'} & \boxed{\hat{y}_{T'+1}} & & & & \\
 & y_2 & y_3 & \cdots & & \cdots & y_{T'} & y_{T'+1} & \boxed{\hat{y}_{T'+2}} & & & \\
 & & & & & \cdots & & & \cdots & & & \\
 & & & y_H & & \cdots & y_{T'} & y_{T'+1} & y_{T'+2} & \cdots & y_{T'+H} & \boxed{\hat{y}_{T'+H+1}}.
\end{array}$$

The one-step ahead forecasts can effectively be written as a function of w since the WML estimator $\hat{\psi}_{2:T'+1}$ maps every weight w to a point in the parameter space that defines a forecast value $\hat{y}_{T'+i}(w) \equiv \hat{y}_{T'+i}(\hat{\psi}_{i:T'+i}(w))$.

Finally, we define the H out-of-sample one-step ahead forecast errors as follows

$$e_i(w) = \hat{y}_{T'+i}(w) - y_{T'+i} \quad , \quad i = 1, ..., H$$

and use these to obtain a cross-validation criteria for selecting the weight w that minimizes the one-step ahead mean squared forecast error ($\text{MSE}_1(w)$)

$$\hat{w}_H = \arg \min_{w \in [0,1]} \frac{1}{H} \sum_{i=1}^{H} e_{+1}^{i}(w)^2 = \arg \min_{w \in [0,1]} \text{MSE}_1(w)$$

Naturally, the criterion can be easily redesigned for w to minimize the h-step ahead forecast error (MSE_h). Since w is directly chosen to minimize the forecast error, it is clear that any estimate $\hat{w}_H \neq 1$ will only occur if the WML estimator can improve the error compared to the ML estimator. However, it is important to take into account the possibility of spurious reductions in the MSE that occur only because H is small. For this reason we propose the use of a Diebold-Mariano test statistic that can be used to assess if the improvement in forecasting accuracy is statistically significant. Lemma 4.1 highlights that the asymptotic Gaussian distribution derived in Diebold and Mariano (1995c) is valid under the conditions of Theorem 4.5. The assumptions of the Diebold-Mariano test hold in the current setting, for any given pair (w, w'), since Theorem 4.5 ensures that the data is SE with four bounded moments. The squared residuals are therefore covariance stationary and so are their differences. Of course, the question whether these assumptions hold in practice is ultimately an empirical issue for which there exist tests that one may wish to employ; see also the discussion in Diebold (2012). Next we let $\bar{d}_H(w, w')$ and $\Sigma_H(d_i(w, w')$ denote the sample average and standard error based on H differences in MSE obtained under the weights w and w',

$$d_i(w, w') := e_i(w)^2 - e_i(w')^2 \qquad i = 1, ..., H.$$

Lemma 4.1. *Let the conditions of Theorem 4.5 hold. Then*

$$\bar{d}_H(w, w')/\Sigma_H(d_i(w, w') \xrightarrow{d} N(0, 1) \quad as \quad H \to \infty$$

under the null hypothesis $H_0 : \mathbb{E}d_i(w, w') = 0$*, and*

$$\bar{d}_H(w, w')/\Sigma_H(d_i(w, w') \to \infty \quad as \quad H \to \infty$$

under the alternative hypothesis $H_1 : \mathbb{E}d_i(w, w') > 0$.

We stress that the Diebold-Mariano test is the most natural tool for comparing the forecasting performance of our model under any two WML estimates. The more recent tests proposed in the literature, for example, West (1996b) and Clark and McCracken (2001b, 2015)Clark and McCracken (2015) are not appropriate for our comparisons as they focus on testing the forecasting performance of different models evaluated at their pseudo-true parameters, rather than testing different forecasts; see also the discussions in Giacomini and White (2006b) and Diebold (2012).

4.2.4 Small Sample Properties of WML: A Monte Carlo Study

In our Monte Carlo study, we investigate the finite sample effects of different choices for the value of W on the in-sample fit for two different data generation processes (DGPs). The first DGP for $z_t = (y_t', x_t')'$ is the dynamic factor model (4.1) – (4.2) and the second DGP is a stationary vector autoregressive (VAR) model. The panel dimension for y_t is $N_y = 1$ (we have an univariate time series y_t) and we consider three different panel dimensions for x_t, specifically $N_x = 2$, $N_x = 5$ and $N_x = 10$. The time series length in all simulations is set to $T = 120$. The first 80 observations are used for parameter estimation while the remaining 40 observations are used for one-step ahead forecast evaluation. The mean squared error (MSE) of the 40 forecast errors of the univariate time series y_t is used for the selection of the optimal value for W in the WML method. In our Monte Carlo study we only consider these 40 "in-sample" forecasts.

The first DGP is the DFM (4.1) – (4.2) with a single dynamic factor f_t, that is $r = 1$. The parameter matrices of the DFM are chosen as follows: the factor loading matrix is given by $\Lambda = (1, 1, 1/2, 1/3, \ldots, 1/N_x)'$, the diagonal elements of Σ_ε are set to 0.5 and the persistence coefficient for the single factor f_t is given by where $\Sigma_\varepsilon = 0.8$. The second DGP for z_t is the stationary vector autoregressive process of order 1, VAR(1), and is given by $z_t = \Psi z_{t-1} + \xi_t$ with normally independently distributed $N_z \times 1$ disturbance vector ξ_t, that has mean zero and variance matrix Σ_ξ, and with $N_z \times N_z$ autoregressive coefficient matrix Ψ. The stationary conditions with respect to Ψ apply. The parameter matrices of the VAR(1) are chosen as follows: the matrix Ψ is an upper-triangular matrix with its diagonal elements equal to 0.8, its upper-triangular elements are randomly and uniformly chosen within the range $(-0.5, 0.5)$, and the variance matrix Σ_ξ is set equal to $0.5 \cdot I_{N_z}$. These choices for the parameters imply a stationary VAR(1) model.

In our simulation experiment, we consider two different settings of correct specification and misspecification. In the first case, we adopt the DFM as the DGP and we consider the same model for estimation and forecasting. We generate three databases of 500 vector time series z_t with $N_y = 1$, and each with $N_x = 2$, $N_x = 5$ and $N_x = 10$. For each simulated vector time series z_t, we estimate the parameters, using the first 80 observations, via the WML method for a range of values for W, including $1, \ldots, 5, 10, 50$ and 100. Notice that for $W = 1$, the WML method reduces to the ML method. Based on the parameter estimates and on the next 40 observations, we compute the mean

squared error (MSE) of the one-step ahead forecast errors for the target variable y_t. The forecast errors are obtained from the Kalman filter. Under correct specification, we expect that increasing the value of W will not improve the forecasting accuracy for the variable of interest y_t. Theorem 4.2 implies that asymptotically the different values of W must yield the same results since the WMLE is consistent to the true parameter for any W. Any improvements in the correct specification setting are thus only finite sample improvements.

In the second case, the VAR(1) model is the DGP as described above while the DFM is considered for estimation and forecasting. Similarly as in the first case, we generate three databases of 500 vector time series from the VAR(1) process, each with $N_z = 3$, $N_z = 6$ and $N_z = 11$. Subsequently, we estimate the DFM parameters with the values of W as given above. In the misspecification case, we also consider the W values 250, 500 and 1000, since we expect that an increasing weight W will be beneficial for the forecasting accuracy of y_t. Theorem 4.3 implies that such large improvements are explained by the fact that we can use the weight W to let the estimated parameter vector converge to the pseudo-true parameter value.

In Table (4.1), for each entry, we present the average of the MSE for the 500 generated time series of the target variable y_t, scaled by the average MSE obtained by the ML method, for the corresponding case, that is $W = 1$. From the misspecification case (right-hand side panel), we learn that increasing W leads to a better in-sample forecasting accuracy for y_t, for all three dimensions N_x. It is not necessary to choose a very large value of W. The improvements in the average MSE for larger W appear to converge to some upper limit, for all three dimensions N_x. Furthermore, we observe that more gains are made when more variables are included in the model so that the misspecification is more pronounced as the number of factors remains one. The correct specification case (left-hand side panel of Table (4.1) reveals that improvements of the in-sample forecasting accuracy are negligible for an increasing W. For instance, the value of MSE is the smallest when $W = 3$ for $N_x = 5$, while the improvement is only about 0.11% compared to the benchmark of $W = 1$. Overall, a larger value for W does not lead to a better forecasting accuracy for y_t when the model is correctly specified.

In panel I of Table (4.2), we present the frequencies of optimal W values in 500 Monte Carlo simulations. The results confirm the findings from Table (4.1). Under correct specification, the weights W that are close to unity provide more accurate in-sample forecasts. On the other hand, when the model is misspecified, the results suggest that we need to choose a large W in order to guarantee a better in-sample forecasting accuracy of y_t. In panel II of Table (4.2) we also report the sample rejection rates of the Diebold-Mariano test at 90% confidence level over the 500 simulations for the forecasts obtained from WML parameter estimates against the forecasts obtained from ML parameter estimates. Under correct model specification, the rejection rate can be viewed as the size of the DM test in our setting. While under misspecification, the rejection rate can be viewed as the power of the DM test. For $N_x = 2$, the rejection

rate of the DM test is 4.4% and the test is clearly undersized. However, for $N_x = 5$ and $N_x = 10$, the rejection rates are 8.4% and 11.8%, respectively. The results roughly suggest that we obtain the correct size when N_x gets larger. On the other hand, the rejection rate of the DM test is 91.4% when $N_x = 2$ in the misspecification case. The rejection rate increases when we include more variables in x_t. These results indicate that the power of the test is large when N_x gets larger. Hence we support the conclusion from Stock and Watson (2002a) that large-scale dynamic factor models are preferred when accurate forecasts are required. We can conclude that for a larger dimension N_x, the DM test for the WML method is slightly oversized but the power of the DM test is strong.

4.3 Mixed frequency dynamic factor model

In empirical forecasting studies based on dynamic factor analyses, we typically analyze variables that are observed at different frequencies. More specifically, and most relevant for economic forecasting, we consider settings for which the variable of interest y_t is observed at a low frequency, quarter by quarter, while the related variables in x_t are observed at high frequencies such as monthly, weekly and daily. To accommodate mixed frequency time series panels, we need to modify the dynamic factor model (4.1) – (4.2) appropriately. We will discuss a number of solutions.

4.3.1 Low versus high frequency updating

When a time series panel of mixed frequency is analysed, one may consider the highest frequency to formulate the stationary dynamic process of the factor f_t in (4.1). For the low frequency variables, the entries corresponding to the high frequency time points, for which no observations are available, can be treated as missing observations. It is well established that the Kalman filter is designed to handle missings observations. This solution has been adopted in, among others, Mariano and Murasawa (2003) and Bańbura and Rünstler (2011), although these contributions consider different dynamic factor model specifications. An alternative solution is to specify the state vector of the underlying linear state space model as a low frequency variable. We show that the low frequency state vector can still accommodate the high frequency dynamics in the model, simultaneously with the low frequency dynamics. Also, in many cases of practical relevance, this solution can lead to a much lower computational burden. In the Online Appendix we show how monthly dynamics can be treated by a linear state space model with a low frequency time index, say a quarterly index, in which the monthly entries are stacked into a quarterly vector. Similar ideas can be adopted for even higher frequency dynamics for weekly and daily observation. The idea of stacking the series observed at higher frequencies into vectors of the lowest frequency have also been explored by Ghysels (2012) in a vector autoregressive context and by Marcellino et al. (2014) for explanatory variables in their

	Correct Specification			Misspecification		
	$N_x = 2$	$N_x = 5$	$N_x = 10$	$N_x = 2$	$N_x = 5$	$N_x = 10$
W	I : In-sample forecast mean squared errors					
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.0119	0.9993	0.9964	0.9527	0.9439	0.9310
3	1.0124	0.9989	0.9960	0.9281	0.9081	0.8870
4	1.0134	1.0001	0.9961	0.9105	0.8806	0.8551
5	1.0141	1.0006	0.9974	0.8981	0.8538	0.8294
6	1.0150	1.0002	0.9972	0.8872	0.8329	0.8097
10	1.0190	1.0023	1.0001	0.8578	0.7794	0.7572
15	1.0224	1.0040	1.0024	0.8384	0.7388	0.7012
25	1.0267	1.0064	1.0043	0.8177	0.6914	0.6547
50	1.0319	1.0092	1.0062	0.7955	0.6330	0.5862
100	1.0367	1.0110	1.0084	0.7751	0.5930	0.5056
250				0.7648	0.5530	0.4551
500				0.7561	0.5359	0.4168
1000				0.7511	0.5209	0.3880

Table 4.1: Monte Carlo results I. In-sample one-step ahead forecasting of target variable y_t using the dynamic factor model with parameter estimates obtained from the WML method for different values of W. We report the mean squared error (MSE), averaged over 500 Monte Carlo simulations. The time series sample size is $T = 120$, a single target variable is considered, $N_y = 1$, and the panel dimensions for x_t are $N_x = 2, 5, 10$. In the case of 'correct specification', the DFM model (4.1) – (4.2), with $r = 1$, $\Lambda_y = 1$, $\Lambda_x = (1, 1/2, \ldots, 1/N_x)'$, $\Sigma_\varepsilon = 0.5 I_{N_z}$ and $\Phi = 0.8$, is used for data generation of y_t and x_t. In the case of 'misspecification', y_t and x_t are generated by the stationary vector autoregressive model $z_t = \Psi z_{t-1} + \xi_t$, with diagonal elements of Ψ equal to 0.8, its lower-diagonal elements are set to zero, its upper-diagonal elements are uniformly sampled within the range (-0.5,0.5), and the variance matrix of the zero-mean disturbance vector ξ_t is a diagonal matrix with its diagonal elements equal to 0.5. In both cases, the forecasts are obtained from the DFM as described above with estimated Λ, Σ_ε (diagonal) and Φ by the WML method. The smallest average MSE in each column is highlighted.

