Essays in Empirical Macroeconomics

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von
HARUN MIRZA
aus Bonn

Bonn 2013

Dekan:	Prof. Dr. Klaus Sandmann
Erstreferent:	Prof. Dr. Jürgen von Hagen
Zweitreferent:	Prof. Dr. Jörg Breitung

Tag der mündlichen Prüfung: 11.12.2013

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn (http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

Acknowledgments

Writing this thesis would have been an unsuccessful endeavour without the help of numerous Professors, colleagues and friends. In the following I would like to express my gratitude towards certain (groups of) people in particular.

I would first like to thank my supervisors Prof. Dr. Jürgen von Hagen and Prof. Dr. Jörg Breitung for continuous support and important suggestions and ideas. Thank you for your guidance and support regarding my various undertakings. Further, I am indebted to Gernot Müller for his help and insightful discussions. I would also like to thank my "mentor" Michael Evers.

I want to thank my coauthors Lidia, Marta and Paul. You are great people to work with and I owe you for fruitful discussions, excellent cooperation and continuous efforts to improve our work. I also thank Luca Onorante who worked with Marta and me on a project related to our study. Further, I thank Thomas Westermann, who was the responsible line manager during my stay at the ECB, as well as the whole Euro Area Macroeconomic Developments Division for its hospitality. I also benefited from helpful discussions with Barbara Rossi during my stay at the University Pompeu Fabra in Barcelona.

I also wish to thank the BGSE, especially Prof. Dr. Urs Schweizer, Dr. Silke Kinzig and Pamela Mertens for their efforts in making the PhD programme such a success. I appreciate Anita Suttarp's administrative support and her efforts in managing any activities related to the Institute for International Economic Policy. I would also like to thank the German Research Foundation (DFG) for financial support. I am grateful for interesting ideas and suggestions of my colleagues from the BGSE and participants at various seminars. Especially, the BGSE 2009 class deserves thanks for making the time in Bonn such a great experience both in terms of research and in terms of various activities unrelated to the university.

Finally, a few persons deserve special attention. First, Thomas, thank you for being the best office mate one can imagine. You were a great source of knowledge, a superb listener and above all an excellent friend. I would like to thank my family for supporting me whatever I am doing. And last I wish to thank you, Nici, for always being there for me and having endured all frustrations and successes that have been associated with me writing this dissertation.

Contents

List of Figures

List of Tables

Introduction

This thesis consists of three chapters in empirical macroeconomics. In particular it treats important issues related to monetary policy. The focus is on inflation dynamics and all three studies involve some approach of forecasting.

The first chapter deals with estimation problems of a monetary policy rule and inflation dynamics. In light of potential weak instrument problems, the study suggests to use factors as a summary of the relevant information typically available at central banks. These factors are used to instrument the public's unobserved inflation expectations prevalent at each point in time. The second chapter more explicitly focuses on forecasting inflation and in contrast to the first chapter it discusses out-of-sample forecasting with a large set of Phillips curve models for Euro area inflation. A particular emphasis is on forecast combination techniques and various approaches in detrending inflation prior to estimation. While also touched upon in the first two chapters, the final chapter explicitly deals with survey inflation expectations, i.e., it analyses how professional forecasters form their expectations as reported in a well-known survey. We examine how much weight these forecasters attach to backward- and how much to forward-looking information and suggest a Phillips curve model to explain their expectations formation process.

CHAPTER 1.[1] It is common to use instrumental variable approaches in estimating the parameters of Phillips curve and Taylor rule models with forward-looking expectations. Weak identification or weak instrument problems often arrise in this context. They describe a situation where the parameters of a model are not well-

[1]This chapter is based on joint work with Lidia Storjohann. Our paper is forthcoming in the Journal of Money, Credit and Banking (see Mirza and Storjohann, 2013)

identified as the instruments do not contain sufficient information for the endogenous variable they are supposed to predict. In such a case instrumental variable estimation (or GMM) can be biased. Since weak-identification robust inference methods often yield inconclusively large confidence sets, we suggest to use factors generated from a large macroeconomic data set – as is typically available at central banks – as additional instruments. We show that the combination of robust methods and the use of factors improves the estimation results substantially in a way that allows us to show first that the Taylor principle is satisfied over the Volcker-Greenspan regime in the US. Second, we find that forward-looking dynamics dominate backward-looking dynamics in the estimation of a hybrid New Keynesian Phillips Curve (NKPC).

CHAPTER 2.[2] The second chapter more explicitly deals with forecasting, while in contrast to chapter one, we consider out-of-sample forecasts. We analyse a large variety of Phillips curve models and evaluate their performance in forecasting euro area inflation against popular benchmarks. In light of considerable model heterogeneity and instability we resort to two approaches proposed in the literature: Model averaging and detrending prior to estimation. We show that the performance of Phillips curve models is episodic and univariate models seem to be tough benchmarks. Models where inflation is detrended by long-run survey expectation seem to lead to the lowest forecast errors, while model-based alternatives for the trend perform comparably only in some instances. Model combination is a promising strategy in light of a huge set of candidate models and we show that the simple average over forecasts from this set is seldomly improved upon by more sophisticated performance-based averages.

CHAPTER 3.[3] This chapter analyses how survey inflation expectations are formed. In particular, we suggest to use a hybrid New Keynesian Phillips Curve model to evaluate the private inflation expectations formation process. We show that professional forecasters attach a larger weight to forward-looking information, while backward-looking information is also relevant. The contribution of the marginal cost measure is small and often insignificantly different from zero. As the forecasting hori-

[2]This chapter is based on joint work with Marta Banbura. Special thanks go to Luca Onorante, who has worked with us on a related project.

[3]This chapter is based on joint work with Paul Hubert. The research project benefited from funding of the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°320278.

zon varies or one considers further-ahead forward-looking information, the weights remain relatively constant. The findings suggest that the central bank can help to anchor short-run inflation expectations through anchoring longer-term inflation expectations. The parameter estimates of our NKPC-based inflation expectations formation model are comparable to what the literature finds on the actual NKPC, suggesting that professional forecasters indeed employ this model in forming their expectations. In contrast, consumers do not seem to attach particular weight to forward-looking information and their inflation expectations formation process cannot be explained well by means of the same Phillips curve model.

Chapter 1

Making Weak Instrument Sets Stronger: Factor-Based Estimation of Inflation Dynamics and a Monetary Policy Rule

1.1 Introduction

This paper combines the insights from the literature on factor models and from studies on the weak-identification problem in the estimation of single-equation time-series models. We show that adding factors, generated from a large macroeconomic data set, as additional instruments in Generalized Method of Moments (GMM) estimation yields more precise results for a forward-looking Taylor rule and the hybrid New Keynesian Phillips Curve (NKPC).

In a recent paper, Mavroeidis (2010) reassesses the seminal work by Clarida, Galí, and Gertler (2000). Given that their analysis of monetary policy rules in the US might suffer from weak instrumental variables (IV),[1] which can lead to biased estimators and inference, he evaluates their model using methods that are robust against weak IVs. In constructing joint confidence sets for the parameters on expected future inflation and the output gap, he empirically confirms the conclusion that pre-Volcker monetary policy was accommodative to inflation. In contrast to Clarida, Galí, and Gertler (2000) though, he claims that with the use of robust

[1]Note that for ease of reference we denote the case of weak identification also as a problem of weak instruments.

methods it cannot be shown whether monetary policy during the Volcker-Greenspan tenure was adherent to the Taylor principle or not due to inconclusive confidence sets. Similarly, Kleibergen and Mavroeidis (2009) estimate the hybrid NKPC, as introduced by Galí and Gertler (1999), using weak-identification robust methods. They find confidence sets that are so large as to be consistent with both dominant forward- and backward-looking inflation dynamics.

We follow a different route in this paper. Rather than relying solely on typical instruments such as own lags of variables in the model that can result in uninformatively large robust confidence sets, we construct additional instruments by estimating factors from a comprehensive macroeconomic data set (Stock and Watson, 2008). We employ these factors in the first stage of the respective estimation, an approach applied to point estimates of the NKPC by Beyer, Farmer, Henry, and Marcellino (2008) and Kapetanios and Marcellino (2010) and to Taylor rules by Bernanke and Boivin (2003) and Favero, Marcellino, and Neglia (2005). In contrast to these studies, we consider confidence sets of the parameters in order to derive conclusions with respect to the Taylor principle and the joint behavior of the parameters of the NKPC. In addition, we rely on the weak-identification robust statistic suggested by Kleibergen (2005) given that it is not known a priori whether factors will be strong instruments.

The literature on factor analysis has shown that dimension-reduction techniques can be successful in summarizing a vast amount of information in few variables (e.g. Stock and Watson, 2002, 2008). These variables, i.e. the factors, can perform well as additional instruments in IV and GMM estimation as has been shown in formal evaluations by Bai and Ng (2010) and Kapetanios and Marcellino (2010), respectively. Kapetanios, Khalaf, and Marcellino (2011) analyze factor-based weak IV robust statistics for linear IV estimation.

Our empirical results illustrate that the use of factors substantially reduces the size of the two-dimensional weak IV robust confidence sets, as the factor-augmented instrument set is stronger in the estimation procedure. First, this leads to evidence of dominant forward-looking dynamics in the NKPC, while the coefficient on the marginal cost measure is not significantly different from zero. Second, the results with respect to the Taylor rule allow us to conclude that in the Volker-Greenspan period, monetary policy satisfied the Taylor principle. For this period, we also evaluate the usefulness of survey-based expectations as instruments and find that they can somewhat improve precision of the Taylor rule estimates if added to the factor-based instrument set or to the variable set of the factor model.

The structure of the paper is as follows. In Section 1.2, we introduce the hybrid NKPC, as well as the assumed Taylor rule and the corresponding transmission mechanism. Section 1.3 presents our approach and Section 1.4 corresponding results. Section 1.5 concludes.

1.2 Model

1.2.1 The Hybrid New Keynesian Phillips Curve

We analyze the hybrid version of the NKPC as used by Galí and Gertler (1999) and Kleibergen and Mavroeidis (2009), among others. This version of inflation dynamics includes both forward- and backward-looking elements:

$$\pi_t = \delta \, mc_t + \gamma_f \mathbb{E}_t \pi_{t+1} + \gamma_b \pi_{t-1} + u_t, \tag{1.1}$$

where π_t and mc_t are the inflation rate and a measure of marginal costs, respectively, and $\mathbb{E}_t$ is the expectation operator with respect to information up to time t. The parameter δ is the slope and γ_f and γ_b can be interpreted as the respective weights on forward- versus backward-looking dynamics in the economy. The variable u_t is an unobserved cost-push shock with $\mathbb{E}_{t-1}\, u_t = 0$. The estimation equation is obtained by replacing expected future inflation by its realization:

$$\pi_t = \delta \, mc_t + \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + e_t^{(1)}, \tag{1.2}$$

where the resulting error $e_t^{(1)} = u_t - \gamma_f(\pi_{t+1} - \mathbb{E}_t \pi_{t+1})$ may be autocorrelated at lag 1.

1.2.2 A Model of Monetary Policy

A Forward-Looking Taylor Rule

The conduct of monetary policy we assume is the Clarida, Galí, and Gertler (2000) version of a forward-looking Taylor rule with a certain degree of interest rate smoothing, which is also used in Mavroeidis (2010):

$$r_t = \alpha + \rho(L)\, r_{t-1} + (1-\rho)(\psi_\pi\, \mathbb{E}_t \pi_{t+1} + \psi_x\, \mathbb{E}_t x_t) + \varepsilon_t, \tag{1.3}$$

where the variables r_t, π_{t+1} and x_t are the policy interest rate, the one-period-ahead inflation rate and the output gap, respectively.[2] The monetary policy shock is an i.i.d. innovation such that $\mathbb{E}_{t-1}\, \varepsilon_t = 0$. The intercept α is a linear combination of the inflation and the resulting interest rate target and (ψ_π, ψ_x) are the feedback coefficients of the policy rule. $\rho(L) = \rho_1 + \rho_2 L + \ldots + \rho_n L^{n-1}$ displays the degree of policy smoothing, where L is the lag operator, and $\rho = \rho_1 + \rho_2 + \ldots + \rho_n$.

The estimation equation is once more obtained by replacing the expected values by their realizations:

$$r_t = \alpha + \rho(L)\, r_{t-1} + (1-\rho)(\psi_\pi \pi_{t+1} + \psi_x x_t) + e_t^{(2)}, \tag{1.4}$$

where the resulting error $e_t^{(2)} = \varepsilon_t - (1-\rho)[\psi_\pi(\pi_{t+1} - \mathbb{E}_t \pi_{t+1}) + \psi_x(x_t - \mathbb{E}_t x_t)]$ may exhibit first-order autocorrelation.

Transmission Mechanism

The transmission mechanism used to interpret the results is fully characterized by two equilibrium conditions which are derived from a standard New Keynesian sticky-price model by log-linearization around the steady state (see e.g. Clarida, Galí, and Gertler, 2000; Lubik and Schorfheide, 2004). Together with equation (1.3) these two conditions, namely an Euler equation for output, $y_t = \mathbb{E}_t y_{t+1} - \sigma(r_t - \mathbb{E}_t \pi_{t+1}) + g_t$, and a version of the New Keynesian Phillips Curve, $\pi_t = \beta\, \mathbb{E}_t \pi_{t+1} + \lambda(y_t - z_t)$, capture the dynamics of the model. The output elasticity of inflation $\lambda > 0$ reflects the degree of nominal rigidities, $0 < \beta < 1$ is the discount factor, y_t stands for output and $z_t = y_t - x_t$ captures variation in the marginal cost of production. In the Euler equation σ is the intertemporal elasticity of substitution and g_t represents exogenous shifts in preferences and government spending.