	Correct Specification			Misspecification		
	$N_x = 2$	$N_x = 5$	$N_x = 10$	$N_x = 2$	$N_x = 5$	$N_x = 10$
W	II : Frequencies of optimal W values					
1	59.8	41.2	36.8	1.4	2.2	0.8
2	15	10.0	12.0	0.4	0.2	0.4
3	1.6	4.0	7.2	0.2	0.0	0.0
4	0.8	3.0	3.8	0.4	0.2	0.0
5	1.6	4.0	3.6	0.2	0.2	0.0
6	2.8	4.0	6.6	0.0	0.0	0.0
10	2.4	4.8	7.2	1.2	0.0	0.2
15	3.4	5.2	3.4	0.4	0.0	0.2
25	3.2	3.6	5.4	0.8	0.0	0.0
50	2.4	4.8	3.6	1.8	0.6	0.0
≥ 100	7.0	15.4	10.4	93.2	96.6	98.4
	III : Rejection rates of Diebold-Mariano test					
RRR	0.044	0.084	0.118	0.914	0.928	0.934

Table 4.2: Monte Carlo results II and III. Results II : For the same Monte Carlo study as carried out for Table 4.1, we report the frequencies of optimal W values in 500 Monte Carlo simulations. The optimal W is this value of W that produces the smallest MSE within each simulation; the WML method is used for the estimation of the parameters in the DFM from which the in-sample one-step ahead forecasts are computed. Results III : We report the finite sample realized rejection rate (RRR) of the Diebold-Mariano test in 500 Monte Carlo simulations based on the optimized value of W, see Results II. We adopt a 90% confidence level for the Diebold-Mariano test concerning forecasts obtained from WML against those from ML estimates. The results for 'correct specification can be considered as the *size* of the test and for 'misspecification' as the *power* of the test.

Bayesian regression model with stochastic volatility. In our case we adopt these ideas for all high frequency variables in a dynamic factor model.

4.3.2 Low versus high frequency updating: computing times

To illustrate the computational consequences of low versus high frequency formulations, we consider a monthly autoregressive model of order 1, AR(1) model, for the variable x_τ^m, where τ is the high frequency time index and the superscript "m" is to indicate it is a monthly variable x, that is $x_\tau^m = \varphi x_{\tau-1}^m + \varepsilon_\tau^m$ where ε_τ^m is the corresponding zero-mean disturbance and φ is the autoregressive coefficient. We can stack three consecutive values of the monthly variable x_τ^m in a specific quarter t into the quarterly 3×1 vector x_t, where t is the low frequency time index. We have

$$x_t = \begin{pmatrix} x_{t,1} \\ x_{t,2} \\ x_{t,3} \end{pmatrix} = \begin{pmatrix} x_\tau^m \\ x_{\tau+1}^m \\ x_{\tau+2}^m \end{pmatrix}. \tag{4.5}$$

where $x_{t,i}$ is the i-th element of x_t and corresponds to the value of the x variable in month i of quarter t. The high frequency dynamic process for x_τ^m can be expressed by the low frequency vector process

$$\begin{pmatrix} x_{t,1} \\ x_{t,2} \\ x_{t,3} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \phi_x \\ 0 & 0 & \phi_x^2 \\ 0 & 0 & \phi_x^3 \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \\ x_{t-1,3} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ \phi_x & 1 & 0 \\ \phi_x^2 & \phi_x & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{t,1} \\ \varepsilon_{t,2} \\ \varepsilon_{t,3} \end{pmatrix}, \tag{4.6}$$

for $t = 1, \ldots, T$, where $(\varepsilon_{t,1}, \varepsilon_{t,2}, \varepsilon_{t,3})'$ is similarly defined as x_t in relation to the monthly disturbance ε_τ^m. The Online Appendix provides the derivation of this formulation and more detailed general results, such as for autoregressive models of higher orders p, that is AR(p) models. The dynamic properties of the monthly series x_τ^m are not altered by this quarterly formulation: the quarterly vector process is just a different formulation of the monthly process. Whether the monthly dynamic process is represented by x_τ^m or by the stacked quarterly 3×1 vector x_t, or even by a yearly 12×1 vector, it also has no effect on the value of the loglikelihood function for a given parameter vector. The low and high frequency representations are observationally equivalent as the derivations in the Online Appendix only rely on equalities. In all cases, the Kalman filter can be used for loglikelihood evaluation. Therefore the maximized loglikelihood value and the parameter estimates are the same for low and high frequency representations of autoregressive, or any other linear, dynamic processes.

However, the different representations have an effect on computing times. For example, for 100 years of data and for a monthly representation, we have a time series dimension of $T = 1200$. When the data is stacked into quarterly 3×1 vectors we have $T = 400$ and with yearly 12×1 vectors we only have $T = 100$. On the other hand,

the stacked vector x_t will have a larger dimension. Hence the different representations will have an effect on the Kalman filter computations when evaluating the loglikelihood function. To illustrate this, we have evaluated the loglikelihood value 10,000 times for simulated AR(p) models of a length of 1,000 years, for different orders p and using different representations: daily, weekly, monthly, quarterly and yearly. For example, for a weekly AR(p) process, time series are generated consisting of $T = 52,000$ weekly observations. The loglikelihood value is then calculated 10,000 times using the parameter values that maximizes the loglikelihood function, for different lower frequency representations. We have verified that likelihood evaluations for these different representations resulted in the same value.

The computing times for different combinations of AR(p) processes and frequencies are presented in Table (4.3). We focus here on the results for weekly and daily autoregressive processes; the results for monthly processes are presented in the Online Appendix. It is clear that for weekly and daily AR(1) and AR(2) processes, the representations based on weekly and daily models, respectively, are computationally most efficient. For these cases a lower-dimensional stacked vector outweighs the fact that the Kalman filter has to go through 52,000 (weekly) and 364,000 (daily) iterations instead of 1,000 iterations in the yearly representations. But for a weekly AR(3) process, or for any order $p > 2$, the 13-monthly representation (we assume that each month consists of 4 weeks) leads to a faster computation of the likelihood function. Here the smaller time dimension is beneficial while the size of the stacked vectors are of the same size given the number of lags that need to be accommodated in the state vector. Similar effects take place when p increases further and even the yearly representation become computationally more efficient. The results for weekly and daily processes in Table (4.3) clearly illustrate that stacking observations into low frequency vectors can lead to large computational gains, especially when many lagged dependent variables are part of the model. The considered lag lengths may appear in some cases artificially large in Table (4.3), but we notice that a yearly AR(1) process can only be represented by a weekly AR(52) process.

4.3.3 Mixed frequency dynamic factor model: low frequency updating

We provide the details of the low frequency representation of a dynamic factor model that contains both low and high frequency variables. To keep the discussion simple, and tailored to the DFM that we adopt for our empirical study in Section 4.4, we consider the DFM for the quarterly target variable y_t, with $N_y = 1$, and for a monthly $N_x \times 1$ vector x_τ^m, with quarterly index t and monthly index τ. We define the $(3 \cdot N_x) \times 1$ vector x_t in terms of x_τ^m as in (4.5).

The dynamic factor model (4.1) is formulated for a quarterly time index t. In case the autoregressive process of the dynamic factor is formulated as a monthly process, we

	Weekly time series					Daily time series			
p	Week	"Month"	Quarter	Year	p	Day	Week	"Month"	Quarter
1	2,5	3,5	6,9	69,9	1	2,6	3,9	15,3	108,4
2	3,4	3,8	7,5	72,5	2	3,6	4,2	16,6	112,7
3	4,4	4,0	8,3	76,1	3	4,6	4,5	17,7	116,4
6	14,0	7,8	10,6	85,4	6	14,6	7,0	21,8	130,7
7	18,1	10,1	11,5	87,3	7	18,8	9,2	29,7	132,0
8	24,3	11,9	12,2	89,9	8	24,9	11,5	24,4	135,3
11	47,0	22,4	14,4	100,2	18	379,4	57,2	36,6	174,5
12	54,2	24,9	15,2	103,2	19	287,3	68,5	39,9	179,0
13	65,6	30,1	18,2	106,0	20	319,8	68,3	38,8	182,2
46	6504,6	609,5	325,3	209,9	58	13333,8	930,8	435,4	316,3
47	8064,1	647,5	354,2	209,3	59	16124,3	970,4	455,5	318,6
48	9252,8	684,8	363,3	220,5	60	15937,9	1005,2	470,9	323,1

Table 4.3: Computing times. The left panel of this table presents the total computing time that is required to filter $1,000$ weekly time series with $T = 52,000$ using Kalman filter for the corresponding weekly AR(p) model. Four different approaches are used: treating the data as weekly observations, stacking the data into 13 "monthly" 4×1 vectors, stacking the data into quarterly 13×1 vectors and stacking the data into yearly 52×1 vectors. The right panel of this table presents the total computing time that is required to filter $1,000$ daily time series with $T = 364,000$ using Kalman filter for the corresponding daily AR(p) model. Four different approaches are used: treating the data as daily observations, stacking the data into weekly 7 by 1 vectors, stacking the data into 13 "monthly" 28×1 vectors and stacking the data into quarterly 91×1 vectors. Each value presents the aggregate computing time over 1000 simulations. For each p, the shortest time of the four approaches is highlighted.

introduce the monthly $r \times 1$ vector f_τ^m for month τ. The vector autoregressive process of order p^m is given by

$$f_\tau^m = \Phi_1^m f_{\tau-1}^m + \Phi_2^m f_{\tau-2}^m + \ldots + \Phi_p^m f_{\tau-p^m}^m + \eta_\tau^m, \qquad \eta_\tau^m \sim \mathrm{N}(0, \Sigma_\eta),$$

where Φ_j^m is the autoregressive coefficient matrix for lag j and η_τ^m is a monthly disturbance vector. The stack of three consecutive months of f_τ^m, corresponding to a specific quarter t, is f_t as in (4.5). In particular, we have $f_t = (f_{t,1}', f_{t,2}', f_{t,3}')' = (f_\tau^{m\prime}, f_{\tau+1}^{m\prime}, f_{\tau+2}^{m\prime})'$. The monthly dynamic processes of the elements of the stacked vector f_t can be formulated as a quarterly vector autoregressive process, similar to the equation for f_t in (4.1). The

mixed frequency dynamic factor model is given by

$$\begin{pmatrix} y_t \\ x_{t,1} \\ x_{t,2} \\ x_{t,3} \end{pmatrix} = \begin{pmatrix} \Lambda_y & \Lambda_y & \Lambda_y \\ \Lambda_x & 0 & 0 \\ 0 & \Lambda_x & 0 \\ 0 & 0 & \Lambda_x \end{pmatrix} \begin{pmatrix} f_{t,1} \\ f_{t,2} \\ f_{t,3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t,1} \\ \varepsilon_{x,t,2} \\ \varepsilon_{x,t,3} \end{pmatrix} \tag{4.7}$$

where the $N_z \times r$ factor loading matrix $\Lambda = [\Lambda_y', \Lambda_x']'$ is used in (4.1) and where $\varepsilon_{x,t,j}$ is the disturbance associated with $x_{t,j}$ for month j in quarter t. The similarity with (4.1) becomes even clearer by formulating the model as

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \iota_3' \otimes \Lambda_y \\ I_3 \otimes \Lambda_x \end{bmatrix} f_t + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix}, \tag{4.8}$$

where f_t is the $(3 \cdot r) \times 1$ vector of monthly factors and $\varepsilon_{x,t} = (\varepsilon_{x,t,1}', \varepsilon_{x,t,2}', \varepsilon_{x,t,3}')'$, with ι_k being the $k \times 1$ vector of ones. The dynamic specification for f_t, implied by any linear dynamic process for the monthly vector f_τ^m, can be formulated in state space form; see the Online Appendix. More general specifications can also be considered. For example, we can replace $\iota_3' \otimes \Lambda_y$ by a matrix with three different loading matrices for each month, that is $\Lambda_y = [\Lambda_{y,1}\ ,\ \Lambda_{y,2}\ ,\ \Lambda_{y,3}]$, where $\Lambda_{y,i}$ is the $N_y \times r$ factor loading matrix for month i. Other mixed frequencies than for monthly and quarterly variables can be considered. For example, a mix of yearly and weekly variables can be jointly modelled using the above stacking approach in a similar way.

Alternatively, we can specify the dynamic factor in our mixed frequency model as a quarterly process rather than a monthly process. In this case, the vector f_t is not a stacked vector; it just contains the r dynamic factors. For example, we can specify the dynamic process for f_t by the quarterly vector autoregressive model of order p, VAR(p),

$$f_t = \Phi_1 f_{t-1} + \ldots + \Phi_p f_{t-p} + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \Sigma_\eta),$$

with autoregressive coefficient matrix Φ_j, for lag $j = 1, \ldots, p$ in quarters. The observation equation of the dynamic factor model for quarterly factors is then simply given by

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \Lambda_y \\ \iota_3 \otimes \Lambda_x \end{bmatrix} f_t + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix}, \tag{4.9}$$

In this case, we have f_t being applicable to all three months in quarter t.

4.4 Empirical study: forecasting U.S. GDP growth

We present the results of our empirical study where we investigate whether the WML method is able to improve the nowcasting and forecasting of a relevant economic time series, in our case quarterly growth in U.S. gross domestic product, by means of a mixed frequency dynamic factor model.

4.4.1 Data

In our empirical analysis, we have constructed a database of 15 U.S. economic time series that are listed in Table (4.4). The target series is quarterly gross domestic product (GDP) growth. The time index t refers to a quarterly index. The other economic time series are observed at a monthly frequency except the series "initial unemployment claims" that is observed at a weekly frequency. The data is collected from the "FRED2" website. Our dataset can be regarded as a subset of the data used in the study by Bańbura et al. (2013). The time series length is 40 years, starting from 1970 and ending at 2009.

We consider three different panel selections from the database that are included in the monthly vector x_τ^m and the weekly variable $x_{\tau^*}^w$, where τ is the monthly index and τ^* is the weekly index. We assume that we have 52 of 53 weeks in one year; mostly we have 13 weeks in one quarter but occasionally we have 14 weeks in one quarter. Hence the dimension of x_t can vary in some occurrences. The three different selections are:

(i) The first four monthly related variables with labels EMP, DPI, IPI and SLS; see Table (4.4). This selection of variables is similar to the one considered by Mariano and Murasawa (2003) and Bańbura et al. (2013).

(ii) All monthly related variables listed in Table (4.4). This selection of 13 series is a subset of the 16 monthly variables used in the analysis of Bańbura et al. (2013).

(iii) The first four monthly related variables plus the last weekly variable IUC listed in Table (4.4). This selection mimics to some extent the analysis of Aruoba et al. (2009) where also quarterly, monthly and weekly variables are used.

All time series are transformed and/or demeaned so that no intercept coefficients are required in the model. Detected outliers in each series are replaced by their median values of the previous five observations; here we follow Stock and Watson (2005).