[2] As the output gap x_t is not known at the time the interest rate is set in period t, we use its expected value.

As highlighted in Woodford (2003, ch. 4), determinacy in this model requires:

$$\psi_\pi + \frac{1-\beta}{\lambda}\,\psi_x - 1 \geq 0. \tag{1.5}$$

Further, the interest rate response should not be too strong – a condition that is not binding for the empirical results in this paper.[3]

Equation (1.5) is a generalized version of Taylor's principle that the policy rate should be raised more than one for one with inflation to guarantee macroeconomic stability and can be seen as a benchmark to evaluate monetary policy (see Taylor (1999) for a qualitative and Clarida, Galí, and Gertler (2000) for a more quantitative perspective on this principle).

1.3 Factor-GMM Methodology

1.3.1 Benchmark Specifications

As the realizations of future inflation and the output gap are unknown at time t, we estimate both models with GMM assuming rational expectations, where the moment conditions are $\mathbb{E}Z_t^{(i)}e_t^{(i)} = 0$ for any predetermined instrument set $Z_t^{(i)}$ and $i = 1, 2$. For both models we use an estimation sample consisting of quarterly data from 1961:I to 2006:I (see the data appendix for details). This corresponds exactly to the specifications in Mavroeidis (2010) and is similar to that in Kleibergen and Mavroeidis (2009).[4]

New Keynesian Phillips Curve

In accordance with the paper by Kleibergen and Mavroeidis (2009) we estimate the NKPC with the labor share as a proxy for marginal costs and a benchmark instrument set that comprises three lags of inflation and the labor share.[5]

[3] Recent studies show that other factors might also be important in guaranteeing determinacy (see e.g. Davig and Leeper, 2007; Coibion and Gorodnichenko, 2011). Cochrane (2011) argues that the existence of a unique equilibrium in a New Keynesian model with a Taylor rule requires imposing strong assumptions. Further, he shows analytically that the forward-looking version we analyze in this paper can be identified.

[4] The data set in the latter study goes until 2007:4 which, however, would not be possible in our context given limited data availability for the factor model.

[5] In order to guarantee comparability with the study by Kleibergen and Mavroeidis (2009) we

Point estimates by Galí and Gertler (1999) indicate a dominance of forward- over backward-looking dynamics and further that the coefficient on the labor share is positive and significantly different from zero.[6] Recent criticism of such an approach emphasizes that the parameters of the NKPC could be weakly identified, and thus researchers should rely on weak-instrument robust inference (see e.g. Ma, 2002; Mavroeidis, 2004, 2005). It has been shown that conventional GMM methods can be biased in the single-equation context, when the expected Jacobian of the moment equation is not of full rank as the instruments are insufficiently correlated with the relevant first-order conditions (see Stock and Wright, 2000; Mavroeidis, 2004, among others).

Hence, Kleibergen and Mavroeidis (2009) base their interpretations on one- and two-dimensional confidence sets that are found by inverting weak-identification robust statistics such as Stock and Wright's S or Moreira's MQLR, which are applications to GMM of the Anderson-Rubin and Morereira's CLR statistic, respectively, as well as the K-LM and the JKLM statistic from Kleibergen (2005).[7] In our analysis we rely on the combined K-LM test discussed in Kleibergen (2005) and also used in Mavroeidis (2010) that is a combination of a 9 percent level K-LM test and a 1 percent level JKLM test, which improves the power of the former test against irrelevant alternatives.[8] Further, Newey and Windmeijer (2009) show that this version of the K-LM test and the test based on the MQLR statistic are asymptotically valid even under many weak moment conditions. These results, however, do not apply to the finite sample case if many moments are arbitrarily weak (e.g. if the instruments are irrelevant).

Kleibergen and Mavroeidis (2009) find confidence intervals that are so wide as to accommodate both dominant backward- and dominant forward-looking dynamics, i.e. values of γ_f both larger and smaller than 0.5, respectively. Further, they provide evidence that the coefficient on the labor share is statistically indistinguishable from

treat the labor share as endogenous.

[6]Note that the estimation sample in the study by Galí and Gertler (1999) only goes until 1997:IV and that their instrument set also contains lags of the long-short interest rate spread, output gap, wage inflation and commodity price inflation.

[7]For a discussion of the behavior of these statistics see the latter paper.

[8]To have more reliable results, we actually use a combination of a 4.5 percent level K-LM test and a 0.5 percent level JKLM test. Henceforth, whenever we mention the K-LM test we refer to this combined version.

zero.

Taylor Rule
For the Taylor rule the benchmark instrument set consists of four lags of each the Federal Funds rate, inflation and the output gap. The estimation sample is split such that the pre-Volcker and Volcker-Greenspan periods run from 1961:I to 1979:II and 1979:III to 1997:IV, respectively. We also briefly consider a third period from 1987:III to 2006:I which corresponds to the mandate of Alan Greenspan. Mavroeidis (2010) uses the same instrument set and time periods and in order to guarantee comparability of our results, we stick with the additional assumption that $n = 2$ for the first and $n = 1$ for the following time periods, i.e. $\rho(L) = \rho_1 + \rho_2 L$ and $\rho(L) = \rho_1$, respectively.[9]

Clarida, Galí, and Gertler (2000) find evidence that in the pre-Volcker period monetary policy was accommodative to inflation and therefore might have allowed for sunspot fluctuations in inflation, while in the second era it satisfied the Taylor principle, as depicted by inequality (1.5).

It has been pointed out, however, that estimation of DSGE models may be subject to the weak-identification problem (see e.g. Lubik and Schorfheide, 2004; Canova and Sala, 2009). Therefore, Mavroeidis (2010) reconsiders the empirical evidence of Clarida, Galí, and Gertler (2000) by testing different joint parameter specifications for the feedback coefficients of the Taylor rule using the K-LM test that is weak-instrument robust and for a high degree of overidentification more powerful than a test based on Stock and Wright's S statistic (see Kleibergen, 2005).

For the pre-Volcker period Mavroeidis' results support the previous finding that monetary policy did not satisfy the Taylor principle. For the second subsample, on the other hand, he shows that there is inconclusive evidence whether a determinate equilibrium exists or not due to uninformative confidence sets.

[9] Clarida, Galí, and Gertler (2000) use four lags of commodity price inflation, M2 growth and the spread between the long-term bond rate and the three-month Treasury bill rate as additional instruments and consider slightly different time periods, where the first period spans 1960:I to 1979:II and the second 1979:III to 1996:IV.

1.3.2 A Factor Model

The size of the weak IV robust confidence sets by Kleibergen and Mavroeidis (2009) and Mavroeidis (2010) suggests that in both models instruments are indeed weak and therefore stronger instruments are called for. Thus, we follow the approach of generating factors from a large macroeconomic data set and using them in the first stage of the estimation as discussed for the NKPC by Beyer et al. (2008) and Kapetanios and Marcellino (2010) and for Taylor rules in Bernanke and Boivin (2003) and Favero, Marcellino, and Neglia (2005). In contrast to these authors, who consider only point estimates, we also analyze joint confidence sets of the parameter estimates. This enables us to make inference with respect to the Taylor principle. Further, we provide a discussion on the comparison of forward- and backward-looking dynamics in the NKPC jointly with an analysis of the coefficient on the labor share. The rationale underlying the use of Factor GMM is that a central banker relies on a large information set in his forecasts of important macroeconomic variables. While each individual variable in this data set is only weakly correlated with future inflation, the output gap or the labor share and therefore contains only little information, the factors serve as a summary of that information and are thus better predictors for our variables of interest (Bernanke and Boivin, 2003).

The results by Stock and Watson (2002, 2008) indicate that the factors derived from their data sets contain important information with respect to inflation and output. Consequently, they have the potential to make the benchmark instrument set stronger. In order for the factors to be appropriate instruments, we need to make sure that they are uncorrelated with the error terms in equation (1.2) and (1.4). Therefore, the validity of the overidentifying restrictions is discussed in Section 1.4.

The properties of Factor-IV and Factor-GMM estimation are analyzed with Monte-Carlo simulations by Bai and Ng (2010) and Kapetanios and Marcellino (2010), respectively. Kapetanios, Khalaf, and Marcellino (2011) evaluate factor-based weak IV robust statistics. Favero, Marcellino, and Neglia (2005) compare two different ways to construct factors in a dynamic factor model: dynamic and static principal components (for the two approaches see Forni, Hallin, Lippi and Reichlin, 2000 and Stock and Watson, 2002, respectively). The authors report that the results for the two methods are comparable. Overall the static factors perform slightly better in their applications, while the dynamic factors seem to provide a better summary of information as fewer factors explain as much variation in the variables from the data set. For simplicity we rely on static principle components, given that the performance of both methods seems comparable.

Principal component analysis relies on the assumption that the set of variables is driven by a small set of factors and some idiosyncratic shocks. We assume the data-generating process underlying the variables to admit a factor representation:

$$X_t = \Lambda F_t + \nu_t, \tag{1.6}$$

where X_t is an $N \times 1$ vector of zero-mean, $I(0)$ variables, Λ is an $N \times k$ matrix of factor loadings, F_t is an $k \times 1$ vector of the factors and ν_t is an $N \times 1$ vector of idiosyncratic shocks, where N, the number of variables, is much larger than the number of factors k. Static factors can be estimated by minimizing the following objective function:

$$V_{N,T}(F, \Lambda) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \Lambda_i' F_t)^2, \tag{1.7}$$

where $F = (F_1, F_2, \dots, F_T)'$, Λ_i' is the i-th row of Λ, X_{it} is the i-th component of X_t and T is the number of time periods.

1.3.3 Data Set

To construct the factors we employ the data set by Stock and Watson (2008), which is an updated version of the data they use for former papers, e.g. Stock and Watson (2002). The subset of this data set relevant for the estimation of factors includes 109 quarterly time series that have strong information content with respect to inflation and output, consisting of disaggregated price and production data, as well as indices, among others. The time series span 1959:III to 2006:IV with $T = 190$ observations. We use principal component analysis to extract the factors from the transformed data series, where we carried out the same transformations as indicated in Stock and Watson (2008) to guarantee stationarity of both the time series and the resulting factors (see the data appendix for details).

Stock and Watson (2008) use the factors for forecasting and provide evidence that if potential changes in the factor model are sufficiently small there is a particular benefit in calculating the factors for the whole data set by principal components, even if there exists a structural break in the forecasting equation.[10] Moreover, in the construction of the factors having more observations increases the signal-to-noise ratio.

So far there is no general consensus on how to determine the number of factors k. We rely on the criteria that are recommended by Bai and Ng (2002) in this context (PC_1, PC_2, IC_1, IC_2) and are frequently used in the literature on factor models as they seem to perform well for large N. The PC criteria, which are shown to rather overestimate the true number of factors, are consistent with five or six factors, whereas the IC criteria are consistent with two or four factors for the whole data set. Based on these results and the canonical correlations between subsample and full-sample estimates of the factors, Stock and Watson (2008) make a case for using four factors, and we follow their suggestion. Using more factors does not improve our estimation results significantly, while it introduces even more instruments, and with fewer factors the results are somewhat less accurate; in either case the main conclusions would persist.[11]

1.4 Results

1.4.1 New Keynesian Phillips Curve

We estimate equation (1.2) as described in Section 1.3.1 and employ the same data set as Kleibergen and Mavroeidis (2009) for the benchmark results. However, in

[10]If one interprets the factor model as a set of policy functions, where the factors can be seen as states, a structural break in the Taylor rule has the potential to cause a break in the factor model. However, as Stock and Watson (2008) show, the factor model is relatively stable such that any potential regime change in monetary policy conduct would have only affected the dynamics of the benchmark instruments while the factor model implied policy functions are relatively unchanged.

[11]More recently proposed criteria like those by Onatski (2009) or Ahn and Horenstein (2009) are in line with our choice. The criterion by Onatski as well as the two criteria by Ahn and Horenstein predict two factors. Simulations by the respective authors have shown that these criteria tend to rather underestimate the true number of factors. As underestimation of the number of factors is more severe than overestimation in this context, the use of four factors seems a reasonable choice.

Table 1.1: Point estimates for the parameters of the NKPC

	Time period (in quarters) 1961:I-2006:I	
	BM	Factor GMM
δ	0.02	0.03*
	(0.03)	(0.02)
γ_f	0.73***	0.65***
	(0.08)	(0.02)
γ_b	0.27***	0.34***
	(0.08)	(0.02)

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Standard errors are in brackets. Estimation of the NKPC, equation (1.2), is conducted by GMM using Newey-West weight matrix. BM refers to the results based on the benchmark instrument set comprising three lags of each inflation and the labor share. The Factor-GMM results are generated extending the instrument set by lags one to four of the factors derived before.

order to have more information with respect to the two endogenous variables and thus more precise estimation results, we expand the benchmark instrument set by the four factors we generated from the Stock and Watson (2008) data set. As the contemporaneous values of the factors may be correlated with the error term $e_t^{(1)}$, we use only their first four lags as instruments. To investigate whether the overidentifying restrictions are satisfied, we calculate the weak-identification robust S sets for both instrument sets considered. These confidence sets are based on the S statistic that equals the value of the GMM objective function at the parameter values of the null hypothesis. They contain all parameter values, where one cannot jointly reject the null hypothesis and the validity of the overidentifying restrictions. The fact that the S sets are indeed not empty provides evidence that our identifying assumptions are reasonable (see Stock and Wright, 2000).