4.4.2 Design of the empirical study

In our empirical study, we investigate the forecasting performances of mixed frequency dynamic factor models (mfDFMs) with their parameters estimated by standard maximum likelihood (ML) and by our weighted maximum likelihood (WML) method. We aim to establish whether the WML method for different mfDFMs can improve the accuracy of nowcasts and forecasts of quarterly growth in U.S. gross domestic product (quarterly GDP growth). Our study is a stylized exercise to investigate the role of the WML method in empirical settings; we do not present results of a real-time forecasting study.

We consider three different mfDFM specifications corresponding to the three database selections described above. The mfDFM is formulated in terms of the quarterly time index t for y_t and x_t where x_t consists of stacks of three monthly observations (and, possibly, of a stack of thirteen, occasionally fourteen, weekly observations), which correspond to quarter t, for each related variable in the selected panel. Apart from the three panel selections, we also consider two variants of which the first variant formulates the

Variable	Description	Frequency	Model selections		
	Target variable y_t		(i)	(ii)	(iii)
DGP	U.S. Real GDP (billions of chained 1996)	Q	✓	✓	✓
	Related variables x_t				
EMP	Employees on non-agricultural payrolls	M	✓	✓	✓
DPI	Real disposable personal income	M	✓	✓	✓
IPI	Industrial production index	M	✓	✓	✓
SLS	Manufacturing and trade sales	M	✓	✓	✓
PMI	Purchasing manager index, manufacturing	M		✓	
UR	Unemployment rate	M		✓	
PCE	Personal consumption expenditures	M		✓	
HS	Housing starts total	M		✓	
NRS	New residential sales	M		✓	
PPI	Producer price index, finished goods	M		✓	
MNO	Manufacturer new orders	M		✓	
CPI	Consumer price index, all urban consumers	M		✓	
PFS	Philadelphia Fed survey, business conditions	M		✓	
IUC	Initial unemployment claims	W			✓

Table 4.4: Database variables. We present the labels and definitions of the quarterly, monthly and weekly variables that are used in our empirical study for the U.S. economy. All time series are available from 1970 upto 2009: fourty years of data. The third column indicates the frequency of the series: monthly (M), quarterly (Q) or weekly (W). The last three columns indicate which variables are selected for the models (i), (ii) and (iii). Data is obtained from "FRED2".

dynamic factor as a monthly variable (M) as in (4.8) and the second has the dynamic factor formulated as a quarterly variable (Q) as in (4.9). The label $mf\text{DFM}[j, q]$ refers to the mfDFM specification for panel j and frequency q, with $j = (i), (ii), (iii)$ and $q = Q, M$. In all cases, the vector autoregressive processes for the dynamic factors are based on one lag as in (4.1).

In our empirical study we assess the improvements in the forecasting and nowcasting accuracy by adopting the WML method for different weights W. The forecast evaluation period for the quarterly target time series y_t starts from 2000, quarter 1, until 2009, quarter 4, that is forty quarterly observations. For each forecast, the earlier thirty year sample is used for parameter estimation and determining the W in our WML method. The first twenty years are used for parameter estimation and the last ten years are used for determining the optimal weights, within a rolling sample scheme; see the discussion in Section 4.2.3. Then the nowcasts, the h-steps ahead (measured in months) forecasts and their subsequent errors are computed, for $h = 1, 2, 3, 6, 12$. The details are as follows.

The forecasts are calculated using a rolling window of ten years of data to estimate the parameters and determine the weights. We evaluate both the nowcasting ($h = 1, 2$ months) and the forecasting ($h = 3, 6, 12$ months) performance of all the competing models. When $h = 1$, the values of x_t are known until the first two months of the quarter that needs to be forecasted. When $h = 2$, only the first month of the quarter to be forecasted is observed. When $h = 3$, we are forecasting one quarter ahead, no observations are available for the quarter that we forecast. All values until the previous quarter are observed. Similarly, when $h = 6$ and $h = 12$ we are forecasting two and four quarters ahead, respectively.

We evaluate the accuracy of the forecasting and nowcasting of quarterly GDP growth by means of the mean squared error (MSE) based on the last fourty quarters (out-of-sample). In case of nowcasting ($h = 1, 2$) we have 40 errors, in case of one-quarter ahead forecasting ($h = 3$) we have 39 errors, etc. We compare the resulting MSEs but we also assess whether the nowcasts and forecasts based on WML estimates outperform those based on ML estimates by means of the Diebold-Mariano (DM) test.

4.4.3 Empirical results: forecast performance

The forecasting results for our mixed frequency dynamic factor models with different selections of related variables, with different dynamics for the dynamic factors, and with parameters estimated by two different methods are presented in Table (4.5). The forecast accuracy is measured by the mean squared forecasting errors (MSE). In the case of the WML estimation method, we take different integer values for W and determine its optimal value over a large range. The selected optimal value is denoted by W^{opt} and is used for the forecasting exercise. The most accurate model for each forecasting horizon is determined by the smallest MSE. In particular we focus on comparing the forecasting accuracy between ML and WML methods. Some of the differences reported in Table (4.5) are large. For example, in the case of model $mf\text{DFM}[(i), M]$ and one-step ahead forecasting, $h = 1$, the MSE drops from 0.81 when parameters are estimated by ML, to 0.55 when estimated by WML. In almost all cases, amongst all forecast horizons, the WML method leads to a smaller MSE when the same model is considered but with parameters estimated by ML. However, differences in MSEs also occur when we compare amongst different models and different frequencies in which the dynamics are modelled. We may conclude from Table (4.5) that the most accurate nowcasts ($h = 1, 2$) are obtained from the WML method applied to the larger panel model (ii) with quarterly factors. The most accurate forecasts ($h > 2$) are also obtained from WML but with monthly or weekly factors.

In earlier studies, amongst others, Aruoba et al. (2009) and Bańbura et al. (2013) argue that incorporating weekly data into a mixed frequency dynamic factor model is beneficial to nowcasting. Our results are clearly consistent with these findings. The improvements delivered by model (iii), in comparison to model (i), and especially when

j	q	W	$h=1$	$h=2$	$h=3$	$h=6$	$h=12$
(i)	M	1	0.8138	0.8139	0.8138	0.8138	0.8138
		W^{opt}	0.5493	0.8138	0.6580	0.7715	0.8100
	Q	1	0.6399	0.6654	0.7258	0.8040	0.8199
		W^{opt}	0.6084	0.8138	0.6949	0.7859	0.8173
(ii)	M	1	0.5306	0.8138	0.8138	0.8138	0.8619
		W^{opt}	0.5318	0.5760	0.6184	0.7407	0.8138
	Q	1	0.5353	0.5905	0.6686	0.7842	0.8138
		W^{opt}	0.5259	0.5597	0.6345	0.7813	0.8539
(iii)	W	1	0.7848	0.7847	0.7847	0.7847	0.7847
		W^{opt}	0.5438	0.5917	0.7847	0.7847	0.7847
	Q	1	0.7848	0.7848	0.6768	0.7848	0.8149
		W^{opt}	0.5255	0.5735	0.7847	0.7677	0.8082

Table 4.5: Forecasting comparisons for the quarterly U.S. DGP growth rate between 2000 and 2009 at forecast horizons $h = 1, 2, 3, 6, 12$ months (a 10 years rolling window). We present the mean squared forecast errors (MSEs) for the mfDFM(j, q) models with parameters estimated by maximum likelihood (weight $W = 1$) and by weighted maximum likelihood (weight $W = W^{opt}$). The optimal weight W^{opt} is determined as described in section 4.4.2. For each forecasting horizon the most accurate model (with smallest MSE) is highlighted.

parameters are estimated by WML, are substantial in all cases. We should notice that model (iii) is the same as model (i) with only the weekly variable "initial unemployment claims" added to it. The improvements are less convincing for forecasting. In case of within the year forecasting ($h = 3, 6$), model (ii) produces the smallest MSE overall. In case of one-year ahead forecasting ($h = 12$), model (iii) is preferred overall.

4.4.4 Empirical results: relative forecast performance

The relative MSE ratios between the two estimation methods are presented in Table (4.6). The entries are relative to the MSE from the ML method: a value smaller than 1 implies that the WML method provides more accurate forecasts (in terms of a smaller MSE) when compared to the MSE from the ML method, based on the same model specification for mfDFM$[j, q]$. We may conclude from the results presented in Tables (4.5) and (4.6) that the WML method improves the forecasting accuracy in most cases. The MSE reductions are even large when the ML method provides relative inaccurate forecasts. For example, the nowcasting MSE value for $h = 1$ and model mfDFM$[(iii), Q]$,

j	q	$h=1$	$h=2$	$h=3$	$h=6$	$h=12$
(i)	M	0.6749*	0.9999	0.8086	0.9480	0.9953
		(0.099)	(0.191)	(0.138)	(0.168)	(0.416)
	Q	0.9508	1.2231	0.9574	0.9775	0.9968
		(0.171)	(0.802)	(0.138)	(0.193)	(0.336)
(ii)	M	1.0023	0.7079	0.7599	0.9102	0.9442
		(0.519)	(0.103)	(0.142)	(0.213)	(0.144)
	Q	0.9826	0.9479*	0.9491*	0.9962	1.0494
		(0.344)	(0.057)	(0.094)	(0.412)	(0.869)
(iii)	W	0.6930*	0.7540	0.9999	0.9999	0.9999
		(0.052)	(0.140)	(0.108)	(0.126)	(0.136)
	Q	0.6696*	0.7308	1.1595	0.9782	0.9918
		(0.038)	(0.116)	(0.757)	(0.356)	(0.238)

Table 4.6: Relative forecasting comparisons for the quarterly U.S. DGP growth rate between 2000 and 2009 at forecast horizons $h = 1, 2, 3, 6, 12$ months (a 10 years rolling window). We present the MSE ratios of forecasts based on WML estimates relative to forecasts based on ML estimates, considering the same $mf\text{DFM}(j, q)$ model specifications. The values in parentheses are the p-values of the corresponding Diebold-Mariano test; its significance at the 10% confidence level is indicated by an asterix.

with estimated parameters by the ML method, is 0.78 which is a relatively large value amongst the MSE values for the other competing models. But when the parameters are estimated by our WML method, the MSE is reduced to 0.53, which is the smallest value amongst all other models. The forecasting accuracy is improved by about 33% and this improvement is significant at the 10% confidence level. Such results are consistent with the argument in Section 4.2. When the model is misspecified, achieving a good in-sample fit does not lead to a good out-of-sample forecasting accuracy. Our WML method is designed to perform well when the model is misspecified and interest lies in forecasting only the target variables.

When concentrating on models with quarterly factors ($q = Q$), we may conclude from the reported MSEs in Table (4.5) as follows. Amongst the three mixed frequency DFMs, as indicated in Table (4.4), model (ii) is superior in nowcasting ($h = 1, 2$) compared to models (i) and (iii), when parameters are estimated by the ML method. In case of forecasting ($h > 2$), the performances of the models (ii) and (iii) are similar overall but outperform model (i), when the ML method is used. When we focus on these comparisons for the WML method, our findings change somewhat. For both nowcasting and forecasting, the models (ii) and (iii) perform equally well overall but clearly outperform

model (i).

We adopt the Diebold-Mariano (DM) test to verify whether the forecasts obtained from the WML method are significantly more accurate compared to those obtained from the ML method, when considering the same model. The p value of the DM test for each model is also presented in Table (4.6). Almost all MSE ratios are smaller than unity for the different forecasting horizons, but for a number of entries we have forecasts from the WML method that are significantly better than those from the ML method, at the 10% confidence level. In particular, the WML method has significantly improved the forecasting accuracies for models (i) and (iii) with high frequency factors M and W, for the nowcasting horizon $h = 1$.

To obtain a more detailed picture of the empirical relevance of our WML method, we also study the improvements in forecasting accuracy for two different forecasting periods: (a) from 2000 to 2007, and (b) from 2008 to 2009. The second forecasting period (b) covers the enduring period of the financial crisis. The relative MSE ratios for the WML method with respect to the ML method are presented in Table (4.7), in the same way as in Table (4.6). The most significant forecasting accuracy improvements by the WML method are realized in the financial crisis period (b). We already have concluded that the WML method is expected to improve forecast accuracy when models are misspecified. Given that after the financial crisis the forecasting ability of most models have detoriated, the WML method can be quite effective in reducing the nowcast and forecast MSEs. It can also be concluded from the results presented in Table (4.7) that overall the WML method is most effective for short-term forecasting or nowcasting $(h = 1, 2)$. At forecasting horizon $h = 1$, many of the reported WML improvements are high or present a significant Diebold-Mariano test statistic, especially for models (i) and (iii).

4.5 Conclusions

We have introduced a new weighted maximum likelihood (WML) estimation procedure for dynamic factor models where some target variables are of key importance while the other, related, variables are only used to facilitate the forecasting or analysis of target variables. The WML method introduces variable-specific weights in the likelihood function to let the target variables have more importance than the related variables in the parameter estimation process. We have derived the asymptotic properties of the WML estimator of the parameter vector and have provided an information-theoretic characterization based on the Kullback-Leibler divergence. Furthermore, we have proposed a cross-validation method to estimate the weights that optimize the forecasting performance for the target variable. The Monte Carlo study for investigating the finite sample performance of the WML estimator highlights its good properties overall. In empirical studies, the dynamic factor model may need to be modified to accommodate a panel

of time series in which variables are observed at different seasonal frequencies such as monthly, quarterly and weekly. We have argued that a low frequency representation of the mixed frequency dynamic factor model leads to a computationally efficient treatment. The representation is highly flexible and can allow for low and high frequency dynamic factors. The empirical study shows that our WML method can lead to significant improvements in the accuracy of nowcasting and forecasting U.S. GDP growth. We expect that our proposed solutions also have consequences in other applications and in other modeling frameworks. Interesting future research may focus on applying the WML method to other models and settings as well as providing a more detailed comparison with other estimation approaches in which selections of variable equations can be penalized. For example, a possible alternative approach is to adopt Bayesian estimation methods where different prior conditions are given to parameters that are associated with different variable equations.

		(*a*) period 2000–2007					(*b*) period 2008 and 2009				
		$h=1$	$h=2$	$h=3$	$h=6$	$h=12$	$h=1$	$h=2$	$h=3$	$h=6$	$h=12$
(i)	M	0.8245*	0.9999	1.0023	0.9407	1.0326	0.5966*	0.9999	0.7058*	0.9512*	0.9783*
		(0.054)	(0.177)	(0.510)	(0.327)	(0.815)	(0.068)	(0.154)	(0.049)	(0.006)	(0.095)
	Q	0.8925*	0.9971	0.9576*	0.9840	0.9924	0.9995	1.3885	0.9572	0.9744	0.9990
		(0.035)	(0.492)	(0.065)	(0.222)	(0.186)	(0.497)	(0.811)	(0.288)	(0.226)	(0.467)
(ii)	M	0.9859	0.8912	1.0077	1.0525	0.9580*	1.0259	0.6090*	0.6189*	0.8393*	0.9383
		(0.340)	(0.135)	(0.541)	(0.660)	(0.000)	(0.598)	(0.053)	(0.013)	(0.045)	(0.288)
	Q	0.9632	0.9374*	0.9735	0.9926	1.0578	1.0089	0.9574	0.9257	0.9981	1.0456
		(0.231)	(0.056)	(0.101)	(0.165)	(0.967)	(0.546)	(0.206)	(0.112)	(0.480)	(0.668)
(iii)	W	0.9567	0.9595	0.9999	0.9999	1.0000	0.5367*	0.6426	0.9999	0.9999	0.9999
		(0.345)	(0.346)	(0.123)	(0.365)	(0.579)	(0.052)	(0.118)	(0.136)	(0.176)	(0.187)
	Q	0.9133	0.9107	0.9356	1.0410	0.9985	0.5273*	0.6353	1.3273	0.9492*	0.9885
		(0.190)	(0.162)	(0.232)	(0.639)	(0.465)	(0.044)	(0.108)	(0.856)	(0.003)	(0.221)

Table 4.7: Relative forecasting comparisons for quarterly U.S. DGP growth rate, for periods (*a*) between 2000 and 2007, and (*b*) 2008 and 2009. For further details, see Table 4.6.

Chapter 5

Summary

This thesis focuses on the topic of forecasting macroeconomic time series. First, I evaluate and compare the predictive ability among various forecasting models. The evaluation and comparison is made using a time-series of inflation. Particularly, I compare the forecasting ability of the Phillips curve against its extensions with time-varying parameters. The empirical results show that adding time-varying components, such as unobserved component and score-driven volatility, does not always lead to a forecasting model with greater precision. I find that the argument that a complex time-varying model can beat the simple linear model in terms of predictive accuracy is unfounded. In the second part, a weighted maximum likelihood estimation (WMLE) method is developed for forecasting macroeconomic time series. The WMLE is designed to optimise forecasting accuracy and recognises the potential misspecification of the model at hand. For misspecified time-series models, achieving a good in-sample fit over the entire data set becomes a separate problem from that of achieving good out-of-sample forecasting accuracy. I show that the WMLE outperforms the classical MLE in terms of forecasting accuracy when the model is misspecified. The WMLE method is applied to different macroeconomic data sets and different forecasting models.

I consider also two types of WMLE. The first weights each observation differently in the likelihood function. The second weights each variable differently in multivariate models.

For the first WMLE I show how to estimate the optimal weights using a cross-validation technique. An application of the WMLE to the U.S. industrial production index reveals that the forecasting accuracy during the latest global recession can be significantly improved by increasing the weights of observations corresponding to past re-

cession periods. This highly intuitive empirical finding shows that one should look more carefully at past recessions when the objective is to produce accurate forecast during the recent global financial crisis.

The second WMLE, is applied to high-dimensional time-series models. High-dimensional time-series models, such as dynamic factor models, are widely used to forecast some key macroeconomic variables with a panel of time series data. The WML method introduces variable-specific weights in the likelihood function to let the target variables have more importance than the related variables in the parameter estimation process. I find that, by applying the WMLE method, I can significantly improve the forecasting precision of key variables. The empirical study shows that the WML method for multivariate models can lead to significant improvements in the accuracy of nowcasting and forecasting U.S. GDP growth.

Bibliography

Akcoglu, Mustafa A., K. U. (1979). A differentiation theorem in lp. *Mathematische Zeitschrift 169*, 31–40.

Ang, A., G. Bekaert, and M. Wei (2007, May). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics 54*(4), 1163–1212.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009). Real-Time Measurement of Business Conditions. *Journal of Business & Economic Statistics 27*(4), 417–427.

Atkeson, A. and L. E. Ohanian. (2001). Are Phillips curves useful for forecasting inflation? *Quarterly Review of Federal Reserve Bank of Minneapolis* (Win), 2–11.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-casting and the real-time data flow. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, pp. 195–237. Elsevier.

Bańbura, M. and M. Modugno (2014). Maximum Likelihood Estimation Of Factor Models On Datasets With Arbitrary Pattern Of Missing Data. *Journal of Applied Econometrics 29*(1), 133–160.

Bańbura, M. and G. Rünstler (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting 27*(2), 333–346.

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society 12*(5), 788–792.

Bittanti, S. and P. Colaneri (2000). Invariant representations of discrete-time periodic systems. *Automatica 36*(12), 1777 – 1793.

Bittanti, S. and P. Colaneri (2009). *Periodic Systems: Filtering and Control*. Springer-Verlag.

Blasques, F., S. J. Koopman, and A. Lucas (2014a). Maximum likelihood estimation for generalized autoregressive score models. Technical Report 14-029/III, Tinbergen Institute Discussion Paper.

Blasques, F., S. J. Koopman, and A. Lucas (2014b). Stationarity and ergodicity of univariate generalized autoregressive score processes. *Electron. J. Statist. 8*(1), 1088–1112.

Bollerslev, T. (1986, April). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization 31*(4), 942–959.

Bräuning, F. and S. J. Koopman (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International Journal of Forecasting 30*(3), 572–584.

Chan, J. C. C., G. Koop, and S. M. Potter (2013, January). A New Model of Trend Inflation. *Journal of Business & Economic Statistics 31*(1), 94–106.

Chen, W., B. D. Anderson, M. Deistler, and A. Filler (2012). Properties of blocked linear systems. *Automatica 48*(10), 2520–2525.

Clark, T. E. and M. W. McCracken (2001a, November). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics 105*(1), 85–110.

Clark, T. E. and M. W. McCracken (2001b, November). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics 105*(1), 85–110.

Clark, T. E. and M. W. McCracken (2015). Nested forecast model comparisons: A new approach to testing equal accuracy. *Journal of Econometrics 186*(1), 160–177.

Clements, M. and D. Hendry (1998). *Forecasting Economic Time Series*. Number 9780521634809 in Cambridge Books. Cambridge University Press.

Creal, D., S. J. Koopman, and A. Lucas (2008, November). A General Framework for Observation Driven Time-Varying Parameter Models. Tinbergen Institute Discussion Papers 08-108/4, Tinbergen Institute.

Creal, D., S. J. Koopman, and A. Lucas (2013, 08). Generalized Autoregressive Score Models With Applications. *Journal of Applied Econometrics 28*(5), 777–795.

Diebold, F. X. (2012). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. NBER Working Papers 18391, National Bureau of Economic Research, Inc.

Diebold, F. X. and R. S. Mariano (1995a, July). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Diebold, F. X. and R. S. Mariano (1995b, July). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Diebold, F. X. and R. S. Mariano (1995c). Comparing predictive accuracy. *Journal of Business and Economic Statist. 13*(3), 253–63.

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

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Eguchi, S. and J. Copas (1998). A class of local likelihood methods and near-parametric asymptotics. *Journal of the Royal Statistical Society: Series B 60*(4), 709–724.

Engle, R. F. and M. W. Watson (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association 76*, 774–81.

Ferrara, L., M. Marcellino, and M. Mogliani (2013, January). Macroeconomic forecasting during the Great Recession: The return of non-linearity? CEPR Discussion Papers 9313, C.E.P.R. Discussion Papers.

Foroni, C., P. Gurin, and M. Marcellino (2015). Markov-switching mixed-frequency VAR models. *International Journal of Forecasting 31*(3), 692 – 711.

Foroni, C., M. Marcellino, and C. Schumacher (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society, Series A 178*, 57–82.

Fuhrer, J. C. (1995). The Phillips curve is alive and well. *New England Economic Review* (Mar), 41–56.

Gallant, R. and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Cambridge University Press.

Gavin, W. T. and K. L. Kliesen (2008). Forecasting inflation and output: comparing data-rich models with simple rules. *Review* (May), 175–192.

Ghysels, E. (2012). Macroeconomics and the Reality of Mixed Frequency Data.

Ghysels, E., J. B. Hill, and K. Motegi (2015). Testing for Granger causality with mixed frequency data. *Journal of Econometrics*, forthcoming.

Giacomini, R. and B. Rossi (2009). Detecting and Predicting Forecast Breakdowns. *Review of Economic Studies 76*(2), 669–705.

Giacomini, R. and H. White (2006a, November). Tests of Conditional Predictive Ability. *Econometrica 74*(6), 1545–1578.

Giacomini, R. and H. White (2006b). Tests of Conditional Predictive Ability. *Econometrica 74*(6), 1545–1578.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: the real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*(4), 665–676.

Giraitis, L., G. Kapetanios, and S. Price (2012). Adaptive Forecasting in the Presence of Recent and Ongoing Structural Change. Working Papers 691, Queen Mary University of London, School of Economics and Finance.

Goyal, A. and I. Welch (2003). Predicting the Equity Premium with Dividend Ratios. *Management Science 49*(5), 639–654.

Groen, J. J. J., R. Paap, and F. Ravazzolo (2013, January). Real-Time Inflation Forecasting in a Changing World. *Journal of Business & Economic Statistics 31*(1), 29–44.

Hansen, P. R., A. Lunde, and J. M. Nason (2011, 03). The Model Confidence Set. *Econometrica 79*(2), 453–497.

Harvey, A. (2011). Modelling the Phillips curve with unobserved components. *Applied Financial Economics 21*(1-2), 7–17.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Hu, F. and J. Zidek (1997). The relevance weighted likelihood. Unpublished.

Inoue, A., L. Jin, and B. Rossi (2014). Window Selection for Out-of-Sample Forecasting with Time-Varying Parameters. CEPR Discussion Papers 10168, C.E.P.R. Discussion Papers.

Jungbacker, B. and S. J. Koopman (2015). Likelihood-based dynamic factor analysis for measurement and forecasting. *Econometrics Journal 18*, C1–C21.

King, R. G. and M. W. Watson (1994). The post-war U.S. Phillips curve: a revisionist econometric history: response to Evans and McCallum. Working Paper Series, Macroeconomic Issues 94-17, Federal Reserve Bank of Chicago.

Koop, G. and S. Potter (2004). Forecasting in dynamic factor models using Bayesian model averaging. *Econometrics Journal 7*(2), 550–565.

Koop, G. and S. M. Potter (2007). Estimation and Forecasting in Models with Multiple Breaks. *Review of Economic Studies 74*(3), 763–789.

Koopman, S. J. and J. Durbin (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis 21*, 281–96.

Koopman, S. J., A. Lucas, and M. Scharth (2015). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics*.

Krengel, U. (1985). *Ergodic theorems*. Berlin: De Gruyter studies in Mathematics.

Marcellino, M. (2008). A linear benchmark for forecasting GDP growth and inflation? *Journal of Forecasting 27*(4), 305–340.

Marcellino, M., A. Carriero, and T. E. Clark (2014). Real-Time nowcasting with a Bayesian mixed frequency model with stochastic volatility. Working paper.

Marcellino, M., J. H. Stock, and M. Watson (2003). Macroeconomic forecasting in the euro area: country specific versus euro wide information. *European Economic Review 47*, 1–18.

Mariano, R. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics 18*, 427–443.

Markatou, M., A. Basu, and B. G. Lindsay (1997). Weighted likelihood estimating equations: The discrete case with applications to logistic regression. *Journal of Statistical Planning and Inference 57*(2), 215–232.

Markatou, M., A. Basu, and B. G. Lindsay (1998). Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association 93*(442), 740–750.

McCracken, M. W. and T. E. Clark (2006, August). The Predictive Content of the Output Gap for Inflation: Resolving In-Sample and Out-of-Sample Evidence. *Journal of Money, Credit and Banking 38*(5), 1127–1148.

Mehra, R. (1970). On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control 15*(2), 175–184.

Mittnik, S. and P. A. Zadrozny (2005). Forecasting quarterly german gdp at monthly intervals using monthly ifo business conditions data. In J.-E. Sturm and T. Wollmershauser (Eds.), *Ifo survey data in business cycle and monetary policy analysis*, pp. 19–48. Heidelberg: Physica-Verlag.

Molodtsova, T. and D. H. Papell (2009). Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics 77*(2), 167–180.

Paye, B. S. and A. Timmermann (2006). Instability of return prediction models. *Journal of Empirical Finance 13*(3), 274–315.

Plackett, R. L. (1950). Some theorems in least squares. *Biometrika 37*(1/2), 149–157.

Quah, D. and T. J. Sargent (1993). A dynamic index model for large cross sections. In J. H. Stock and M. Watson (Eds.), *Business cycles, indicators and forecasting*, pp. 285–306. Chicago: University of Chicago Press.

Rao, R. R. (1962). Relations between Weak and Uniform Convergence of Measures with Applications. *The Annals of Mathematical Statistics 33*(2), 659–680.

Rossi, B. (2013). *Advances in Forecasting under Instability*, Volume 2 of *Handbook of Economic Forecasting*, Chapter 0, pp. 1203–1324. Elsevier.

Rünstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. V. Nieuwenhuyze (2009). Short-term forecasting of GDP using large datasets: a pseudo real-time forecast evaluation exercise. *Journal of Forecasting 28*(7), 595–611.

Schinasi, G. J. and P. A. V. B. Swamy (1989). The out-of-sample forecasting performance of exchange rate models when coefficients are allowed to change. *Journal of International Money and Finance 8*(3), 375–390.

Schumacher, C. and J. Breitung (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting 24*(3), 386–398.

Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models. *Journal of the American Statistical Association 84*(405), 276–283.

Stock, J. H. and M. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*, 1167–79.

Stock, J. H. and M. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statist. 20*, 147–62.

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

Stock, J. H. and M. W. Watson (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

Stock, J. H. and M. W. Watson (1999, October). Forecasting inflation. *Journal of Monetary Economics 44*(2), 293–335.

Stock, J. H. and M. W. Watson (2002c, December). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association 97*, 1167–1179.

Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*(6), 405–430.

Stock, J. H. and M. W. Watson (2007, 02). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking 39*(s1), 3–33.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroeskedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.

Swanson, N. R. (1998, May). Money and output viewed through a rolling window. *Journal of Monetary Economics 41*(3), 455–474.

Tibshirani, R. and T. Hastie (1987). Local likelihood estimation. *Journal of the American Statistical Association 82*(398), 559–567.