Point estimates are presented in Table 1.1. As discussed in Kleibergen and Mavroeidis (2009), results based on the benchmark instrument set indicate a dominance of forward- over backward-looking dynamics with parameter values of $(\gamma_f, \gamma_b) = (0.73, 0.27)$ both being significant at the 1 percent level. This is in line with the findings by Galí and Gertler (1999). The coefficient on the labor share is positive and – unlike in the latter study – insignificant. Including the factors in the instrument set yields more precise estimates of the parameters with all standard errors reduced substantially. In the Factor GMM model the labor share is positive and significant at the 10 percent level. However, one needs to keep in mind that in the case of weak instruments point estimates are unreliable. Further, it needs to be taken into account that using conventional two-step procedures after pretesting for identification is not recommended, as the size of such methods cannot be controlled (see e.g. Andrews, Moreira and Stock, 2006). Similar to Kleibergen and Mavroeidis (2009) we thus rely on two-dimensional confidence sets found by inverting the weak IV robust K-LM statistic (see Section 1.3.1), which does not seem to display a serious power loss in the case of strong instruments (Kleibergen, 2005). The fact that the factor-based confidence sets are smaller than the benchmark results provides evidence that our point estimates are more likely to be reliable.

Galí and Gertler (1999) and Kleibergen and Mavroeidis (2009) emphasize that a restricted model, where $\gamma_f + \gamma_b = 1$, performs well. Given that our point estimates support these findings we follow the approach by Kleibergen and Mavroeidis (2009) and from here on focus on the restricted model.[12]

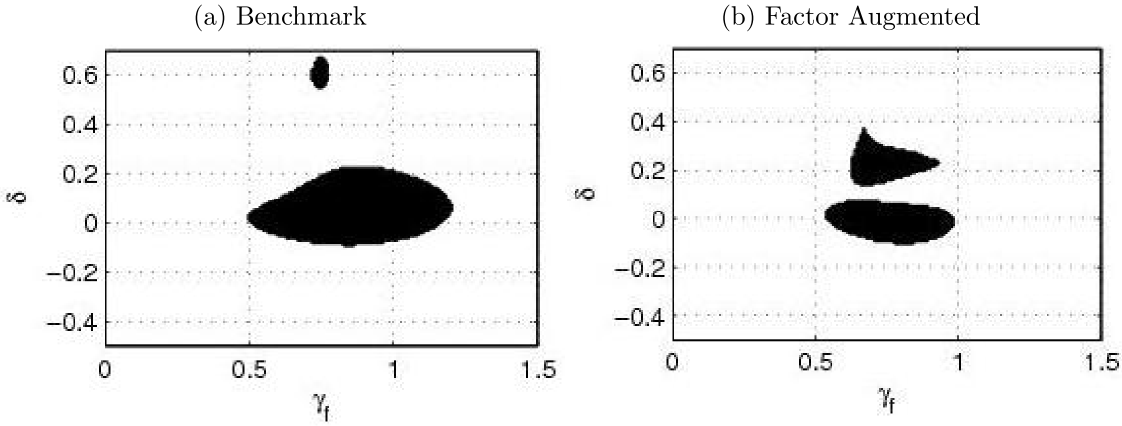
Figure 1.1 shows the joint confidence sets at 95 percent significance for both the benchmark and the factor-based instrument set.[13] These sets contain all values of (γ_f, δ) that cannot be rejected by the K-LM test. The shape of the K-LM sets may seem unconventional. However, note that confidence sets can be nonconvex and unbounded if based on the K-LM statistic as explained by Kleibergen (2005).

The robust confidence set based on the benchmark instrument as shown in Figure 1.1(a) is so large as to be in line with both dominant forward- and backward-looking dynamics. Further, the K-LM test cannot reject parameter values of $1 < \gamma_f \leq 1.2$ which would imply a negative backward-looking coefficient. The largest part of the confidence set lies around a value of zero for the coefficient on the labor share δ, indicating that the NKPC is relatively flat and that identification problems are present as explained in Kleibergen and Mavroeidis (2009). A small outlier part

[12]Kleibergen and Mavroeidis (2009) argue that inflation can be nonstationary and hence for the restricted model the use of lags of π_t as instruments may violate the conditions necessary for asymptotic theory to apply. In order to control for this possibility we instead use lags of $\Delta\pi_t$ in the restricted model as suggested by the authors.

[13]Figure 1.1 is constructed using MATLAB and the code by Kleibergen and Mavroeidis (2009). The factors are added as additional instruments.

Figure 1.1: 95 percent weak-identification robust confidence sets
for the coefficients of the NKPC



Note: The figure shows weak identification robust confidence sets for the coefficients (γ_f, δ) of the NKPC, as specified in equation (1.2) under the restriction that $\gamma_f + \gamma_b = 1$ for the period 1961:I to 2006:I using quarterly data. The left part shows the K-LM set using the benchmark instrument set comprising two lags of the first difference in inflation and three lags of the labor share. The right part depicts the K-LM set with lags one to four of the factors as additional instruments.

of the K-LM set lies around a value of $\delta = 0.6$.

Figure 1.1(b) provides evidence that adding factors to the instrument set can improve on the estimation as the resulting confidence set is smaller than in the benchmark case. Containing only values of γ_f between 0.54 and 0.98 it provides evidence for dominant forward-looking dynamics. Further, the outlier region has vanished from the confidence set such that the range of values for δ not rejected by the K-LM test is greatly reduced. However, as before a value of $\delta = 0$ cannot be rejected at 95 percent significance. This finding highlights that the NKPC is relatively flat resulting in identification problems for the coefficient on the marginal cost measure, as stressed in the previous literature: e.g. Woodford (2003); Kleibergen and Mavroeidis (2009); Kapetanios and Marcellino (2010).

1.4.2 Taylor Rule

We estimate equation (1.4) using the same time periods and methods as Mavroeidis (2010), i.e. GMM with Newey-West weight matrix, and expand the benchmark instrument set by lags of the factors in order to achieve more precise estimation

Table 1.2: Point estimates for the parameters of the Taylor rule

	Time period (in quarters)					
	1961:I-1979:II		1979:III-1997:IV		1987:III-2006:I	
	BM	Factor GMM	BM	Factor GMM	BM	Factor GMM
α	0.54***	0.76***	0.16	0.36***	−0.18	−0.07
	(0.18)	(0.08)	(0.19)	(0.13)	(0.18)	(0.12)
ψ_π	0.86***	0.83***	2.24***	1.91***	2.80***	2.80***
	(0.07)	(0.03)	(0.32)	(0.18)	(0.65)	(0.68)
ψ_x	0.29***	0.19***	0.82*	0.84***	1.43***	1.54***
	(0.10)	(0.04)	(0.43)	(0.20)	(0.28)	(0.26)
ρ	0.68***	0.57***	0.83***	0.83***	0.89***	0.92***
	(0.10)	(0.04)	(0.05)	(0.03)	(0.02)	(0.01)

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Standard errors are in brackets. Estimation of the Taylor rule, equation (1.4), is conducted by GMM using Newey-West weight matrix. BM refers to the results based on the benchmark instrument set comprising four lags of each inflation, the interest rate and the output gap. The Factor-GMM results are generated extending the instrument set by lags one to four of the factors derived before.

results.[14] The S sets are nonempty for both instrument sets and both periods considered providing evidence for the validity of the overidentifying restrictions.

For illustrative purposes point estimates for our specification are presented in Table 1.2. Note, that the Factor-GMM results closely resemble the evidence by Favero, Marcellino, and Neglia (2005).[15] The results based on the benchmark instrument set are similar in spirit to Clarida, Galí, and Gertler (2000).[16] The confidence sets based on the K-LM statistic discussed below provide evidence that the new instrument set is stronger and hence factor-based point estimates are more likely to be reliable. One should keep in mind, though, that in the presence of weak instruments point estimates are inconsistent and standard errors are not reliable. What stands out from the results is the substantial reduction in standard errors by roughly 50 percent for the first and second period and all coefficients. Consequently, in our specification all estimated coefficients (but α) are significant at the 1 percent level. The point estimates indicate that there is a shift in the conduct of monetary policy from the

[14]Note that there are papers stressing the importance of using real-time rather than final revised data, e.g. Orphanides (2001). This is not a concern for our study, as we are interested in the actual feedback coefficients rather than the intended ones.

[15]Favero, Marcellino, and Neglia (2005) estimate a forward-looking Taylor rule for the US from 1979:I to 1998:IV. In contrast to them, however, we use a different benchmark instrument set, a different data set for generating the factors and also consider the pre-Volcker and Greenspan period.

[16]In contrast to Clarida, Galí, and Gertler (2000), though, we leave out the three additional instruments commodity price inflation, M2 growth and the spread between the long-term bond rate and the three-month Treasury Bill rate, as Mavroeidis (2010) does in his analysis. We verify that this does not influence the main results significantly.

first period to the second. While the feedback coefficients (ψ_π, ψ_x) in the pre-Volcker regime are estimated to be $(0.83, 0.19)$, their estimates increase to $(1.91, 0.84)$ in the Volcker-Greenspan regime. These results already point to a more aggressive response of monetary policy to inflation and the output gap in the second period. To get information about the more recent stance of monetary policy, we also include a third period, which coincides with the Greenspan regime, 1987:III to 2006:I. Monetary policy under Greenspan seems to be characterized by a high degree of smoothing ($\rho = 0.92$), as also noted by Mavroeidis (2010), and an even stronger response to inflation and the output gap. The standard errors of the feedback coefficients are larger for this period, which is probably a result of the increased persistence of the policy rate (see Mavroeidis, 2010).

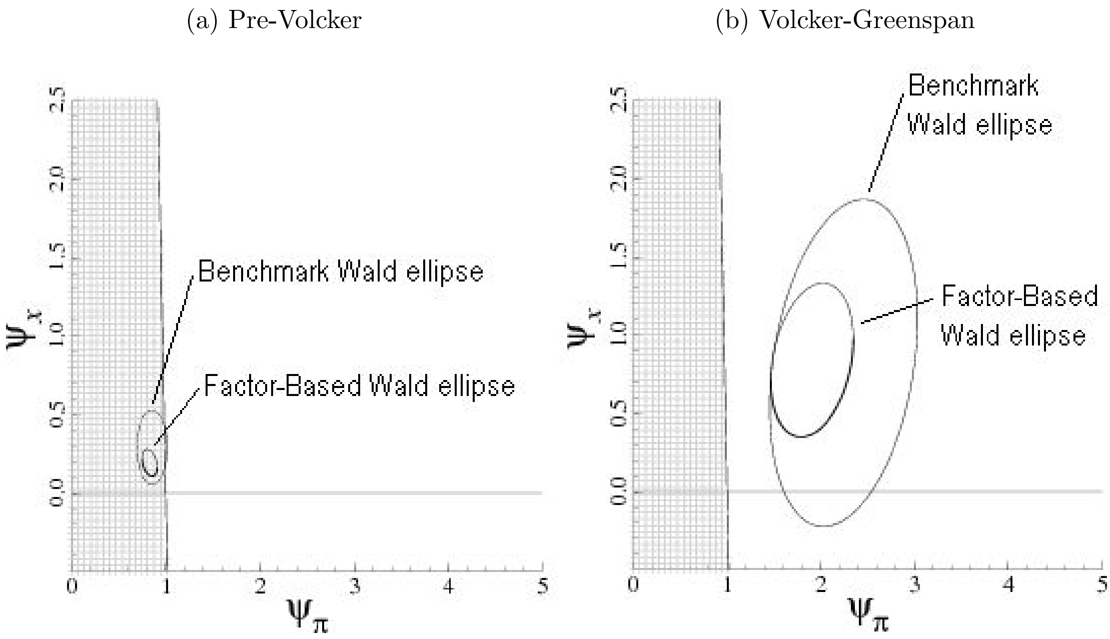
In order to be able to draw conclusions with respect to the Taylor principle, however, we consider joint estimates of the feedback coefficients. Figure 1.2 shows the Wald ellipses for the two parameters of interest, i.e. ψ_x and ψ_π, based on the point estimates presented before.[17] Interpreting their results Clarida, Galí, and Gertler (2000) and Mavroeidis (2010) assume that the degree of nominal rigidities λ and the discount factor β are equal to 0.3 and 0.99, respectively. They argue that these assumptions are in line with empirical evidence and we stick to them for comparability, verifying that they do not influence our main conclusions. The almost vertical line represents equation (1.5), i.e. the Taylor principle, under these assumptions, and is thus the boundary between indeterminacy (to the left) and determinacy (to the right).

For both periods discussed the factor-based Wald ellipse lies firmly within the ellipse based on the original instrument set. As presented in Figure 1.2(a), the pre-Volcker regime Wald ellipses are both located in the indeterminacy region. In contrast to that, the ellipses for the Volcker-Greenspan period have shifted to the determinacy region, as shown in Figure 1.2(b). These results provide evidence that the Taylor principle is satisfied under Volcker-Greenspan, while it has been violated before.