Timmermann, A. (2006, January). *Forecast Combinations*, Volume 1 of *Handbook of Economic Forecasting*, Chapter 4, pp. 135–196. Elsevier.

van der Vaart, A. W. (2000, June). *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press.

Wang, X., C. van Eeden, and J. V. Zidek (2004). Asymptotic properties of maximum weighted likelihood estimators. *Journal of Statistical Planning and Inference 119*(1), 37–54.

Wang, X. and J. V. Zidek (2005). Selecting likelihood weights by cross-validation. *The Annals of Statistics 33*(2), 463–500.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression. *Journal of Econometrics 23*, 385–400.

West, K. D. (1996a, September). Asymptotic Inference about Predictive Ability. *Econometrica 64*(5), 1067–84.

West, K. D. (1996b). Asymptotic Inference about Predictive Ability. *Econometrica 64*(5), 1067–84.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge Books. Cambridge University Press.

Wolff, C. C. P. (1987). Time-Varying Parameters and the Out-of-Sample Forecasting Performance of Structural Exchange Rate Models. *Journal of Business & Economic Statistics 5*(1), 87–97.

Appendix A

Deviations of Chapter 4

A.1 Proofs of Theorems and Propositions

Notation, variables, definitions and model formulations are all presented in the main paper.

Proposition A.1 below states well known conditions for the strict stationarity and ergodicity (SE) of the true processes $\{f_t\}_{t\in\mathbb{Z}}$, $\{x_t\}_{t\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ generated by our basic dynamic factor model (DFM), initialized in the infinite past. Below, we let $\|\cdot\|$ denote the Euclidean spectral norm, i.e. the p-norm that sets $p = 2$. The results extend naturally to other norms.

Proposition A.1. *Let $\{x_t\}_{t\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ be generated according to the DFM with*
(i) $\|\Phi\| < 1$ *and* $0 < \|\Sigma_\eta\| < \infty$*;*
(ii) $\|\Lambda_x\| < \infty$ *and* $0 < \|\Sigma_\varepsilon\| < \infty$*;*
(iii) $\|\Lambda_y\| < \infty$*.*
Then $\{x_t\}_{t\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ are SE sequences with bounded moments of any order; i.e. $\mathbb{E}|x_t|^r < \infty$ *and* $\mathbb{E}|y_t|^r < \infty\ \forall\ r > 0$*.*

Proof of Proposition A.1. Let $\{f_t(f_1^*)\}_{t\in\mathbb{N}}$ be generated according to the DFM with initialization f_1^* and $\|T_f\| < 1$, $\|R_f\| < \infty$ and $\|\Sigma_\eta^2\| < \infty$. Then by Theorem 3.1 in Bougerol (1993), $\{f_t(f_1^*)\}_{t\in\mathbb{N}}$ converges to an SE sequence $\{f_t\}_{t\in\mathbb{Z}}$ satisfying $\mathbb{E}|f_t|^r < \infty\ \forall\ r$. Uniqueness of the limit SE sequence is obtained in Straumann and Mikosch (2006). Furthermore, since $\{f_t\}$ is a linear Gaussian process with $\mathbb{E}|x_t|^r < \infty\ \forall\ r > 0$. The bounds $\|\beta_x\| < \infty$ and $0 < \sigma_\epsilon^2 < \infty$, together with the iid Gaussian nature of the innovations

$\{\epsilon_t\}$ ensure that $\{x_t\}$ is SE and Gaussian with bounded moments of any order. Similarly, the bounds $\|\beta_y\| < \infty$ and $0 < \sigma_\zeta^2 < \infty$ and the iid Gaussian nature of $\{\zeta_t\}$ ensure the SE linear Gaussian nature of $\{y_t\}$ with $\mathbb{E}|y_t|^r < \infty \; \forall \; r > 0$. □

Theorem A.1 ensures the existence of the weighted maximum likelihood (WML) estimator as a random variable that takes values in the arg max set of the random likelihood function.

Theorem A.1. (Existence) *For given $w \in [0,1]$, let $(\Psi, \mathfrak{B}(\Psi))$ be a compact measurable space. Then there exists a.s. a measurable map $\hat{\psi}_T(w, \tilde{f}_1^*) : \Omega \to \Psi$ satisfying*

$$\hat{\psi}_T(w, \tilde{f}_1^*) \in \arg\max_{\psi \in \Psi} \mathcal{L}_T(\psi, w, \tilde{f}_1^*),$$

for all $T \in \mathbb{N}$ and every filter initialization $\tilde{f}_1^*$.

Proof of Theorem A.1. For every given $w \in [0,1]$, the random likelihood function $\mathcal{L}_T(\cdot, \tilde{f}_1^*)$ is trivially almost surely continuous on Ψ. The compactness of Ψ implies by Weierstrass' theorem that the arg max set is almost surely non-empty. As a result, $\hat{\psi}_T$ exists almost surely $\forall \; T \in \mathbb{N}$. The continuity of the likelihood function in f_t, x_t and y_t for every $\psi \in \Psi$ implies also measurability of the likelihood under the Borel σ-algebra. For every given $w \in [0,1]$ the measurability of the WML estimator can now be obtained by application of Theorem 2.11 of White (1994) or Lemma 2.1 and Theorem 2.2 in Gallant and White (1988). □

Theorem A.2 establishes the strong consistency of the WML estimator of the true parameter vector $\psi_0 \in \Psi$ for any choice of weight $w \in (0,1]$ for the likelihood.

Theorem A.2. (Consistency) *Let $\{x_t\}$ and $\{y_t\}$ be generated by the dynamic factor model defined in the DFM under some $\psi_0 \in \Psi$, and suppose that the conditions of Propositions A.1 and Theorem A.1 hold. Suppose furthermore that*

$$\mathcal{L}_\infty(\psi_0, 1) > \mathcal{L}_\infty(\psi, 1) \; \forall \; \psi \neq \psi_0$$

and there exists a unique SE sequence such that

$$\|\tilde{f}_t(\tilde{f}_1^*) - \tilde{f}_t\|_\Psi \overset{e.a.s.}{\to} 0 \; \forall \; \tilde{f}_1^* \quad as \quad t \to \infty \qquad with \qquad \mathbb{E}|\tilde{f}_t|^2 < \infty.$$

Then the WML estimator $\hat{\psi}_T(w, \tilde{f}_1^*)$ satisfies

$$\hat{\psi}_T(w, \tilde{f}_1^*) \overset{a.s.}{\to} \psi_0 \quad as \quad T \to \infty$$

for any choice of weight $w \in (0,1]$ and any initialization $\tilde{f}_1^*$.

Proof of Theorem A.2. The consistency of the WML estimator can be obtained by appealing to the classical extremum estimation theory found e.g. in Theorem 3.4 of White (1994) or Theorem 3.3 of Gallant and White (1988). In particular, for any weight $w \in (0, 1]$ and initialization $\tilde{f}_1^*$, the consistency $\hat{\psi}_T(w, \tilde{f}_1^*) \xrightarrow{a.s.} \psi_0$ follows from the uniform convergence of the weighted likelihood

$$\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_1^*) - \mathcal{L}_\infty(\psi, w)| \xrightarrow{a.s.} 0 \ \forall \ \tilde{f}_1^* \in \mathbb{R}_+ \ \text{ as } \ T \to \infty, \tag{A.1}$$

and the identifiable uniqueness of the true parameter $\psi_0 \in \Psi$ defined e.g. in White (1994).

To establish the uniform convergence of $\mathcal{L}_T(\psi, w, \tilde{f}_1^*)$ we use the norm sub-additivity inequality

$$\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_1^*) - \mathcal{L}_\infty(\psi, w)| \ \leq \ \sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_1^*) - \mathcal{L}_T(\psi, w)| + \sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w) - \mathcal{L}_\infty(\psi, w)|$$

where $\mathcal{L}_T(\psi, w)$ denotes the likelihood evaluated at the filtered $\tilde{f}_t(\psi)$ starting in the infinite past. The term

$$\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_1^*) - \mathcal{L}_T(\psi, w)|$$

vanishes by the assumption that $\|\tilde{f}_t(\psi, w, \tilde{f}_1^*) - \tilde{f}_t(\psi)\| \xrightarrow{a.s.} 0$, the continuity of the likelihood function and the continuous mapping theorem.

The ergodic theorem for separable Banach spaces of Rao (1962) ensures that

$$\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w) - \mathcal{L}_\infty(\psi, w)| \xrightarrow{a.s.} 0$$

(see also Theorem 2.7 in Straumann and Mikosch (2006)) for the sequence $\{\mathcal{L}_T(\cdot, w)\}$ of points in $\mathbb{C}(\Psi, \mathbb{R})$ under:

(i) the SE nature of $\{\mathcal{L}_T(\cdot, w)\}_{T \in \mathbb{Z}}$ which is ensured by SE nature of $\{\tilde{f}_t\}_{t \in \mathbb{Z}}$, $\{x_t\}_{t \in \mathbb{Z}}$ and $\{y_t\}_{T \in \mathbb{Z}}$, by the continuity of and Proposition 4.3 in Krengel (1985);

(ii) the moment bound $\mathbb{E} \sup_{\psi \in \Psi} |\ell_t(\psi, w)| < \infty$ ensured by the Gaussian log likelihood under the bounded second moment of $\tilde{f}_t$, x_t and y_t.

The identifiable uniqueness of the true parameter $\psi_0 \in \Psi$, typically defined as

$$\sup_{\psi : \|\psi - \psi_0\| > \epsilon} \ell_\infty(\psi, w) < \ell_\infty(\psi_0, w) \ \forall \ \epsilon > 0$$

is ensured by the uniqueness of ψ_0, the compactness of Ψ, and the continuity of $\mathbb{E}\ell_t(\psi, w)$ on Ψ, which is obtained through the continuity of $\mathcal{L}_T$ on Ψ for every $T \in \mathbb{N}$ and the uniform convergence of the likelihood; see e.g. White (1994). The uniqueness of ψ_0 as the maximizer $\mathcal{L}_\infty(\cdot, w)$ for any $w \in (0, 1]$ is ensured by Theorem A.3 which shows that the maximizer $\mathcal{L}_\infty(\psi_0, w) = 0$ if and only if $\mathcal{L}_\infty(\psi_0, 1) = 0$. □

Theorem A.3 characterizes the limit pseudo-true parameter $\psi_0^*(w)$ as the minimizer of a transformed KL divergence for every given $w \in (0,1]$. Similar to the KL divergence, this new transformed divergence is also a pre-metric on the space of probability measures. Below we let p denote the true joint density of the vector $z_t := (y_t, x_t')'$, where x_t is the stacked vector of monthly variables x_t, and let $p(z_t) = p_1(y_t|x_t) \cdot p_2(x_t)$ so that p_1 denotes the true conditional density and y_t given x_t and p_2 the true marginal of x_t. Similarly, we let $q(\cdot;\psi)$ denote the joint density of z_t as defined by our parametric model under $\psi \in \Psi$, and let $q_1(\cdot;\psi)$ and $q_2(\cdot;\psi)$ be the counterparts of p_1 and p_2 for the parametric model density. Finally, given any two densities a and b, we let $\mathrm{KL}(a,b)$ denote the KL divergence between a and b.

Theorem A.3. (Consistency) *Let $\{x_t\}$ and $\{y_t\}$ be SE and satisfy $\mathbb{E}|x_t|^2 < \infty$ and $\mathbb{E}|y_t|^2 < \infty$. Furthermore, let the conditions of Theorem A.1 hold and suppose that*

$$\mathcal{L}_\infty(\psi_0^*(w), w) > \mathcal{L}_\infty(\psi, w) \;\forall\; \psi \neq \psi_0^*(w)$$

and there exists a unique SE sequence such that

$$\|\tilde{f}_t(\tilde{f}_1^*) - \tilde{f}_t\|_\Psi \xrightarrow{e.a.s.} 0 \;\; \textit{for every initialization } \tilde{f}_1^* \quad \textit{as} \;\; t \to \infty \quad \textit{with} \quad \mathbb{E}|\tilde{f}_t|^2 < \infty.$$

Then

$$\hat{\psi}_T(w, \tilde{f}_1^*) \xrightarrow{a.s.} \psi_0^*(w) \;\; \textit{as } T \to \infty$$

for any initialization $\tilde{f}_1^*$ and any weight $w \in (0,1]$. Furthermore, the pseudo-true parameter $\psi_0^*(w)$ minimizes a transformed KL divergence

$$\mathrm{TKL}_w\big(q(\cdot;\psi), p\big) \;=\; \mathrm{KL}\big(q_1(\cdot;\psi), p_1\big) + w\mathrm{KL}\big(q_2(\cdot;\psi), p_2\big)$$

which is a pre-metric on the space of distributions satisfying for any $w \in (0,1]$,

$$\mathrm{TKL}_1\big(q(\cdot;\psi), p\big) = \mathrm{KL}\big(q(\cdot;\psi), p\big) \quad , \quad \mathrm{TKL}_0\big(q(\cdot;\psi), p\big) = \mathrm{KL}\big(q_1(\cdot;\psi), p_1\big) \;,$$

$$\mathrm{KL}\big(q_1(\cdot;\psi), p_1\big) \;\leq\; \mathrm{TKL}_w\big(q(\cdot;\psi), p\big) \;\leq\; \mathrm{KL}\big(q(\cdot;\psi), p\big) \;,$$

$$\textit{and} \quad \mathrm{TKL}_w\big(q(\cdot;\psi), p\big) = 0 \quad \textit{if and only if} \quad \mathrm{KL}\big(q_1(\cdot;\psi), p_1\big) = 0.$$

Proof of Theorem A.3. The consistency statement follows by the same steps as the proof of Theorem A.2 with the exception that the SE nature of $\{y_t\}$ is assumed rather than derived through Proposition A.1.

Let z_t denote a d_z-variate random vector with joint density $p(z_t)$. Furthermore, consider a family of parametric joint densities indexed by the parameter vector ψ, defined as $Q(\Psi) := \{q(z_t;\psi), \psi \in \Psi\}$. Note that it is possible but not necessary that $p(z_t) \in Q(\Psi)$. If $\psi_0^*(1)$ is the pseudo-true parameter that maximizes the limit log likelihood function

$$\psi_0^* := \arg\max_{\psi \in \Psi} \mathbb{E}_0 \log q(z_t;\psi),$$

then it is well known that ψ_0^* also minimizes the the Kullback–Leibler divergence $\mathrm{KL}\big(p, q(\cdot;\psi)\big)$ between $p(z_t)$ and $q(z_t;\psi)$ because

$$\arg\max_{\psi\in\Psi} \mathbb{E}_0 \log q(z_t;\psi) = \arg\min_{\psi\in\Psi} \mathbb{E}_0 \log p(z_t) - \mathbb{E}_0 \log q(z_t;\psi) = \arg\min_{\psi\in\Psi} \mathrm{KL}\Big(p, q(\cdot;\psi)\Big).$$

Let now the joint density $q(x_t;\psi)$ be factorized into

$$q_1(z_{1,t}|z_{2:d_z,t};\psi_0) \times q_2(z_{2:d_z,t};\psi_0) := q_1(z_{1,t}|z_{2,t},...,z_{d_z,t};\psi_0) \times q_2(z_{2,t},z_{3,t},...,z_{d_z,t};\psi_0)$$

and define $\psi_0^*(w)$ as the pseudo-true parameter that maximizes the weighted limit log likelihood function

$$\psi_0^*(w) := \arg\max_{\psi\in\Psi} \mathbb{E}_0 \log q_1(z_{1,t}|z_{2:d_z,t};\psi) + w \log \mathbb{E}_0 q_2(z_{2:d_z,t};\psi).$$

Then it follows naturally that $\psi_0^*(w)$ is the minimizer of the weighted average of KL divergences

$$\psi_0^*(w) = \arg\min_{\psi\in\Psi} \mathrm{KL}(q_1,p_1) + w\mathrm{KL}(q_2,p_2)$$

because

$$\begin{aligned}
&\arg\max_{\psi\in\Psi} \Big[\mathbb{E}_0 \log q_1(z_{1,t}|z_{2:d_z,t};\psi) + w \log \mathbb{E}_0 q_2(z_{2:d_z,t};\psi)\Big] \\
&= \arg\max_{\psi\in\Psi} \Big[\mathbb{E}_0 \log q_1(z_{1,t}|z_{2:d_z,t};\psi) - \mathbb{E}_0 \log p_1(z_{1,t}|z_{2:d_z,t}) \\
&\qquad + w \log \mathbb{E}_0 q_2(z_{2:d_z,t};\psi) - w \log \mathbb{E}_0 p_2(z_{2:d_z,t})\Big] \\
&= \arg\min_{\psi\in\Psi} \Bigg[\mathbb{E}_0 \log p_1(z_{1,t}|z_{2:d_z,t}) - \mathbb{E}_0 \log q_1(z_{1,t}|z_{2:d_z,t};\psi) \\
&\qquad + w\Big(\log \mathbb{E}_0 p_2(z_{2:d_z,t}) - \log \mathbb{E}_0 q_2(z_{2:d_z,t};\psi)\Big)\Bigg] \\
&= \arg\min_{\psi\in\Psi} \Big[\mathrm{KL}(q_1(\cdot;\psi),p_1) + w\mathrm{KL}(q_2(\cdot;\psi),p_2)\Big].
\end{aligned}$$

Clearly, if $w = 1$, then we obtain the usual ML pseudo-true parameter since

$$\psi_0^*(1) = \arg\min_{\psi\in\Psi} \mathrm{KL}(q_1(\cdot;\psi),p_1) + \mathrm{KL}(q_2(\cdot;\psi),p_2) = \arg\min_{\psi\in\Psi} \mathrm{KL}(q(\cdot;\psi),p).$$

For $w = 0$ we obtain the ML estimator for the conditional model

$$\psi_0^*(0) = \arg\min_{\psi\in\Psi} \mathrm{KL}(q_1(\cdot;\psi),p_1)$$

Finally, it is also clear that the transformed KL divergence $\text{TKL}(q,p) := \text{KL}(q_1(\cdot;\psi),p_1)+ w\text{KL}(q_2(\cdot;\psi),p_2)$ satisfies

$$\text{KL}(q_1(\cdot;\psi),p(\cdot;\psi)) \leq \text{TKL}(q(\cdot;\psi),p) \leq \text{KL}(q(\cdot;\psi),p)$$

and that $\text{TKL}(q(\cdot;\psi),p)$ is a pre-metric for any $w \in (0,1]$ as it inherits positivity $\text{TKL}(q(\cdot;\psi),p) \geq 0$ from the positivity of the KL divergence, and satisfies also the identity of indiscernibles since

$$\text{TKL}(q(\cdot;\psi),p) = \text{KL}(q_1(\cdot;\psi),p_1) + w\text{KL}(q_2(\cdot;\psi),p_2) = 0$$

if and only if $q_1(\cdot;\psi) = p_1$ and $q_2(\cdot;\psi) = p_2$, and hence, if and only if $q(\cdot;\psi) = p$. □

Theorem A.4 establishes the asymptotic normality of the WML estimator under the assumption that the DFM is well specified. Below we let $\mathcal{J}(\psi_0,w) := (\partial\mathbb{E}\ell_t(\psi_0,w)/\partial\psi)\times (\partial\ell_t(\psi_0,w)/\partial\psi)'$ denote the expected outer product of gradients and $\mathcal{I}(\psi_0,w) := \partial^2\mathbb{E}\ell_t(\psi_0,w)/\partial\psi\partial\psi'$ be the Fisher information matrix. The asymptotic normality proof is written for filters whose derivative processes are asymptotically SE and have bounded moments; see Blasques et al. (2014a) for a wide range of observation-driven filters satisfying such conditions. Below, $\{\tilde{df}_t(\tilde{df}_1^*)\}$ and $\{\tilde{ddf}_t(\tilde{ddf}_1^*)\}$ denote the first and second derivatives of the filter w.r.t. the parameter vector ψ, initialized at $\tilde{df}_1^*$ and $\tilde{ddf}_1^*$, respectively. Their SE limits are denoted $\{\tilde{df}_t\}$ and $\{\tilde{ddf}_t\}$.

Theorem A.4. (Asymptotic Normality) *Let the conditions of Theorem A.2 hold and ψ_0 be a point in the interior of Ψ. Suppose furthermore that there exists a unique SE sequence $\{\tilde{df}_t\}$ such that*

$$\|\tilde{df}_t(\tilde{df}_1^*) - \tilde{df}_t\|_\Psi \overset{e.a.s.}{\to} 0 \;\forall\; \tilde{df}_1^* \;\; as \;\; t\to\infty \quad with \quad \mathbb{E}|\tilde{df}_t|^4 < \infty$$

and a unique SE sequence $\{\tilde{ddf}_t\}$ such that

$$\|\tilde{ddf}_t(\tilde{ddf}_1^*) - \tilde{ddf}_t\|_\Psi \overset{e.a.s.}{\to} 0 \;\forall\; \tilde{ddf}_1^* \;\; as \;\; t\to\infty \quad with \quad \mathbb{E}|\tilde{ddf}_t|^2 < \infty.$$

Then, for every $\tilde{f}_1^*$ and every $w\in(0,1]$, the ML estimator $\hat{\psi}_T(\tilde{f}_1^*)$ satisfies

$$\sqrt{T}\big(\hat{\psi}_T(\tilde{f}_1^*,w) - \psi_0\big) \overset{d}{\to} N\Big(0\;,\; \mathcal{I}^{-1}(\psi_0,w)\mathcal{J}(\psi_0,w)\mathcal{I}^{-1}(\psi_0,w)\Big) \;\; as \;\; T\to\infty.$$

Proof of Theorem A.4. Asymptotic normality of the WML estimator can be obtained by verifying the conditions of Theorem 6.2 of White (1994):

(i) $\hat{\psi}_T(\tilde{f}_1^*,w) \overset{a.s.}{\to} \psi_0 \in \text{int}(\Psi)$;
(ii) $\mathcal{L}_T(\cdot,w,\tilde{f}_1^*) \in \mathbb{C}^2(\Psi)$ a.s.;
(iii) $\sqrt{T}\mathcal{L}_T'\big(\psi_0,w,\tilde{f}_1^*,\tilde{df}_1^*\big) \overset{d}{\to} N(0,\mathcal{J}(\psi_0,w))$;

(iv) $\sup_{\psi\in\Psi}\left\|\mathcal{L}''_T(\psi,w,\tilde{f}_1^*,\tilde{df}_1^*,d\tilde{df}_1^*)-\ell''_\infty(\psi,w)\right\|\xrightarrow{a.s.}0$;
(v) $\ell''_\infty(\psi,w)=\mathbb{E}\ell''_t(\psi,w)=\mathcal{I}(\psi,w)$ is non-singular

Conditions (i) and (ii) follow naturally from Theorem A.2, the additional assumption that $\psi_0\in\text{int}(\Psi)$ and the differentiability of the Gaussian likelihood.

Condition (iii) follows by the asymptotic SE nature of the score $\{\ell'_t(\psi,w,\tilde{f}_1^*,\tilde{df}_1^*)\}$ which is implied by the SE nature of $\{x_t\}$ and $\{y_t\}$, and the asymptotic SE nature of the filtered sequence $\{\tilde{f}_t(\psi,\tilde{f}_1^*)\}$ and its derivative $\{\tilde{df}_t(\psi,\tilde{df}_1^*)\}$. Since the limit score sequence $\{\ell'_t(\psi,w)\}$ is SE we can apply the CLT for SE martingales in Billingsley (1961) to obtain

$$\sqrt{T}\mathcal{L}'_T\big(\psi_0,w\big)\xrightarrow{d}N\big(0,\mathcal{J}(\psi_0,w)\big)\ \text{ as }\ T\to\infty, \tag{A.2}$$

where $\mathcal{J}(\psi_0,w)=\mathbb{E}(\ell'_t\big(\psi_0,w\big)\ell'_t\big(\psi_0,w\big)^\top)<\infty$. By Theorem 18.10[iv] in van der Vaart (2000) it thus follows that

$$\sqrt{T}\mathcal{L}'_T\big(\psi_0,w,\tilde{f}_1^*,\tilde{df}_1^*\big)\xrightarrow{d}N\big(0,\mathcal{J}(\psi_0,w)\big)\ \text{ as }\ T\to\infty,$$

as long as

$$\|\mathcal{L}'_T\big(\psi_0,w,\tilde{f}_1^*,\tilde{df}_1^*\big)-\mathcal{L}'_T\big(\psi_0,w\big)\|\xrightarrow{e.a.s.}0\ \text{ as }\ T\to\infty \tag{A.3}$$

since (A.3) ensures $\sqrt{T}\|\mathcal{L}'_T\big(\psi_0,w,\tilde{f}_1^*,\tilde{df}_1^*\big)-\mathcal{L}'_T\big(\psi_0,w\big)\|\xrightarrow{a.s.}0$ as $T\to\infty$. The e.a.s. convergence in (A.3) follows from

$$|f_t(\psi_0,w,\tilde{f}_1^*)-f_t(\psi_0,w)|\xrightarrow{e.a.s.}0$$

and

$$\|\tilde{df}_t(\psi_0,\tilde{df}_1^*)-\tilde{df}_t(\psi_0)\|\xrightarrow{e.a.s.}0.$$

Furthermore, since the score of the weighted likelihood is differentiable, we can use the mean-value theorem to obtain

$$\|\mathcal{L}'_T\big(\psi_0,w,\tilde{f}_1^*,\tilde{df}_1^*\big)-\mathcal{L}'_T\big(\psi_0,w\big)\|\leq\sum_j\big|d\mathcal{L}'_T\big|\big|\tilde{df}_{j,t}(\psi_0,\tilde{df}_1^*)-\tilde{df}_{j,t}(\psi_0)\big|, \tag{A.4}$$

where $\tilde{df}_{j,t}(\psi_0,\tilde{df}_1^*)$ denotes the j-th element of $\tilde{df}_t(\psi_0,\tilde{df}_1^*)$, and $d\mathcal{L}'_T$ denotes the derivative $\partial\mathcal{L}'_T\big(\psi_0,w,\tilde{f}_1^*,\tilde{df}_1^*\big)/\partial\tilde{df}_j$ evaluated at some appropriate point between $\tilde{df}_{j,t}(\psi_0,\tilde{df}_1^*)$ and $\tilde{df}_{j,t}(\psi_0)$. The bounded moments of the weighted likelihood derivatives and the e.a.s. convergence of the filtered process and its derivatives yield

$$\|\mathcal{L}'_T\big(\psi_0,w,\tilde{f}_1^*,\tilde{df}_1^*\big)-\mathcal{L}'_T\big(\psi_0,w\big)\|=\sum_j O_p(1)o_{e.a.s}(1)=o_{e.a.s.}(1). \tag{A.5}$$

Condition (iv) follows by noting that

$$\begin{aligned}\sup_{\psi\in\Psi}\|\mathcal{L}''_T\big(\psi,w,\tilde{f}_1^*,\tilde{df}_1^*,d\tilde{df}_1^*\big)-\mathcal{L}''_\infty\big(\psi,w\big)\leq\sup_{\psi\in\Psi}\|\mathcal{L}''_T\big(\psi,w,\tilde{f}_1^*,\tilde{df}_1^*,d\tilde{df}_1^*\big)-\mathcal{L}''_T\big(\psi,w\big)\|\\+\sup_{\psi\in\Psi}\|\mathcal{L}''_T\big(\psi,w\big)-\mathcal{L}''_\infty\big(\psi,w\big)\|.\end{aligned} \tag{A.6}$$

Clearly, $\sup_{\psi\in\Psi}\|\mathcal{L}''_T(\psi,w,\tilde{f}_1^*,\tilde{df}_1^*,\tilde{ddf}_1^*)-\mathcal{L}''_T(\psi,w)\|0$ as $t\to\infty$ by the continuous mapping theorem and the e.a.s. convergence of the filtered process and its derivatives, and $\sup_{\psi\in\Psi}\|\mathcal{L}''_T(\psi,w)-\mathcal{L}''_\infty(\psi,w)\|$ vanishes by a ULLN under the uniform moment bound on the weighted likelihood $\mathbb{E}\sup_{\psi\in\Psi}\|\mathcal{L}''_t(\psi,w)\|<\infty$.

Condition (v) is the uniqueness of ψ_0 as a maximum of $\ell''_\infty(\psi,w)$ and ensures the non-singularity of the limit weighted likelihood $\mathcal{L}''_\infty(\psi,w)=\mathbb{E}\mathcal{L}''_t(\psi,w)=\mathcal{I}(\psi,w)$. □

We can extend the asymptotic normality results to the mis-specified DFM by centering the WML estimator at the pseudo-true parameter $\psi_0^*(w)$.