However, in the presence of weak instruments point estimates are inconsistent

[17] Figures 1.2 and 1.3 are constructed using the programming language Ox, see Doornik (2007), and the code by Mavroeidis (2010). The factors are added as additional instruments.

Figure 1.2: 95 percent Wald ellipses for the feedback coefficients of the Taylor rule



Note: The Wald ellipses for the feedback coefficients (ψ_π, ψ_x) of the Taylor rule, as specified in equation (1.4), are constructed using GMM with four lags of the instruments and Newey-West weight matrix. The benchmark Wald ellipses are based on the point estimates similar to those by Clarida, Galí, and Gertler (2000), where the instrument set comprises four lags of each inflation, the interest rate and the output gap. The factor-based results are generated extending the instrument set by lags one to four of the factors derived before. The almost vertical line represents equation (1.5), i.e. the Taylor principle with $\lambda = 0.3$ and $\beta = 0.99$, being the boundary between indeterminacy (to the left) and determinacy (to the right).

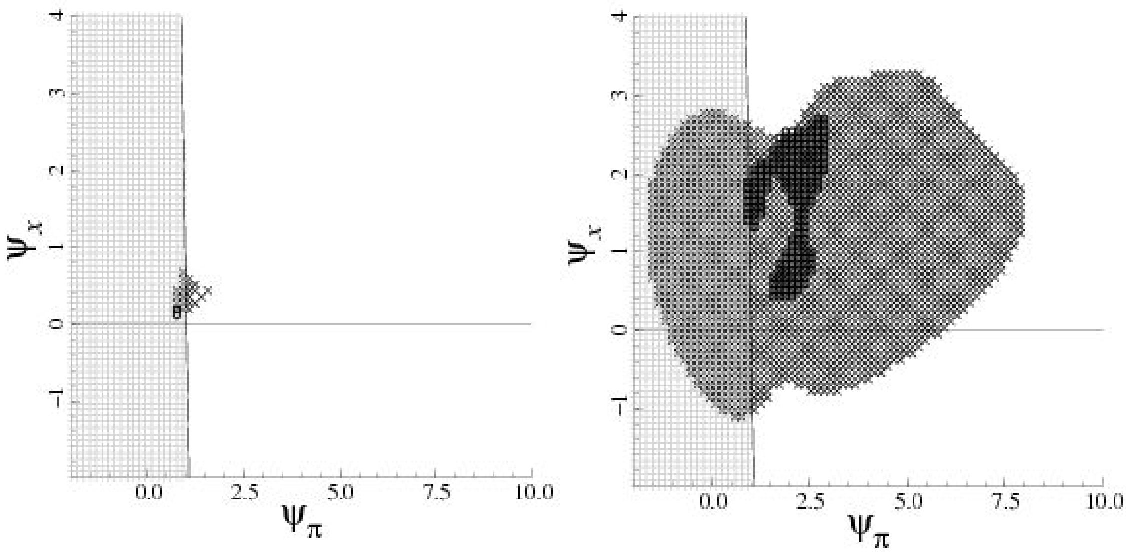
resulting in unreliable Wald ellipses. Therefore, we rely on the weak IV robust K-LM test which guarantees comparability with the results of Mavroeidis (2010). Figure 1.3 shows the factor-based joint confidence sets at 95 percent significance for both subsamples (dark grey areas). For comparison we include the results from Mavroeidis (2010), namely the weak IV robust confidence sets, constructed with the benchmark instrument set (light grey areas). These sets contain all values of (ψ_π, ψ_x) that cannot be rejected by the K-LM test.

Figure 1.3(a) provides further evidence that pre-Volcker monetary policy was not adherent to the Taylor principle, as the Factor-GMM confidence set also lies within the indeterminacy region. The large reduction in the size of the confidence set for the second period corroborates our finding that the factors contain relevant

Figure 1.3: 95 percent weak-identification robust confidence sets
for the feedback coefficients of the Taylor rule

(a) Pre-Volcker

(b) Volcker-Greenspan



Note: The figure shows weak identification robust confidence sets for the feedback coefficients (ψ_π, ψ_x) of the Taylor rule, as specified in equation (1.4). The light grey areas (crosses) represent the K-LM sets as estimated by Mavroeidis (2010) using the benchmark instrument set comprising four lags of each inflation, the interest rate and the output gap. The dark grey areas (circles) are the K-LM sets with lags one to four of the factors as additional instruments. The almost vertical line represents equation (1.5), i.e. the Taylor principle with $\lambda = 0.3$ and $\beta = 0.99$, being the boundary between indeterminacy (to the left) and determinacy (to the right).

information for the estimation. Most importantly, our confidence set clearly lies outside the indeterminacy region, while in contrast to that, Mavroeidis' confidence set for this time period has a considerable part in this very area and his results are even consistent with negative values for both parameters. A substantial part of our confidence set is located around the point estimate of $(\widehat{\psi}_\pi, \widehat{\psi}_x) = (1.91, 0.84)$, whereas another part lies above it, showing that there is some remaining uncertainty with respect to the feedback coefficients of the Taylor rule. Our findings highlight that with the inclusion of additional important information it can be empirically shown that monetary policy conduct under Volcker and Greenspan was more aggressive towards fighting inflation than pre-Volcker and thus satisfied the Taylor principle.[18]

[18] A decrease in λ or β would rotate the boundary of the indeterminacy region counterclockwise around the intersection with the horizontal axis as explained by Mavroeidis (2010). For all admissible values a change in either parameter would not alter our conclusion of determinacy for the second period as our confidence sets are already to the right of the boundary. Similarly, given our

The results with fewer factors or lags are less precise, but go in the same direction, i.e. a shift outwards from the indeterminacy region, while with more factors the results are comparable. Results using the weak IV robust MQLR statistic (see Section 1.3.1) rather than the K-LM statistic are very similar providing evidence for the robustness of our findings. With the use of more recent data, i.e. until 2006:I, the confidence sets shift more towards the indeterminacy region, suggesting that there might have been some time variation in the conduct of monetary policy under Alan Greenspan.[19]

Our results corroborate the empirical evidence by Lubik and Schorfheide (2004), Coibion and Gorodnichenko (2011), Boivin and Giannoni (2006) or Inoue and Rossi (2011), among others. Using Bayesian methods, Lubik and Schorfheide (2004) estimate the parameters of the whole model that underlies our single-equation estimation, whereas Coibion and Gorodnichenko (2011) analyze a similar model under the assumption of a positive and time-varying inflation trend. Boivin and Giannoni (2006) examine the monetary transmission mechanism using a vector autoregressive framework. Albeit the different approaches, these studies find a move of the US economy from indeterminacy to determinacy as a result of a more aggressive monetary policy regime. Inoue and Rossi (2011) use both DSGE models and vector autoregressions allowing for structural breaks in all parameters and show that changes in monetary policy parameters have, among other factors, led to the Great Moderation.

1.4.3 The Number of Instruments

Comparing results based on the benchmark instrument set with those using a larger factor-based instrument set raises the question whether it is the information from the factors or just the increased number of instruments that causes the extra precision in the estimation of the Taylor rule for the Volcker-Greenspan period (Figure 1.3(b)).[20] In order to demonstrate that it is the former rather than the latter, we fix the number of instruments to be equal to the benchmark case for the comparison. These instruments are selected by means of hard thresholding as suggested by Bai and Ng

estimation results, for the first period λ would have to be smaller than 0.01 to change our finding of indeterminacy.

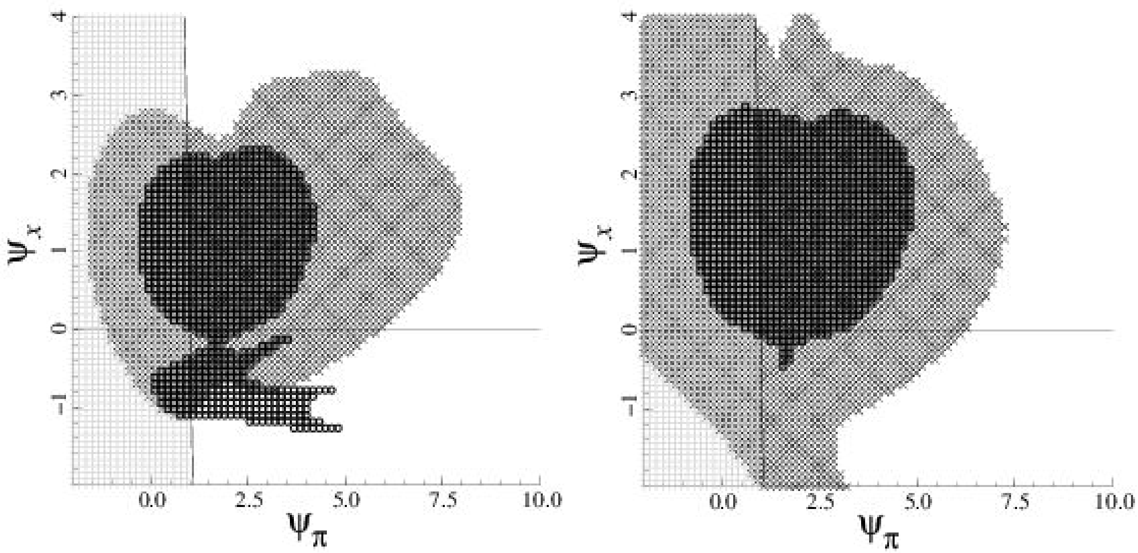
[19] The results for these alternative specifications are available from the authors upon request.

[20] We thank an anonymous referee for pointing this out.

Figure 1.4: 95 percent weak-identification robust confidence sets
for the feedback coefficients of the Taylor rule with selected instruments

(a) Combined K-LM statistic

(b) MQLR statistic



Note: The figure shows weak identification robust confidence sets for the feedback coefficients (ψ_π, ψ_x) of the Taylor rule, as specified in equation (1.4), for the Volcker-Greenspan period. The light grey areas (crosses) represent the confidence sets as estimated by Mavroeidis (2010) using the benchmark instrument set comprising four lags of each inflation, the interest rate and the output gap. The dark grey areas (circles) are the confidence sets with the instruments selected by means of hard thresholding, namely the exogenous first lag of the interest rate, the first four lags of each inflation and the output gap, and the second lag of factor one and two and the forth lag of factor two. Figure 1.4(a) and (b) show results based on the combined K-LM statistic and the MQLR statistic, respectively. The almost vertical line represents equation (1.5), i.e. the Taylor principle with $\lambda = 0.3$ and $\beta = 0.99$, being the boundary between indeterminacy (to the left) and determinacy (to the right).

(2008) which amounts to ranking all instruments by their explanatory power for the endogenous variables (see Appendix B for more details). In the following analysis, the twelve highest-ranked instruments from the factor-based set are used, leading to an instrument set of the same size as in the benchmark case. This procedure yields the following instruments: Apart from the exogenous first lag of the interest rate, the first four lags of inflation and the output gap are included which does not come as a surprise given the relative persistence in either variable. Further, the second lag of the first two factors and the fourth lag of factor four are selected.

Confidence sets for the combined K-LM statistic and the MQLR statistic based on these twelve instruments are presented in Figure 1.4. The results are more precise than the results using the benchmark instrument set of the same size where a higher relative precision is even clearer for the confidence set based on the MQLR statistic. This highlights that the factors contain relevant information for inflation and the output gap and thus it is not just the increased number of instruments which drives the results in Figure 1.3(b).[21]

1.4.4 Using Survey Expectations as Instruments

Results for the Taylor rule estimates during the Volcker-Greenspan period indicate that parameters are still somewhat imprecisely estimated. Given that expectations of future inflation are available from surveys these should have explanatory power for actual realizations. Ang, Bekaert, and Wei (2007) show that inflation surveys are successful in forecasting inflation out-of-sample over the next year. Moreover, Coibion (2010a) and Adam and Padula (2011) estimate different versions of the Phillips Curve, where they replace expected future inflation by expectations from the Survey of Professional Forecasters (SPF) arguing that this approach yields plausible estimates. Similarly, Orphanides (2004) estimates Taylor rules where he replaces expected future inflation by Greenbook forecasts for the specific horizons.

In order to further improve results, we use survey expectations in two different ways in our estimation procedure. On the one hand, we expand the factor-augmented instrument set by one lag of the mean of expected inflation two-periods ahead, i.e. $\mathbb{S}_{t-1}\pi_{t+1}$, and one lag of the mean of expected output growth one-period ahead from the SPF, i.e. $\mathbb{S}_{t-1}g_{y,t}$ (see the data appendix for details).[22] On the other hand, we expand the variable set in the factor model by the two survey variables from the SPF. We estimate four factors from the survey-augmented data set and add their first four lags to the benchmark instrument set.