Theorem A.5. (Asymptotic Normality) *Let the conditions of Theorem A.3 hold and $\psi_0^*(w)$ be a point in the interior of Ψ. Suppose further that $\{x_t\}$ and $\{y_t\}$ are SE and satisfy $\mathbb{E}|x_t|^4<\infty$ and $\mathbb{E}|y_t|^4<\infty$ and there exists a unique SE sequence $\{\tilde{df}_t\}$ such that*

$$\|\tilde{df}_t(\tilde{df}_1^*)-\tilde{df}_t\|_\Psi \xrightarrow{e.a.s.} 0\ \forall\ \tilde{df}_1^*\ \ as\ \ t\to\infty\quad with\quad \mathbb{E}|\tilde{df}_t|^4<\infty$$

and a unique SE sequence $\{\tilde{ddf}_t\}$ such that

$$\|\tilde{ddf}_t(\tilde{ddf}_1^*)-\tilde{ddf}_t\|_\Psi \xrightarrow{e.a.s.} 0\ \forall\ \tilde{ddf}_1^*\ \ as\ \ t\to\infty\quad with\quad \mathbb{E}|\tilde{ddf}_t|^2<\infty.$$

Then, for every $\tilde{f}_1^*$ and every $w\in(0,1]$, the ML estimator $\hat{\psi}_T(w,\tilde{f}_1^*)$ satisfies

$$\sqrt{T}\big(\hat{\psi}_T(\tilde{f}_1^*)-\psi_0^*(w)\big)\xrightarrow{d} N\Big(0\ ,\ \mathcal{I}^{-1}(\psi_0^*(w),w)\mathcal{J}(\psi_0^*(w),w)\mathcal{I}^{-1}(\psi_0^*(w),w)\Big)\ \ as\ \ T\to\infty.$$

Proof of Theorem A.5. Follows the same steps as the proof of Theorem A.4 ith the exception that the required properties of $\{x_t\}$ and $\{y_t\}$ are directly assumed rather than derived through Proposition A.1. □

A.2 Low frequency formulations for autoregressions

A.2.1 Notation

In this Appendix we discuss the stacking approach similar to the one used by Ghysels (2012) and Foroni et al. (2015). In the main paper we adopt the stacking method to treat mixed frequency dynamic factor model within a state space setting. But here we provide the details for the autoregressive model case for monthly time series (high frequency) that can be represented by quarterly (low frequency) vectors within a state space formulation. We use the notation x_τ^m for a variable x that is observed on a monthly (m) basis with monthly time index τ. The observations of the time series x_τ^m can be stacked into a quarterly (q) observed 3×1 vector x_t^q with quarterly time index t

$$x_t^q = \begin{pmatrix} x_{t,1}^q \\ x_{t,2}^q \\ x_{t,3}^q \end{pmatrix} = \begin{pmatrix} x_{3(t-1)+1}^m \\ x_{3(t-1)+2}^m \\ x_{3(t-1)+3}^m \end{pmatrix}, \tag{A.7}$$

where $x_{t,i}^q$ is the i-th element of x_t^q with index t indicating the number of the quarter of the observation and index i indicating the number of the month within the quarter; we have $t = 1, \ldots, T$, $i = 1, 2, 3$ and $\tau = 1, \ldots, 3n$, since each quarter consists of $s = 3$ months.

In a similar way, we can represent the monthly observations into yearly vectors. The monthly observed series x_τ^m can be stacked into a yearly (y) observed 12×1 vector x_t^y with yearly time index t. We then have

$$x_t^y = \begin{pmatrix} x_{t,1}^y \\ x_{t,2}^y \\ \vdots \\ x_{t,12}^y \end{pmatrix} = \begin{pmatrix} x_{12(t-1)+1}^m \\ x_{12(t-1)+2}^m \\ \vdots \\ x_{12(t-1)+12}^m \end{pmatrix}, \tag{A.8}$$

where $x_{t,i}^y$ is the i-th element of x_t^y with the first t indicating the number of the year of the observation and second index i indicating the number of the month within the year; we have $t = 1, \ldots, T$, $i = 1, \ldots, 12$ and $\tau = 1, \ldots, 12n$, since each year consists of $s = 12$ months. Throughout this paper, we will use the superscripts m, q and y to indicate the frequency of time series; we only use this notation where we deem it is necessary. When a variable or vector has only one index, this index typically refers to the number of the month, quarter or year of the observation. The second index refers to the number of the month within the quarter or year of the observation.

A.2.2 Linear State Space Models

The general linear Gaussian state space model can be written in a variety of ways. In this paper we adopt the notation used in Durbin and Koopman (2012), where the model is given as

$$\begin{aligned} x_t &= Z\alpha_t + \epsilon_t, & \epsilon_t &\sim NID\left(0, H\right), \\ \alpha_{t+1} &= T\alpha_t + R\eta_t, & \eta_t &\sim NID\left(0, Q\right), \end{aligned} \tag{A.9}$$

where x_t is a $k \times 1$ vector of observations called the *observation vector* and α_t is an unobserved $m \times 1$ vector called the *state vector*. The system matrices Z, T, R, H and Q are initially assumed to be known and the error terms ϵ_t and η_t are assumed to be serially independent and independent of each other at all time points. In practice, some or all of the matrices Z, T, R, H and Q will depend on elements of an unknown parameter vector ψ.

In the state space model, the state vector α_t cannot be observed directly and hence we base the analysis on observations x_t. These equations hold true for any frequency, as long as the state vector has the same frequency as the observation vector. Therefore, we do not use a superscript to indicate the frequency of the series, although we use these mostly for low frequency models.

The initial state vector α_1 is generated from $\mathrm{N}(a_1,\ P_1)$, independently of $\epsilon_1, \ldots, \epsilon_n$ and $\eta_1, \ldots, \eta_n$, where a_1 and P_1 are assumed known, although P_1 may depend on the parameter vector ψ.

A.2.3 Autoregressive Processes

We now adopt the stacking method in an autoregressive (AR) model setting. For brevity, we illustrate the method using an AR(1). Illustrations with AR(p) models or higher order p are provided in Appendix A. Detailed derivations for all models are provided in the Technical Appendix.

Monthly observations from the AR(1) process $x_\tau^m = \phi x_{\tau-1}^m + \varepsilon_\tau^m$ are stacked into the quarterly 3×1 vector x_t^q of (A.7). The quarterly process of the stacked variable x_t^q is then given by the vector autoregressive process

$$x_t^q = T x_{t-1}^q + R\varepsilon_t^q \tag{A.10}$$

with

$$T = \begin{pmatrix} 0 & 0 & \phi \\ 0 & 0 & \phi^2 \\ 0 & 0 & \phi^3 \end{pmatrix}, \qquad R = \begin{pmatrix} 1 & 0 & 0 \\ \phi & 1 & 0 \\ \phi^2 & \phi & 1 \end{pmatrix}, \tag{A.11}$$

such that the variance matrix of the vector x_t^q, conditional on x_{t-1}^q, is equal to $\sigma_\varepsilon^2 RR'$. We notice that all three elements of x_t^q depend only on the last element of x_{t-1}^q and on the associating elements of the vector of disturbances ε_t^q. The vector ε_t^q is the result of stacking the values of ε_τ^m in similar fashion as in (A.7). The autoregressive process (A.10) is equal to the linear Gaussian state space model (A.9) with state vector $\alpha_t = x_t^q$ and with system matrices T and R given by (A.11) and $Z = I_3$, $H = 0$, $Q = \sigma_\varepsilon^2$ and $\eta_t = \varepsilon_t^q$.

A.2.4 AR(1) process with stacked observations

Consider the AR(1) process of the monthly (m) observed variable x_τ^m with monthly time index τ

$$x_\tau^m = \phi x_{\tau-1}^m + \varepsilon_\tau, \qquad \varepsilon_\tau \sim NID\left(0, \sigma_\varepsilon^2\right) \tag{A.12}$$

When the monthly observations of x_τ^m are stacked into the quarterly (q) 3×1 vectors x_t^q with quarterly time index t, then the equations of the AR(1) process for the stacked observations can be written as

$$x_{t,1}^q = \phi x_{t-1,3}^q + \varepsilon_{t,1}, \qquad x_{t,2}^q = \phi x_{t,1}^q + \varepsilon_{t,2}, \qquad x_{t,3}^q = \phi x_{t,2}^q + \varepsilon_{t,3} \tag{A.13}$$

To develop a low frequency recursion for x_t^q, we substitute the first equation for the value of $x_{t,1}^q$ and the second equation for $x_{t,2}^q$ gives the following set of equations

$$\begin{array}{rcl} x_{t,1}^q & = & \phi x_{t-1,3}^q + \varepsilon_{t,1} \\ x_{t,2}^q & = & \phi(\phi x_{t-1,3}^q + \varepsilon_{t,1}) + \varepsilon_{t,2} \\ & = & \phi^2 x_{t-1,3}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2} \\ x_{t,3}^q & = & \phi(\phi^2 x_{t-1,3}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2}) + \varepsilon_{t,3} \\ & = & \phi^3 x_{t-1,3}^q + \phi^2\varepsilon_{t,1} + \phi\varepsilon_{t,2} + \varepsilon_{t,3} \end{array}, \tag{A.14}$$

which can be written as the autoregressive process

$$x_t^q = T x_{t-1}^q + R\varepsilon_t \tag{A.15}$$

with matrices

$$T = \begin{pmatrix} 0 & 0 & \phi \\ 0 & 0 & \phi^2 \\ 0 & 0 & \phi^3 \end{pmatrix}, \qquad R = \begin{pmatrix} 1 & 0 & 0 \\ \phi & 1 & 0 \\ \phi^2 & \phi & 1 \end{pmatrix} \tag{A.16}$$

where the variance matrix of the vector x_t^q, conditional on x_{t-1}^q, is equal to $\sigma_\varepsilon^2 RR'$. This autoregressive process is equal to the linear Gaussian state space model

$$\begin{array}{ll} x_t = Z\alpha_t + \varepsilon_t, & \varepsilon_t \sim N(0, H) \\ \alpha_{t+1} = T\alpha_t + R\varepsilon_t, & \varepsilon_t \sim N(0, Q) \end{array} \tag{A.17}$$

with $Z = I_3$, with $\alpha_t = x_t^q$ and $H = 0$.

The unconditional variance and covariances (used for initialization of the Kalman Filter) can be obtained from the Yule-Walker equations and are equal to

$$\begin{array}{l} \gamma_0 = \frac{\sigma_\varepsilon^2}{(1-\phi^2)} \\ \gamma_1 = \phi\gamma_0 \\ \gamma_2 = \phi\gamma_1 \end{array} \tag{A.18}$$

A.2.5 AR(2) process with stacked observations

For an AR(2) process with monthly observed variable x_τ^m the transformations are similar. Consider the model

$$x_\tau^m = \phi_1 x_{\tau-1}^m + \phi_2 x_{\tau-2}^m + \varepsilon_\tau, \qquad \varepsilon_\tau \sim NID\left(0, \sigma_\varepsilon^2\right) \tag{A.19}$$

The equations of the AR(2) process for the stacked quarterly observations become

$$\begin{array}{l} x_{t,1}^q = \phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \varepsilon_{t,1} \\ x_{t,2}^q = \phi_1 x_{t,1}^q + \phi_2 x_{t-1,3}^q + \varepsilon_{t,2} \\ x_{t,3}^q = \phi_1 x_{t,2}^q + \phi_2 x_{t,1}^q + \varepsilon_{t,3} \end{array} \tag{A.20}$$

Substitution of the first equation for the value of $x_{t,1}^q$ and the second equation for $x_{t,2}^q$ gives the following set of equations

$$\begin{array}{rcl} x_{t,1}^q & = & \phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \varepsilon_{t,1} \\ x_{t,2}^q & = & \phi_1(\phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,3}^q + \varepsilon_{t,1}) + \phi_2 x_{t-1,2}^q + \varepsilon_{t,2} \\ & = & (\phi_1^2 + \phi_2) x_{t-1,3}^q + \phi_1\phi_2 x_{t-1,2}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2} \\ x_{t,3}^q & = & \phi_1((\phi_1^2 + \phi_2) x_{t-1,3}^q + \phi_1\phi_2 x_{t-1,2}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2}) + \phi_2(\phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \varepsilon_{t,1}) + \varepsilon_{t,3} \\ & = & (\phi_1^3 + 2\phi_1\phi_2) x_{t-1,3}^q + (\phi_1^2\phi_2 + \phi_2^2) x_{t-1,2}^q + (\phi_1^2 + \phi_2)\varepsilon_{t,1} + \phi_1\varepsilon_{t,2} + \varepsilon_{t,3} \end{array} , \tag{A.21}$$

which can be written as the linear Gaussian state space model(A.17) with matrices

$$T = \begin{pmatrix} 0 & \phi_2 & \phi_1 \\ 0 & \phi_1\phi_2 & \phi_1^2 + \phi_2 \\ 0 & \phi_1^2\phi_2 + \phi_2^2 & \phi_1^3 + 2\phi_1\phi_2 \end{pmatrix}, \qquad R = \begin{pmatrix} 1 & 0 & 0 \\ \phi_1 & 1 & 0 \\ \phi_1^2 + \phi_2 & \phi_1 & 1 \end{pmatrix} \tag{A.22}$$

where the variance matrix of the vector x_t^q, conditional on x_{t-1}^q, is equal to $\sigma_\varepsilon^2 RR'$. The unconditional variance and covariances are equal to

$$\begin{array}{l} \gamma_0 = \frac{1-\phi_2}{1+\phi_2}\left(\frac{\sigma_\varepsilon^2}{(\phi_1+\phi_2-1)(\phi_2-\phi_1-1)}\right) \\ \gamma_1 = \frac{\phi_1}{(1-\phi_2)}\gamma_0 \\ \gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0 \end{array} \tag{A.23}$$

A.2.6 AR(3) process with stacked observations

For an AR(3) process with monthly observed variable x_τ^m the transformations are again similar. Consider the model

$$x_\tau^m = \phi_1 x_{\tau-1}^m + \phi_2 x_{\tau-2}^m + \phi_3 x_{\tau-3}^m + \varepsilon_\tau, \qquad \varepsilon_\tau \sim NID\left(0, \sigma_\varepsilon^2\right) \tag{A.24}$$