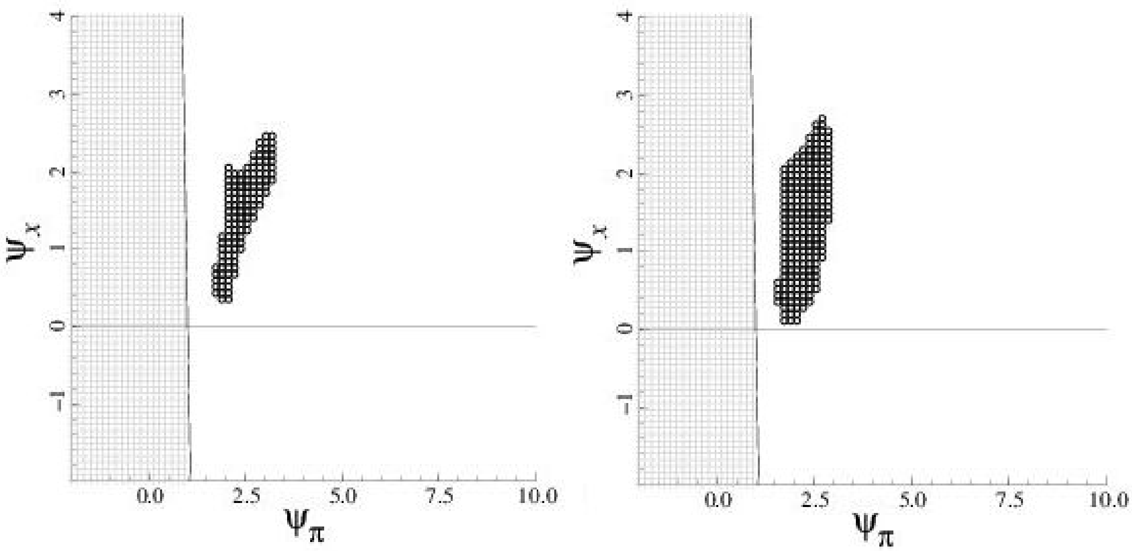
Figure 1.5 shows the results for these two specifications which are rather similar. In comparison to the factor-based results, the estimated output gap coefficient ψ_x is

[21] The oddly-shaped lower part of the confidence region based on the K-LM statistic below the x-axis is related to the fact that the behavior of the K statistic is spurious around inflection points and extrema. Increasing the weight on the J statistic ensures that this region vanishes.

[22] Given that expected output gaps are not provided we also construct expected output gap estimates by using the one-sided Christiano-Fitzgerald filter (2003), however, this does not change the main results. We also use median values rather than means, however, this does not seem to have substantial influence either.

Figure 1.5: 95 percent weak-identification robust confidence sets for the feedback coefficients of the Taylor rule with survey data

(a) Adding survey expectations (b) Survey-augmented factors



Note: The figure shows weak identification robust confidence sets for the feedback coefficients (ψ_π, ψ_x) of the Taylor rule, as specified in equation (1.4), for the Volcker-Greenspan period. The left graph shows the K-LM set estimated using the factor-based instrument set (see notes of Figure 1.3) expanded by $\mathbb{S}_{t-1}\pi_{t+1}$ and $\mathbb{S}_{t-1}g_{y,t}$ taken from SPF (mean values). The right graph depicts the results, where the variable set in the factor model has been expanded by the variables mentioned before. The almost vertical line represents equation (1.5), i.e. the Taylor principle with $\lambda = 0.3$ and $\beta = 0.99$, being the boundary between indeterminacy (to the left) and determinacy (to the right).

essentially unaffected. The estimate of the parameter on expected future inflation ψ_π is more precise resulting in confidence sets that are more clearly located in the determinacy region. We also use the Greenbook forecasts provided by the Federal Reserve for the variables discussed before instead of those from the SPF. Given that the results are very similar, we omit them here.[23]

[23] We also estimate a version, where we extend the benchmark instrument set by lags of the survey variables rather than the factors. However, it turns out that the factors yield much more precise estimates. This finding could be explained by the evidence of Nunes (2010), who shows that rational expectations play a more dominant role in inflation dynamics than do survey expectations. Also, Coibion (2010a) shows that surveys consistently overestimated inflation in the 1980's and 1990's. A different reason could relate to the fact the we use revised data, whereas the surveys contain real-time expectations. It may thus be the case that surveys are more informative in predicting variables in real-time. Finally, expected future output growth does not seem to be very informative with respect to future output gaps.

1.5 Conclusion

In this paper, we conduct factor-based inference of the hybrid New Keynesian Phillips Curve and a forward-looking version of the Taylor rule, as analyzed by Kleibergen and Mavroeidis (2009) and Mavroeidis (2010), respectively. These authors evaluate the models by using weak-identification robust methods. However, both studies find large confidence sets such that reliable interpretation of the estimated parameters is impaired. Therefore, we propose to employ factors generated from a large macroeconomic data set as additional instruments. The inclusion of these factors in the estimation procedure reduces the size of weak-identification robust confidence sets substantially. On the one hand, we show that forward-looking dominate backward-looking dynamics in the NKPC, while the curve is so flat that we cannot exclude a coefficient of zero on the marginal cost measure. On the other hand, our results with respect to the Taylor rule allow us to conclude that monetary policy in the after-1979 Volcker-Greenspan period satisfied the Taylor principle and thus contributed to containing inflation dynamics from there on. Our paper highlights that Factor GMM can be a useful tool to overcome the weak-identification problem common to many macroeconomic applications.

A1 Appendix to Chapter 1

A1.1 Data Appendix

New Keynesian Phillips Curve

For the estimation of the NKPC we use quarterly US data for the GDP deflator and the labor share from 1960:I to 2006:II from Kleibergen and Mavroeidis (2009).

Website:
http://www.econ.brown.edu/fac/Frank_Kleibergen/

Taylor Rule

For the estimation of the Taylor rule we use the same data set as Mavroeidis (2010). It consists of the federal funds rate, the annualized quarter-on-quarter inflation rate based on the seasonally adjusted GDP deflator and the CBO output gap for the US. Data is of quarterly frequency from 1960:I to 2006:II.

Website:
http://www.aeaweb.org/aer/data/mar2010/20071447_data.zip

Factor Data

For generating the factors we use quarterly data for the US from 1959:III to 2006:IV by Stock and Watson (2008), which is an updated version of the data they use for former papers, e.g. Stock and Watson (2002). Details for the 109 quarterly time series that have strong information content with respect to inflation and output, as well as the transformations needed to guarantee stationarity are provided by Stock and Watson (2008) in the data appendix of their paper.

Website:
http://www.princeton.edu/ mwatson/papers/hendryfestschrift_stockwatson_April 282008.pdf

Survey Data

The survey data can be downloaded from the Philadelphia FED. From the SPF we use mean two-quarter ahead expectations of the growth rate of the GDP deflator (dpgdp4) and mean one-quarter ahead expectations of GDP growth (rgdp3). The same variables are used from the Greenbook forecasts (i.e. PGDPdot4 and RGDP-dot3).

Websites:
http://www.phil.frb.org/research-and-data/real-time-center/
survey-of-professional-forecasters/
http://www.philadelphiafed.org/research-and-data/real-time-center/
greenbook-data/philadelphia-data-set.cfm

A1.2 Hard Thresholding

To order the instruments for the Taylor rule we conduct hard thresholding as suggested by Bai and Ng (2008). Hard thresholding amounts to ranking the instruments by their explanatory power for the endogenous variables. The estimation equation for this is:

$$X_{end,t} = \gamma_0 + \gamma_1 X_{exo,t} + \gamma_{2,i} Z_{i,t} + \eta_{i,t} \tag{1.8}$$

The endogenous variable $X_{end,t}$ is regressed on a constant, the exogenous variables $X_{exo,t}$ (the lagged policy rate in our case) and an instrument $Z_{i,t}$. The error term $\eta_{i,t}$ is assumed to be i.i.d. This equation is estimated for both endogenous variables π_{t+1} and x_t and for all instruments $i = 1, \ldots, 27$. For both endogenous variables we develop a ranking of all instruments according to the t statistic for their respective coefficients $\gamma_{2,i}$. For the instrument set in the estimation of the Taylor rule we always include the exogenous variable and first add the highest ranked variable of the regression on π_{t+1} followed by the highest ranked from the regression on x_t that is not yet included. We proceed in this way until we have the number of instruments desired. We start with an instrument from the regression on π_{t+1} given that from the first stage R^2 is seems that it is more difficult to predict inflation than the output gap (see Table 1.3 for the resulting ranking of the instruments).

Table 1.3: Hard thresholding for the Taylor rule

No of instruments	instrument name	variable	ranking
1	fyff_l1		exogenous
2	infl_l1	infl	1
3	gap_l1	gap	1
4	infl_l2	infl	2
5	gap_l2	gap	2
6	infl_l3	infl	3
7	gap_l3	gap	3
8	infl_l4	infl	4
9	gap_l4	gap	4
10	fac2_l2	infl	5
11	fac1_l2	gap	5
12	fac2_l4	infl	6
13	fyff_l4	gap	6
14	fac2_l1	infl	7
15	fac1_l1	gap	7
16	fac2_l3	infl	8
17	fac1_l3	gap	8
18	fac4_l2	infl	13
19	fac1_l4	gap	9
20	fac3_l1	infl	17
21	fyff_l3	gap	10
22	fac4_l4	infl	18
23	fyff_l2	gap	11
24	fac4_l1	infl	19
25	fac4_l3	gap	14
26	fac3_l2	infl	22
27	fac3_l4	gap	22
28	fac3_l3	infl	27

Abbreviations: infl=inflation, gap=output gap, fyff=interest rate, var_li=i-th lag of var, fac$i$=i-th factor. This table shows the ranking from hard thresholding of the instruments for the Taylor rule. The first column presents the final ranking, the second gives the name of the variable and the last two show the ranking of it for either inflation or the output gap.

Chapter 2

Forecasting Euro Area Inflation with the Phillips Curve

2.1 Introduction

The purpose of this study is to investigate the forecasting ability of various Phillips curve specifications for euro area inflation. By Phillips curve we understand here a reduced form relationship between inflation and economic activity or economic slack.

The recent financial crisis followed by a world-wide recession, particularly visible in large output losses in many countries, has been often accompanied by relatively resilient inflation rates. This has renewed interest in academia, central banks as well as the private sector in the Phillips curve relationship and what it might imply for future inflation (see e.g. Stock and Watson, 2010; Fuhrer, Olivei, and Tootell, 2009, 2011; Dale, 2012; IMF, 2013). Issues of interest include, for example, the stability of the relationship, the role of inflation expectations, implications for monetary policy or, which is the focus of this paper, whether the relationship can be exploited to forecast inflation.

Several studies of the forecasting performance of Phillips curves have been recently undertaken for the US (see e.g., Stock and Watson, 2009, 2010; Faust and Wright, 2012; Dotsey, Fujita, and Stark, 2011; Clark and Doh, 2011). The evidence suggests that superior performance of this type of models relative to simple benchmarks is episodic and often related to certain states of the economy (e.g., extreme values of output/unemployment gap, recessions). The choice of "activity" variable is secondary but some variables are more useful than others. Given the uncertainty with respect to the best specification, forecast combinations help. Finally, accounting for a changing trend in inflation can improve forecast accuracy.

The evidence for the euro area is much scarcer, with the available studies often having a different focus. The euro area is an interesting case for various reasons. It is a relatively young economy with around 15 years of common monetary policy, which have been preceded by gradual declines in inflation in many countries in the run-up to the euro adoption. It has particular structural features related to, e.g., rigidities of the labour market. Last but not least, it is a challenging case to study due to shorter history of available data.

The contribution of this paper is to provide comprehensive evidence on the forecasting performance of Phillips curve type models for the euro area economy. We consider different inflation measures, evaluation samples and model specifications. Some important issues in Phillips curve modelling relate to uncertainty with respect to the relevant explanatory variables and to stability of the relationship. Various indicators come into consideration when proxying for the unobserved real marginal costs, depending on the cycle, the economy in question or the time frame. For similar reasons there is uncertainty around the variables that should reflect cost push shocks (see e.g. Gordon, 1982, 1990; Stock and Watson, 1999, 2009). We consider a wide range of explanatory variables. We also evaluate different econometric specifications, namely autoregressive distributed lag models and vector autoregressions, different lag selection methods and both recursive and rolling estimation schemes.

In view of many possible model specifications and of potential instability of the Phillips curve relationship, several studies have suggested forecast combinations, see e.g. Stock and Watson (2009). In addition to the combination strategies proposed in that study we also employ information-theoretic averaging that Kapetanios, Labhard, and Price (2008) show to perform well.

An important element we focus on is the role of inflation detrending for forecast accuracy. Some recent studies indicate that accounting for a trend or time-varying mean of inflation can lead to improvements in forecast accuracy (see e.g., Clark and McCracken, 2010; Faust and Wright, 2012; Clark and Doh, 2011). We evaluate various “statistical” trends, namely exponentially-weighted moving averages of past inflation rates, but also consider long-run inflation expectations available from Consensus Economics. While the importance of inflation expectations as determinants of future inflation have been stressed both in the theoretical literature and in central bank communication, they have not been often considered in forecasting applications. Exceptions, apart from the papers just cited, include Ang, Bekaert, and Wei (2007), Wright (2012) or Koop and Korobilis (2012) who include inflation expectations from surveys as a proxy for actual expectations. For the case with constant intercept we consider both the specification with and without the unit root imposed. The former has been adopted in many studies, e.g., Stock and Watson (1999, 2009).

Questions we address are, inter alia, whether the inclusion of marginal cost measures (and supply shocks) leads to forecast improvements relative to univariate models and which are the variables that yield the lowest forecast errors. We examine whether detrending offers improvements in terms of lower forecast errors relative to a model with a constant mean of inflation and which approach is the most promising in modelling the trend. Finally, we aim at answering the question whether an average over different models can help to overcome the problem of model heterogeneity and instability and which combination approach leads to the best results.

In the following we provide a short preview of the results. As expected no one single model (category) consistently outperforms the rest, while a few results on the individual model categories deserve attention. For all evaluation periods considered either a model related to output or related to the unemployment rate yields the lowest forecast errors. This is an interesting result in light of the fact that the theoretical literature often models the Phillips curve equation including either output (or its gap) or the unemployment rate (or its gap) (see e.g., Woodford, 2003). Further, the inclusion of a supply shock proxy can improve the forecasting performance of our models for some episodes, while the best supply shock measure varies over time and differences with respect to models without this additional variable are not large. Finally, the particular econometric approaches considered, namely single-equation versus multiple-equation models, different lag selection methods, or different estimation windows, do not seem to influence the results dramatically.

Subtracting a time-varying mean of inflation before estimating a model and conducting the forecasts is particularly helpful in the early part of the sample, corresponding to the inflation convergence period, while recently the gains are more muted. We show that survey-based detrending yields relatively low forecast errors, while an exponentially-weighted moving average of inflation with low "forgetting" factor provides the best model-based alternative to estimate the trend. The specification with the unit root imposed performs best in the first part of the sample while its accuracy deteriorates strongly in the latter part.

Concerning forecast combination approaches we find that the simple average over all PC models beats the random walk benchmark on the sample after 2000, while before the performances are about comparable.[1] Performance-based averaging offers only temporary improvements over the simple mean of all models, while differences are typically not large. Thus, using a simple average of different model specifications provides a useful approach in an environment of large uncertainty with respect to the best predictor variables and econometric approaches.

The paper is organised as follows. In Section 2.2 we provide a short survey of the related literature. In Section 2.3 we present the different models used in the forecast evaluation and provide details of the different econometric approaches. In particular we discuss the combination strategies and detrending approaches used. In Section 2.4 we describe the data set we employ and in Section 2.5 we report the results. Section 3.6 is the conclusion.

2.2 Literature Review

An extensive review of the literature on forecasting inflation can be found in Stock and Watson (2009) or Faust and Wright (2012). Here we focus on papers more closely related to our work.

The first studies on forecasting inflation by means of Phillips curves have been conducted by Gordon (1982, 1990). He proposes a so-called Triangle model, whose name derives from the concept of inflation having three determinants, namely inflation persistence, a demand variable such as the unemployment rate and a supply shock. One of the best-known papers evaluating a wide range of model specifications

[1]Other univariate benchmarks are harder to beat such that gains from using Phillips curve type models are smaller.

in the spirit of the Phillips curve for forecasting inflation is Stock and Watson (1999). They find that for the period 1970-1996 Phillips curve models outperform univariate benchmark models in predicting one-year-ahead US inflation. The relevant activity variables seem to change over time and the authors find forecast combinations from their different models to perform better than the individual model predictions. The only exception is a model that makes use of a composite activity index based on 168 variables, as it cannot be improved upon by model combination.

Atkeson and Ohanian (2001) challenge the usefulness of Phillips curves in forecasting. They show that for the period 1984-1999 Phillips curves based on the non-accelerating inflation rate of unemployment – that could be related to an unemployment gap – or on an activity index cannot improve on forecasts from a naïve random walk benchmark. Fisher, Liu, and Zhou (2002) qualify this message arguing that the performance of Phillips curves depends on the sample period, the forecasting horizon, as well as the inflation measure chosen. They provide evidence that these models can improve over naïve benchmarks in times of volatile inflation and also they can predict the direction of changes in inflation relatively well. They argue that it is only in times of monetary regime change that model predictions based on economic activity might have no or only low explanatory power. Recent applications such as Stock and Watson (2009) and (2010) for the US provide evidence in favour of the general message of Fisher, Liu, and Zhou (2002), namely that the performance of Phillips curves depends heavily on the specification, the sample period and the phase of the business cycle. Given the latter finding, different studies advocate making Phillips curve specifications conditional on the state of the economy (see e.g., Fuhrer and Olivei, 2010; Dotsey, Fujita, and Stark, 2011).

As already mentioned, there is less work available for the euro area. Many studies focus on in-sample estimation, see e.g., Galí, Gertler, and López-Salido (2001); O'Reilly and Whelan (2005); Doepke et al. (2008); Paloviita (2008); Musso, Stracca, and van Dijk (2009), or more recently Montoya and Döhring (2011). There are less papers studying out-of-sample forecasting performance, see Rünstler (2002); Hubrich (2005); Canova (2007); Marcellino and Musso (2010); Buelens (2012). Present work extends on these studies along a number of dimensions, including the variety of model specifications, the evaluation sample and most importantly the various detrending and forecast combination approaches considered.

Several studies document various forms of time variation in the coefficients of Phillips curves, see e.g., Musso, Stracca, and van Dijk (2009) and the references therein. These authors report evidence supporting a time-variation in the mean and the slope of the euro area Phillips curve and propose to employ a smooth transition model. This is contrary to the results of O'Reilly and Whelan (2005) who do not find sufficient evidence to reject the hypothesis of stability in the reduced form Phillips curve coefficients and, in particular, in those related to inflation persistence. In terms of forecasting applications, Canova (2007) or Fuhrer, Olivei, and Tootell (2009), for example, show that allowing for time variation in the Phillips curve can improve forecast accuracy.

Following the studies by Stock and Watson (1999; 2009) many authors have resorted to forecast combination as a way to deal with model instability. Model averaging seems to be an adequate substitute for time-varying parameter models, while at the same time it provides a way to deal with many candidate models to construct the forecast. While standard approaches to averaging seem to perform well, more sophisticated methods of forecast combination such as Bayesian model averaging (see Wright, 2009) or information-theoretic model averaging (see Kapetanios, Labhard, and Price, 2008) offer improvements only in some cases. Recently, Clark and McCracken (2010) evaluate forecasts for inflation, output and interest rates from VAR using a wide range of estimation techniques and show three interesting results: first, model averaging appears the right strategy to deal with structural instabilities; second, equally weighted averages are consistently among the best averaging strategies; third detrending inflation and interest rates improves forecast accuracy.

In line with this last finding, Clark and Doh (2011) compare different detrending approaches and assess their usefulness in forecasting US inflation. They make use of both model-based trends and long-run survey expectations. They adopt a Bayesian approach and focus on univariate models, although they consider a version of a Phillips curve model. They show that the best approach varies over time and is prone to instabilities. They conclude that a model based on the survey trend is consistently among the best models, as is a local level model. Similarly, Faust and Wright (2012) show that using survey-based trends results in lower forecast errors in their specifications.

Finally, many applications differ in the statistical properties assumed for the inflation process. For example, Stock and Watson (1999, 2009) argue that US inflation is better modelled as an I(1) process, i.e., differences of inflation are used instead of levels in the forecasting equation. On the other hand, many forecasting studies do not impose this constraint such as Hubrich (2005); Canova (2007); Kapetanios, Labhard, and Price (2008); Wright (2009); Giannone, Lenza, Momferatou, and Onorante (2010) or Buelens (2012). Ang, Bekaert, and Wei (2007) compare the results for both assumptions of stationary and difference-stationary inflation. They find that the models under the two different assumptions perform comparably. While most forecasting applications with Phillips curves focus on direct forecasts from a version of the autoregressive distributed lag (ADL) model there are a few applications that consider VAR models (and iterated forecasts) (see e.g, Hubrich, 2005; Giannone et al., 2010; Clark and McCracken, 2010; Garratt, Mitchell, and Vahey, 2010; Benkovskis et al., 2011).

2.3 Econometric Framework

We denote by π_t^h an annualised h-period inflation rate:

$$\pi_t^h = \frac{400}{h} \ln \left(\frac{P_t}{P_{t-h}} \right) \tag{2.1}$$

where P_t is the appropriate (quarterly) price index. For simplicity, $\pi_t := \pi_t^1$ and hence $\pi_t^h = \frac{1}{h} \sum_{i=0}^{h-1} \pi_{t-i}$. The h-step ahead forecast given the information at time t is denoted as $\pi_{t+h|t}^h$.

All the models are estimated by ordinary least squares (OLS).

2.3.1 Phillips Curve Models

Let $\tilde{\pi}_t$ denote the detrended inflation rate, $\tilde{\pi}_t = \pi_t - \pi_t^{TR}$, $\tilde{\pi}_t^h = \frac{1}{h} \sum_{i=0}^{h-1} \tilde{\pi}_{t-i}$, where π_t^{TR} is the trend of inflation. We will be more specific on π_t^{TR} below. After the detrended inflation rate is forecasted, we add back the trend in order to construct the forecast errors with respect to the realised inflation rate.[2]

We consider the following two model classes:

Autoregressive Distributed Lag (ADL) Models

The general version of the model is:

$$\tilde{\pi}^h_{t+h} = \mu_h + \alpha_h(L)\tilde{\pi}_t + \beta_h(L)y_t + \gamma_h(L)z_t + \nu^h_{t+h}, \tag{2.2}$$

where y_t is a proxy of real marginal costs, (e.g., output gap) and z_t captures supply side shocks (e.g., oil prices). $\alpha_h(L)$, $\beta_h(L)$ and $\gamma_h(L)$ are lag polynomials. In some versions z_t is not included (i.e. $\gamma(L) = 0$). y_t and z_t are demeaned prior to estimation.

These models result in *direct* forecasts. This class of models have been the most widely used in forecasting applications, see e.g., Stock and Watson (1999, 2009).

Vector Auto-Regression (VAR) Models

To evaluate also *iterated* or *indirect* forecasts we use vector autoregressions:

$$X_t = \mu + \Phi(L)X_{t-1} + \nu_t\,, \tag{2.3}$$

where $X_t = [\tilde{\pi}_t \; y_t \; z_t]'$. As above, in some versions z_t is not included in the VAR.

For this model class it is more straightforward to do conditional forecasts or scenarios, see e.g., Giannone et al. (2010).

[2] For simplicity we assume that the inflation trend does not change over the forecast horizon, this is, we use the latest available point as the forecast of the trend. For future research it would interesting to relax that assumption and model how the trend evolves into the future.

2.3.2 Inflation Detrending

The trend is supposed to capture a time-varying mean of inflation. We evaluate different approaches considered in the literature (see e.g., Faust and Wright, 2012; Stock and Watson, 2009; Clark and Doh, 2011). On the one hand, we use "statistical" trends, based on past inflation rates, on the other hand, we rely on long-run survey expectations of inflation as an approximation for the current inflation trend. Altogether, we consider the following cases.

Constant Mean

In the first version of models considered we assume a constant mean inflation rate, i.e., at each point in time we just subtract the actual mean of inflation over the estimation sample up to this point from the inflation rate. This corresponds to the specification in equation (2.2) and (2.3) with $\mu_h = \mu \equiv 0$ where π_t^{TR} is just the mean of inflation over the estimation sample. Alternatively, it could be implemented by unconstrained μ_h and μ and $\pi_t^{TR} \equiv 0$.

Stock & Watson Approach

This approach, hereafter SW, amounts to estimating the models with inflation in differences. This is the version of (2.2) and (2.3) with the unit root imposed. For (2.2) it means that $\alpha(1) = 1$ and $\pi_t^{TR} \equiv 0$ while for the VAR class it amounts to setting $X_t = [\Delta\pi_t \; y_t \; z_t]'$. Alternatively, for $h = 1$ this is equivalent to unconstrained (2.2) and $\pi_t^{TR} = \pi_{t-1}$. This is the type of model considered in Stock and Watson (2007a, 2009) and closely related to the "triangle" model by Gordon (1982; 1990).

Unlike Clark and Doh (2011) we introduce a constant into these types of models such that they correspond to the ADL models as discussed in Stock and Watson (2009).[3]

[3] In the Appendix, we also analyse these models under the assumption that the constant is zero in order to find out whether the intercept is essential. This might for example be the case when inflation is on a downward path (as it happened to be, e.g., during 93-99; see Figure 2.1 below).

For the following approaches we take $\mu_h = \mu \equiv 0$ in equations (2.2) and (2.3).

EWMA Trend

The "statistical" or model-based trends are all derived by exponentially-weighted moving averages (EWMA) of inflation. Thus, the trend inflation rate is $\pi_t^{TR} = \phi \sum_{j=0}^{\infty}(1-\phi)^j \pi_{t-j}$, where ϕ is the smoothing parameter and can be thought of as a "forgetting" factor .

We consider cases with a fixed $(1-\phi)$ equal to either 0.95, 0.85 or 0.75[4] as well as ϕ estimated on the basis of an integrated moving average of order 1 (IMA(1,1)) representation of π_t, see the next point.[5]

Local Level Trend

As one of the trend specifications, Clark and Doh (2011) consider the unobserved components - stochastic volatility (UC-SV) model by Stock and Watson (2007a).[6] Here the inflation rate π_t depends on an unobserved random walk trend τ_t and an innovation η_t. Once we assume that the variances of the permanent and transitory shocks in the model ($\sigma_{\epsilon,t}$ and $\sigma_{\eta,t}$, respectively) are in a fixed ratio, inflation has an IMA(1,1) representation and the trend can be estimated as: $\hat{\tau}_{t|t} = \phi \sum_{j=0}^{\infty}(1-\phi)^j \pi_{t-j}$, i.e., once more the trend is an EWMA of inflation (see e.g. Pagan, 2009).