The equations of the AR(3) process for the stacked quarterly observations x_t^q become

$$\begin{array}{l} x_{t,1}^q = \phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \phi_3 x_{t-1,1}^q + \varepsilon_{t,1} \\ x_{t,2}^q = \phi_1 x_{t,1}^q + \phi_2 x_{t-1,3}^q + \phi_3 x_{t-1,2}^q + +\varepsilon_{t,2} \\ x_{t,3}^q = \phi_1 x_{t,2}^q + \phi_2 x_{t,1}^q + \phi_3 x_{t-1,3}^q + +\varepsilon_{t,3} \end{array} \tag{A.25}$$

Substitutions similar to those described in the previous subsections can again be applied. We have

$$\begin{array}{rcl}
x^q_{t,1} &=& \phi_1 x^q_{t-1,3} + \phi_2 x^q_{t-1,2} + \phi_3 x^q_{t-1,1} + \varepsilon_{t,1} \\
x^q_{t,2} &=& \phi_1 x^q_{t,1} + \phi_2 x^q_{t-1,3} + \phi_3 x^q_{t-1,2} + \varepsilon_{t,1} \\
&=& \phi_1(\phi_1 x^q_{t-1,3} + \phi_2 x^q_{t-1,2} + \phi_3 x^q_{t-1,1} + \varepsilon_{t,1}) + \phi_2 x^q_{t-1,3} + \phi_3 x^q_{t-1,2} + \varepsilon_{t,2} \\
&=& (\phi_1^2 + \phi_2) x^q_{t-1,3} + (\phi_1\phi_2 + \phi_3) x^q_{t-1,2} + \phi_1\phi_3 x^q_{t-1,1} + \phi\varepsilon_{t,1} + \varepsilon_{t,2} \\
x^q_{t,3} &=& \phi_1 x^q_{t,2} + \phi_2 x^q_{t,1} + \phi_3 x^q_{t-1,3} + \varepsilon_{t,1} \\
&=& \phi_1(\phi_1 x^q_{t-1,3} + \phi_2 x^q_{t-1,2} + \phi_3 x^q_{t-1,1} + \varepsilon_{t,1}) \\
&+& \phi_2((\phi_1^2 + \phi_2) x^q_{t-1,3} + (\phi_1\phi_2 + \phi_3) x^q_{t-1,2} + \phi_1\phi_3 x^q_{t-1,1} + \phi\varepsilon_{t,1} + \varepsilon_{t,2}) \\
&+& \phi_3 x^q_{t-1,3} + \varepsilon_{t,2} \\
&=& (\phi_1^3 + 2\phi_1\phi_2 + \phi_3) x^q_{t-1,3} + (\phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2) x^q_{t-1,2} \\
&+& (\phi_1^2\phi_3 + \phi_2\phi_3) x^q_{t-1,1} + \phi\varepsilon_{t,1} + \varepsilon_{t,2}
\end{array} , \quad \text{(A.26)}$$

which can be written as the linear Gaussian state space model (A.17) for the quarterly observed x^q_t with

$$T = \begin{pmatrix} \phi_3 & \phi_2 & \phi_1 \\ \phi_1\phi_3 & \phi_1\phi_2 + \phi_3 & \phi_1^2 + \phi_2 \\ \phi_1^2\phi_3 + \phi_2\phi_3 & \phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2 & \phi_1^3 + 2\phi_1\phi_2 + \phi_3 \end{pmatrix} \quad \text{(A.27)}$$

For the AR(3) process, R is the same matrix as for the AR(2) process and the variance matrix of the vector x^q_t, conditional on x^q_{t-1}, is again equal to $\sigma^2_\varepsilon RR'$. The three values of x^q_t now depend on all three observation of x^q_{t-1} and on the 3×1 vector of disturbances ε_t. The unconditional variance and covariances are equal to

$$\begin{array}{l}
\gamma_0 = \frac{\sigma^2_\varepsilon\left(1-\phi_2-\phi_1\phi_3-\phi_3^2\right)}{(1-\phi_2-\phi_3-\phi_1)\left(1+\phi_2+\phi_3\phi_1-\phi_3^2\right)(1+\phi_3+\phi_1-\phi_2)} \\
\gamma_1 = \frac{(\phi_1+\phi_2\phi_3)\gamma_0}{\left(1-\phi_2-\phi_3\phi_1-\phi_3^2\right)} \\
\gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0 + \phi_3\gamma_1
\end{array} \quad \text{(A.28)}$$

A.2.7 AR(p) process with stacked observations

For AR(p) processes of order $p > 3$ the state vector in (A.17) has to be extended with more lags of x^q_t. For example, for the AR(4) process we would have $\alpha_t = \left(x^q_{t-1,3}, x^q_{t,1}, x^q_{t,3}, x^q_{t,3}\right)'$ and the transition matrix becomes

$$T = \begin{pmatrix} 0 & 0 & 0 & 1 \\ \phi_4 & \phi_3 & \phi_2 & \phi_1 \\ \phi_1\phi_4 & \phi_1\phi_3 + \phi_4 & \phi_1\phi_2 + \phi_3 & \phi_1^2 + \phi_2 \\ \phi_1^2\phi_4 + \phi_2\phi_4 & \phi_1^2\phi_3 + \phi_2\phi_3 + \phi_1\phi_4 & \phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2 + \phi_4 & \phi_1^3 + 2\phi_1\phi_2 + \phi_3 \end{pmatrix},$$

where the variance matrix of the process is defined as $\sigma_\varepsilon^2 RR'$ with

$$R = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \phi_1 & 1 & 0 \\ 0 & \phi_1^2 + \phi_2 & \phi_1 & 1 \end{pmatrix}, \tag{A.29}$$

The unconditional variances and covariances can be obtained analytically using the Yule-Walker equations, or by numerically solving the Algebraic Riccati Equation.

A.3 Interpolation and Aggregation Approaches

Below we discuss the original solution by Mariano and Murasawa (2003) in treating mixed frequency data; they analyze all series at the highest frequency. In this approach, artificial missing values are introduced for the series that is observed at the lower frequency and interpolation techniques are used to describe the dynamics of the unobserved 'latent' monthly GDP growth. Furthermore, we explore a second approach where all series are modeled at the lowest frequency by aggregating the high frequency series to quarterly totals. In this approach, no artificial missing values are needed, but information about the high frequency series is lost and the model does not allow the econometrician to address high frequency dynamics.

A.3.1 Interpolation Approach

In the mixed frequency interpolation (MFI) approach, all series are treated at the highest frequency, say as monthly time series. All variables are driven by one unobserved monthly common factor f_τ and by the idiosyncratic factors u_τ and v_τ. The variable quarterly y_t is observed every third month and has missing values for the first two months of every quarter. Therefore, the data matrix has the following structure

$$\begin{bmatrix} \cdot & \cdot & y_3 & \cdot & \cdot & y_6 & \cdot & \dots & y_{3n} \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & \dots & x_{3n} \end{bmatrix}, \tag{A.30}$$

where x_τ is the $k \times 1$ vector $\left(x_\tau^{(1)}, \dots, x_\tau^{(k)}\right)'$ and n is the number of quarters in the sample period. Since in this approach all series are treated as monthly series, we choose to drop the superscript m, in order to avoid cumbersome notation. The contemporaneous and dynamic interactions between y_τ and the vector of monthly observed variables x_τ are specified via the model

$$\begin{pmatrix} \tilde{y}_\tau^m \\ x_\tau \end{pmatrix} = \begin{pmatrix} \beta_y g\left(f_\tau\right) \\ \beta_x f_\tau \end{pmatrix} + \begin{pmatrix} g(u_\tau) \\ v_\tau \end{pmatrix} + \begin{pmatrix} \xi_\tau \\ \varepsilon_\tau \end{pmatrix} \tag{A.31}$$

where $\tilde{y}_\tau^m$ is the latent monthly variable for y and for which we only have observations available in the last month of each quarter, β_y is a scalar coefficient, β_x is a $k \times 1$ vector of coefficients, and

$$g(a_\tau) = \frac{1}{3}a_\tau + \frac{2}{3}a_{\tau-1} + a_{\tau-2} + \frac{2}{3}a_{\tau-3} + \frac{1}{3}a_{\tau-4} \tag{A.32}$$

for $\tau = 1, \ldots, 3n$. The vector u_τ is the stationary sequence of the idiosyncratic factor for y_τ and v_τ is the stationary sequence of the vector of idiosyncratic factors for x_τ and consists of one value per month for each monthly observed variable $x_\tau^{(i)}$.

The factors f_τ, u_τ and v_τ are modeled as AR processes

$$f_\tau \sim \mathrm{AR}(p_f), \qquad u_\tau \sim \mathrm{AR}(p_u), \qquad v_\tau \sim \mathrm{AR}(p_v), \tag{A.33}$$

where $\mathrm{AR}(p)$ refers to the autoregressive process of order p. There are no interactions between the series of f_τ, u_τ and v_τ nor between the series of $v_\tau^{(i)}$ and $v_\tau^{(j)}$ for any $i \neq j$.

We have established the model by Mariano and Murasawa (2003) and implicitly their model-based solution to the mixed frequency problem. They advocate to use the Kalman filter for likelihood evaluation and general analysis. Specifically, they take advantage of the fact that the Kalman filter can treat missing observations without a problem.

A.3.2 Aggregation Approach

An alternative approach, where the introduction of artificial missing values is not required, is the aggregation of the monthly series x_τ^m into quarterly totals $\bar{x}_t^q$ and the treatment of all series as quarterly series. This model then only describes quarterly dynamics. The unobserved common factor must also become quarterly. To avoid cumbersome notation we drop the superscript q and we write the model in the form

$$\begin{pmatrix} y_t \\ \bar{x}_t \end{pmatrix} = \begin{bmatrix} \beta_y & 1 & 0 \\ \beta_x & 0 & I_k \end{bmatrix} \begin{pmatrix} f_t \\ u_t \\ \bar{v}_t \end{pmatrix} + \begin{pmatrix} \xi_t \\ \bar{\varepsilon}_t \end{pmatrix}, \tag{A.34}$$

for $t = 1, \ldots, T$, where $\bar{x}_t$ and β_x are both $k \times 1$ vectors. The common factor f_t and the idiosyncratic factors u_t and $\bar{v}_t$ can still be modeled by AR processes as in (A.33). However, we must take care when interpreting the values of the parameters of these processes, as they now describe the dynamics from quarter to quarter. We will hereafter refer to this approach as the mixed frequency aggregation (MFA) approach.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

637 T.E. YENILMEZ, Three Essays in International Trade and Development

638 L.P. DE BRUIJN, Essays on Forecasting and Latent Values

639 S. VRIEND, Profiling, Auditing and Public Policy: Applications in Labor and Health Economics

640 M.L. ERGUN, Fat Tails in Financial Markets

641 T. HOMAR, Intervention in Systemic Banking Crises

642 R. LIT, Time Varying Parameter Models for Discrete Valued Time Series

643 R.H. KLEIJN, Essays on Bayesian Model Averaging using Economic Time Series

644 S. MUNS, Essays on Systemic Risk

645 B.M. SADABA, Essays on the Empirics of International Financial Markets

646 H. KOC, Essays on Preventive Care and Health Behaviors

647 V.V.M. MISHEVA, The Long Run Effects of a Bad Start

648 W. LI, Essays on Empirical Monetary Policy

649 J.P. HUANG, Topics on Social and Economic Networks

650 K.A. RYSZKA, Resource Extraction and the Green Paradox: Accounting for Political Economy Issues and Climate Policies in a Heterogeneous World

651 J.R. ZWEERINK, Retirement Decisions, Job Loss and Mortality

652 M. K. KAGAN, Issues in Climate Change Economics: Uncertainty, Renewable Energy Innovation and Fossil Fuel Scarcity

653 T.V. WANG, The Rich Domain of Decision Making Explored: The Non-Triviality of the Choosing Process

654 D.A.R. BONAM, The Curse of Sovereign Debt and Implications for Fiscal Policy

655 Z. SHARIF, Essays on Strategic Communication

656 B. RAVESTEIJN, Measuring the Impact of Public Policies on Socioeconomic Disparities in Health

657 M. KOUDSTAAL, Common Wisdom versus Facts; How Entrepreneurs Differ in Their Behavioral Traits from Others

658 N. PETER, Essays in Empirical Microeconomics

659 Z. WANG, People on the Move: Barriers of Culture, Networks, and Language

660 Z. HUANG, Decision Making under Uncertainty-An Investigation from Economic and Psychological Perspective

661 J. CIZEL, Essays in Credit Risk, Banking, and Financial Regulation

662 I. MIKOLAJUN, Empirical Essays in International Economics

663 J. BAKENS, Economic Impacts of Immigrants and Ethnic Diversity on Cities

664 I. BARRA, Bayesian Analysis of Latent Variable Models in Finance

665 S. OZTURK, Price Discovery and Liquidity in the High Frequency World
666 J. JI, Three Essays in Empirical Finance
667 H. SCHMITTDIEL, Paid to Quit, Cheat, and Confess
668 A. DIMITROPOULOS, Low Emission Vehicles: Consumer Demand and Fiscal Policy
669 G.H. VAN HEUVELEN, Export Prices, Trade Dynamics and Economic Development
670 A. RUSECKAITE, New Flexible Models and Design Construction Algorithms for Mixtures and Binary Dependent Variables
671 Y. LIU, Time-varying Correlation and Common Structures in Volatility
672 S. HE, Cooperation, Coordination and Competition: Theory and Experiment
673 C.G.F. VAN DER KWAAK, The Macroeconomics of Banking
674 D.H.J. CHEN, Essays on Collective Funded Pension Schemes
675 F.J.T. SNIEKERS, On the Functioning of Markets with Frictions
676 F. GOMEZ MARTINEZ, Essays in Experimental Industrial Organization: How Information and Communication affect Market Outcomes
677 J.A. ATTEY, Causes and Macroeconomic Consequences of Time Variations in Wage Indexation
678 T. BOOT, Macroeconomic Forecasting under Regime Switching, Structural Breaks and High-dimensional Data
679 I. TIKOUDIS, Urban Second-best Road Pricing: Spatial General Equilibrium Perspectives
680 F.A. FELSŐ, Empirical Studies of Consumer and Government Purchase Decisions
681 Y. GAO, Stability and Adaptivity: Preferences over Time and under Risk
682 M.J. ZAMOJSKI, Panta Rhei, Measurement and Discovery of Change in Financial Markets
683 P.R. DENDERSKI, Essays on Information and Heterogeneity in Macroeconomics
684 U. TURMUNKH, Ambiguity in Social Dilemmas
685 U. KESKIN, Essays on Decision Making: Intertemporal Choice and Uncertainty
686 M. LAMMERS, Financial Incentives and Job Choice