For simplicity we rely on this latter approach to identify the local level trend, as do, e.g., Clark and McCracken (2010). In this case, we estimate $(1-\phi)$ by fitting an IMA(1,1) model. Several studies have shown that the IMA model performs well relative to the full-blown UC-SV model (e.g., Stock and Watson, 2007a; Clark and

[4]These are typical values considered in the literature, e.g., as noted by Stock and Watson (2009), recently it is estimated to be around 0.85 in the US.

[5]An IMA(1,1) model of inflation looks as follows: $\pi_t - \pi_{t-1} = \epsilon_t - (1-\phi)\epsilon_{t-1}$.

[6]The UC-SV model looks as follows:

$$
\begin{aligned}
&\pi_t = \tau_t + \eta_t, \quad \text{where } \eta_t = \sigma_{\eta,t}\zeta_{\eta,t} \\
&\tau_t = \tau_{t-1} + \epsilon_t, \quad \text{where } \epsilon_t = \sigma_{\epsilon,t}\zeta_{\epsilon,t} \\
&ln(\sigma_{\eta,t})^2 = ln(\sigma_{\eta,t-1})^2 + \nu_{\eta,t}, \quad \text{where } \mathbb{E}_t(\nu_{\eta,t}) = 0 \\
&ln(\sigma_{\epsilon,t})^2 = ln(\sigma_{\epsilon,t-1})^2 + \nu_{\epsilon,t}, \quad \text{where } \mathbb{E}_t(\nu_{\epsilon,t}) = 0
\end{aligned}
$$

Doh, 2011).

LRSE Trend

We analyse the relevance of long-run survey expectations (LRSE) as an anchor of inflation. In this model inflation is detrended by long-run inflation expectations, which is aimed to account for a time-varying intercept in the Phillips curve relationship, as explained by Faust and Wright (2012). It is somewhat similar in spirit to Wright (2012), who uses inflation expectations as priors for the mean of inflation in a VAR. Long-run inflation expectations might be better suited than model-based trends to account for expected changes in policies, such as those adopted during the run-up to the introduction of the euro.

LRSE Trend - Bias Corrected

Here the bias-corrected version of long-run survey expectations is used, hereafter LRSEC. In other words, we first calculate the average deviation of the survey variable from actual inflation and then remove this bias from the survey variable and use the resulting measure as the trend.

Some studies use survey expectations with a shorter horizon in their forecasting exercises, see for example Ang, Bekaert, and Wei (2007) or Koop and Korobilis (2012) in studies on US inflation. It is, however, unclear how to incorporate short-term expectations in a Phillips curve forecasting equation and further, data availability poses a serious problem in the euro area.

2.3.3 Benchmarks

As the benchmark we take the random walk (RW) model of Atkeson and Ohanian (2001) where $\pi^h_{t+h|h} = \pi^4_t$. It is the simplest model we consider, as it does not require any additional information than the lagged inflation rate and it does not need to be estimated. Interestingly, according to Atkeson and Ohanian (2001) it provides a forecast for inflation that is hard to beat for other univariate or multivariate (i.e.

Phillips curve) models.

Some studies consider other univariate models than the random walk as a benchmark, e.g., an autoregressive process. Accordingly we also consider univariate versions of equations (2.2) and (2.3). Analysing their performance relative to the prediction errors of the Phillips curve models allows us to identify if the marginal cost measures or supply shocks provide any added value for the forecasting exercise.

2.3.4 Forecast Combination

Our relatively wide range of specifications of the Phillips curve is a reflection of the many existing theoretical frameworks and is, thus, linked to uncertainty about the appropriate formulation and variables to be used in estimation. To take into account this element of model uncertainty, we resort to forecast combination, comparing standard techniques (see Stock and Watson, 2009) and recently proposed information-theoretic averaging (Kapetanios, Labhard, and Price, 2008).

Standard Approaches

We compare standard combination techniques as discussed in Stock and Watson (2009): mean and median of individual model forecasts as well as weighted averages of forecasts based on their past performance, including the trimmed means. For performance-based weighted averaging, forecasts are constructed as $\Sigma_{i=1}^{n} \lambda_{it}\, \pi^h_{i,t+h|t}$, where $\pi^h_{i,t+h|t}$ are forecasts from model i and λ_{it} are the weights. Let $e^h_{i,t} = \pi^h_t - \pi^h_{i,t|t-h}$ denote the forecast errors of model i. The weights are chosen according to:

$$\lambda_{it} = (\frac{1}{\hat{\sigma}^2_{it}})/(\sum_{j=1}^{n} \frac{1}{\hat{\sigma}^2_{jt}}) \qquad \text{or} \qquad \lambda_{it} = (\frac{1}{\hat{\sigma}^2_{it}})^2/(\sum_{j=1}^{n} \frac{1}{\hat{\sigma}^2_{jt}})^2, \qquad \text{where}$$

$$\hat{\sigma}^2_{it}(\omega) = \sum_{j=0}^{\bar{T}} \omega^j \left(e^h_{i,t-j}\right)^2 .$$

For $\omega \neq 1$ more weight is attached to the more recent forecast errors and we consider $\omega \in \{1, 0.95, 0.9, 0.7, 0.5\}$. We take $\bar{T} = 40$. For performance-based trimmed-mean forecasts we average the best, in terms of the lowest Root Mean Squared Forecast Error (RMSFE) over the latest $\bar{T}$ quarters, 90% or 50% forecasts. We also consider model selection in which the forecast is obtained from the model that has performed best in the past (over 4, 8, 16 or 24 quarters).

We also evaluate equally-weighted forecast averages within some model categories, namely ADL versus VAR models, AIC versus BIC or fixed lag selection and rolling versus recursive estimation.

Information-Theoretic Averaging

Kapetanios, Labhard, and Price (2008) show that information-theoretic averaging is a strong rival for Bayesian techniques. They suggest a combination approach based on the AIC criterion (or BIC/SIC). This amounts to calculating the relative likelihood of each model i and model weights are then constructed as:

$$\lambda_i = \frac{exp(-1/2\psi_i)}{\sum_{i=1}^{n} exp(-1/2\psi_i)}, \quad \text{where} \quad \psi_i = \text{AIC}_i - \min_j \text{AIC}_j.$$

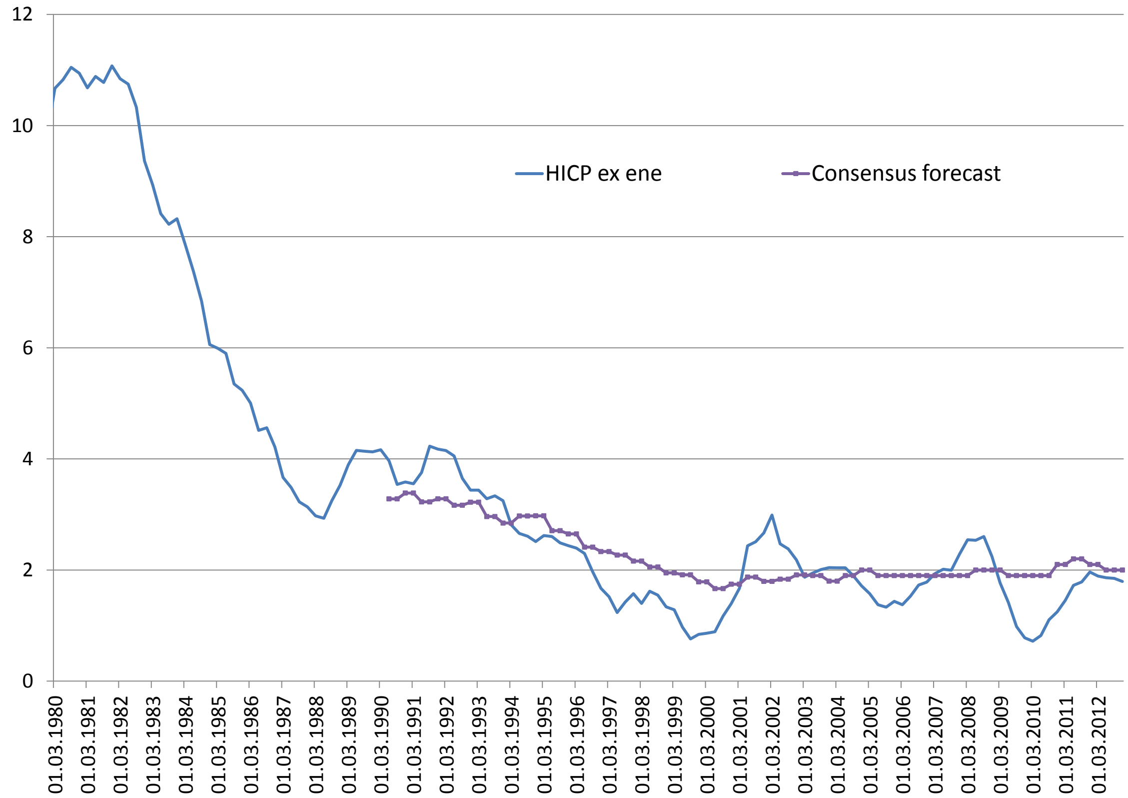
We evaluate the forecast errors of this combination approach based on both the AIC and the BIC criterion.

2.4 Data and Estimation

2.4.1 Data

We use quarterly euro area data. The details related to the available time span or transformations are provided in Table 2.4 in the Appendix. The sample covers 1980 to 2012. Some of the series were backdated using the latest version of the Area-Wide Model database, see Fagan, Hendry, and Mestre (2005) and some are available only later. As the inflation measure we consider the seasonally adjusted harmonized index for consumer prices excluding energy (HEX). Long-run inflation expectations refer to 6-10 years ahead forecasts for euro area inflation provided by Consensus Economics. As these are only published at semi-annual frequency we assume that they remain unchanged in the intermediate quarters.

Figure 2.1: Inflation series - 1980-2012



2.4.2 Details of the Exercise

We evaluate the different specifications in an out-of-sample exercise.[7]

For the estimation we consider both rolling and recursive schemes. For the former case we employ an estimation window of 10 years in order to allow parameters to change over time, while ensuring sufficient observations for reliable estimation.[8]

The lag length of predictors is chosen either by the AIC or the BIC criterion, but it is assumed that at least one lag of inflation and the explanatory variable(s) enters the specification. We allow for up to four lags in the multivariate and in the univariate models. Further, we also consider versions with a fixed number of lags equal to four.

We focus on forecasting performance of the models for the four-quarter-ahead horizon and consider the following evaluation samples: 1993-99, 2000-06 and 2007-

[7]In the evaluation we disregard issues such as data revisions or publication delays; for a discussion see e.g. Bańbura et al. (2012).

[8]For the specifications including predictors that are not available over the entire sample the estimation samples are shorter. In order to assess the robustness of our results to the choice of window size we also average over various window sizes, as suggested by Pesaran and Timmermann (2007).

12. These periods are rather different in terms of inflation dynamics as shown in Figure 2.1. The first sample corresponds to the run-up to the euro introduction and features declines in inflation rates in many euro area countries. The second period is characterised by relatively stable inflation rates. Finally, the last period witnessed elevated inflation rates on account of pass-through from food and oil price shocks followed by significant drops in inflation as a result of the financial crisis. Given that long-run survey expectations for the euro area are only available as of 1990:II, when analysing the survey-based detrending method, we focus only on the latter two evaluation samples.

We evaluate the models for HEX inflation, as mentioned before. Moreover, we consider the following twelve standard measures of marginal costs: unemployment rate (URX, level and difference), unemployment gap (URXgap), output growth (YER), output gap (YERgap), employment growth (LNN), employment gap (LNNgap), capacity utilisation (CPU, level and difference), industrial production growth (IPT) and the industrial production gap (IPTgap). The gaps are produced using the Christiano-Fitzgerald filter that is a nearly optimal one-sided band-pass filter, see Christiano and Fitzgerald (2003), where we keep the cycles shorter than 15 years. For these gaps we use both a demeaned and a not-demeaned version. Further, we make use of the unemployment recession gap (URXrec) that has been suggested by Stock and Watson (2010) and is supposed to work well as a predictor for inflation during recessions: $ur_t = u_t - min(u_t, u_{t-1}, \ldots, u_{t-11})$, where u_t is the unemployment rate URX. In addition, we employ the average of the before mentioned variables as a separate measure, as well as a principal component estimated from this same data set (in both cases the gap variables with non-zero mean are ignored) and finally a principal component of a different set of macroeconomic variables (see the Appendix for details).

Further, we include the following supply shock indicators: the UK Brent Crude Index (POE), the nominal effective exchange rate (EEN) and the imports of goods and services deflator (MTD).

In total we have 912 Phillips curve models and 12 univariate models[9] for each de-

[9]These are 76 ADL and 76 VAR specifications with lag length selection by either the AIC criterion, the BIC criterion or a fixed number of lags equal to four and with either a rolling or a recursive window. The 76 models include, for each of the 19 marginal cost measures, a version

trending approach, i.e., the constant mean specification, the SW approach, EWMA detrending with a fixed smoothing coefficient of either 0.95, 0.85 or 0.75, local level detrending (from an estimated IMA model) and survey-expectations based detrending both with and without bias correction (LRSE and LRSEC, respectively).

The criterion we employ to evaluate the results is the Root Mean Squared Forecast Error (RMSFE). Finally, we test for equal mean squared forecast errors by means of the widely-used Diebold-Mariano (1995) test with the random walk model as the benchmark. We compute the t-tests with a heteroskedasticity- and autocorrelation-consistent (HAC) variance using a quadratic spectral kernel (see Andrews, 1991).

2.5 Results

In this section we discuss the results of our various model specifications and econometric approaches. After presenting individual model results, we put a particular emphasis on the detrending methods used in our study and the averaging approaches employed. Finally, we analyse other aspects relevant for our forecasting exercise related to predictor variables, the estimation window, lag length selection and direct (ADL) versus indirect (VAR) forecasts.[10]

2.5.1 Individual Models

We start by comparing the forecasting performance of individual Phillips curve models as described in Section 2.3.1 and 2.3.2. We focus on a forecast horizon of one year ($h = 4$). Figures 2.2-2.4 report the RMSFE for individual models of the constant mean specification, for the models with inflation in differences (SW approach) and for models where inflation has been detrended by long-run survey expectations, respectively. The results for the remaining detrending approaches are analysed in the subsequent section. The red lines correspond to the RMSFE of the average forecast over all PC models (associated with the respective detrending method) and the

with and without one of the three supply shock indicators. Of the 12 univariate models 6 are ADL and 6 are VAR specifications. The 6 models for each specification come from variation in lag and estimation window selection.

[10] Results for other inflation measures, i.e., headline inflation and the GDP deflator, are qualitatively similar to the findings we get for HEX inflation and are available from the authors upon request.

green line marks the RMSFE of the random walk benchmark. The horizontal axis indicates the various individual models considered.[11]

The first result that stands out from Figure 2.2 is the large variability in performance (in terms of RMSFE) of the individual specifications. Many specifications outperform the RW benchmark, however, this set is not constant over time. There are also many specifications with very poor accuracy. While the simple average of all demeaned PC models clearly beats the benchmark in the last two evaluation samples, it yields a remarkably higher RMSFE over the 1993-1999 sample. Thus, it seems that assuming a constant inflation mean yields particularly bad results for this period, where inflation clearly trends downward (see Figure 2.1).

Analysing our model forecasts with inflation in differences (Figure 2.3) we again find large variability both among models and over time. While this set of models can beat the benchmark over all evaluation samples as indicated by the mean, it does so by a smaller degree than the constant mean class over the last two periods. It, thus, seems that forecasting inflation under the assumption of a unit root, i.e., using inflation in differences, is particularly helpful for the first evaluation sample, while after 2000 it yields on average somewhat higher RMSFE than the constant mean specification.[12]

Detrending by long-run survey expectations is (arguably) the most interesting approach. This is the case as the survey variable can be interpreted as the public's expectation of the inflation trend which signals monetary policy makers how well inflation expectations are anchored. Given that such long-term expectations are not available before 1990 for the euro area, we only discuss results based on this trend for the last two evaluation samples. This approach leads to even better results than the before-mentioned such that the benchmark clearly can be beaten by the average PC forecast. Also, almost all individual models detrended by long-run survey expectations perform better than the random walk.

[11] The first half of models is estimated with a rolling window of 40 quarters, while the second half is estimated recursively. Within each of these classes the first half consists of ADL models while the latter half are VAR models. Finally, variations within these classes are related to different predictor variables and lag selection procedures, see 2.4.2.

[12] In the Appendix we plot the same results based on the SW detrending approach without the constant included. Individual model forecasts do not seem to be affected substantially for the intermediate period, while the average forecast error for the last evaluation sample drops and model heterogeneity is remarkably more contained in the first period.

The relative performance of different detrending approaches seems to vary over time. In particular, the models imposing a unit root in inflation appear to perform better in the beginning of the sample, whereas models in levels of inflation seem to have lower RMSFE towards the end of the sample. The relative performance of VAR versus ADL models (first and third quarter of models, respectively) and rolling versus recursive schemes also varies over time. In either case, there is no consistent evidence in favour of one specification over the other (more evidence on these different specifications is provided below).

Figure 2.2: RMSFE - individual models - constant mean

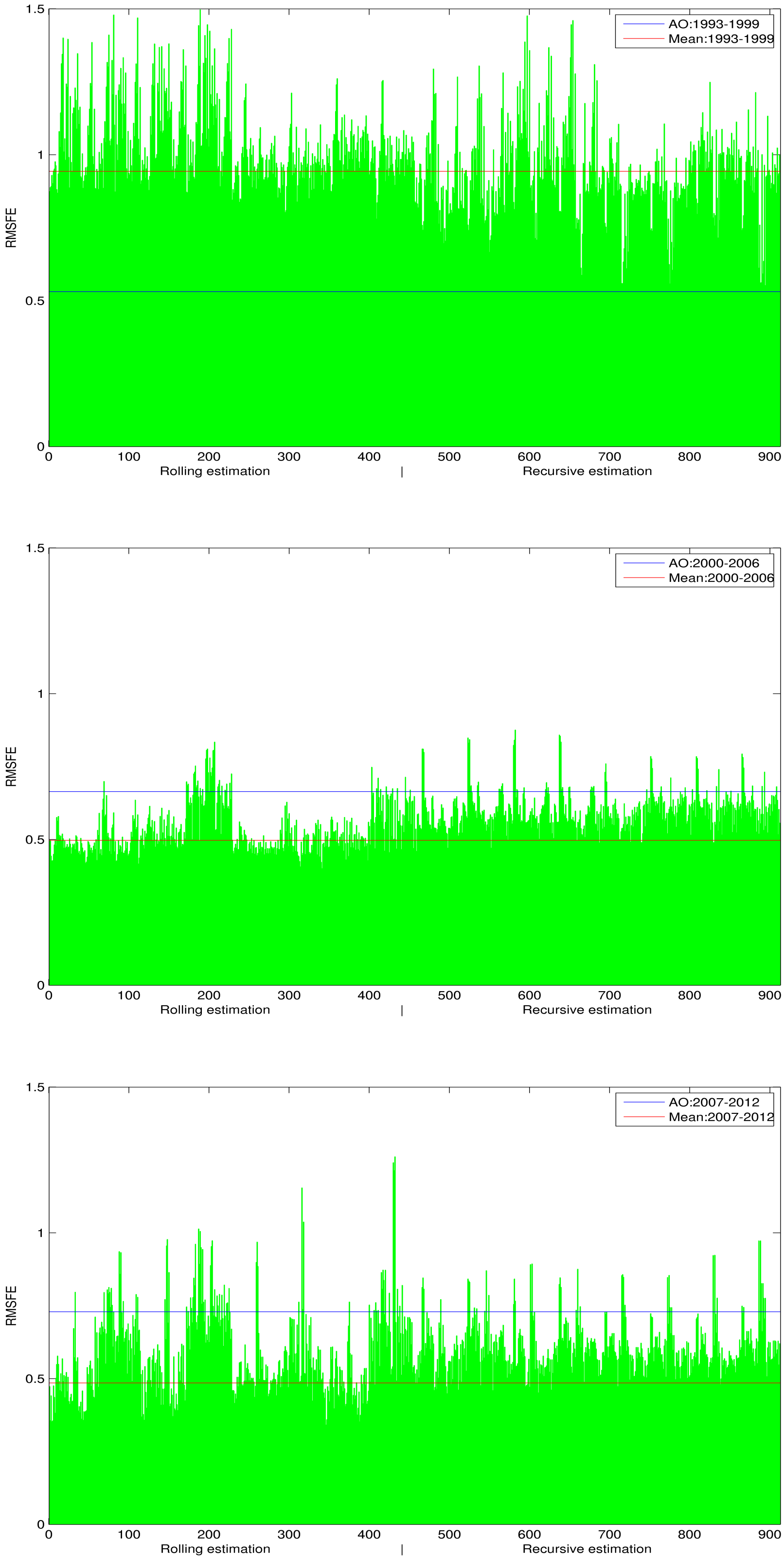


Figure 2.3: RMSFE - individual models - SW approach

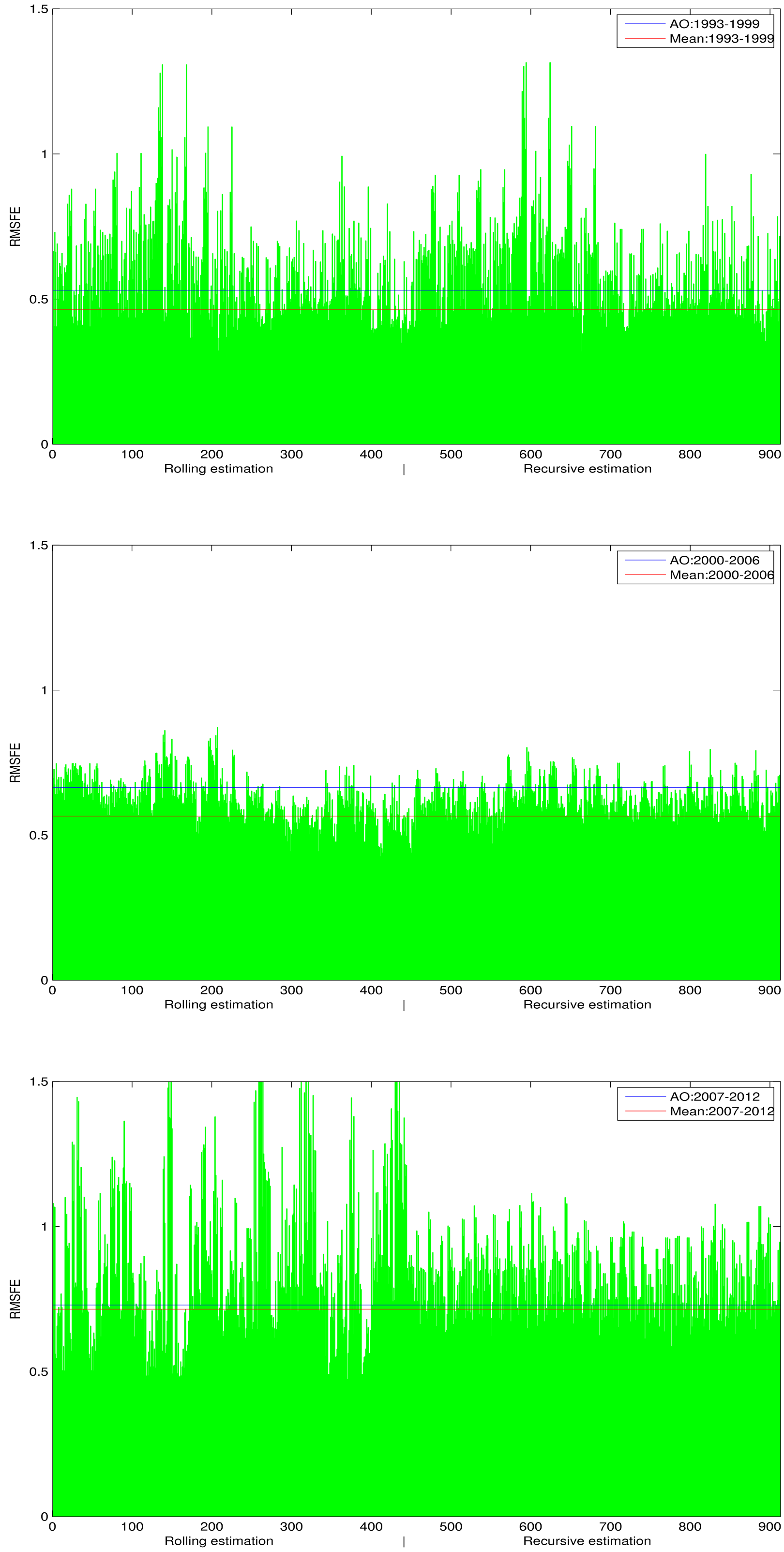
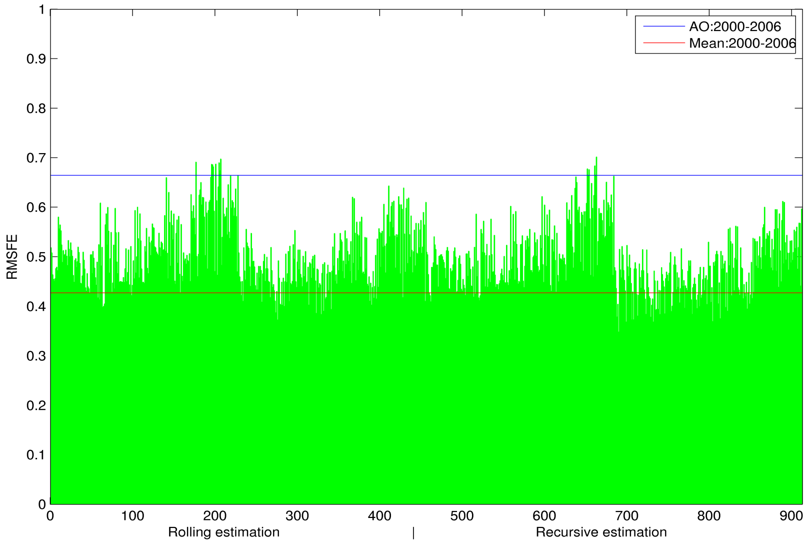
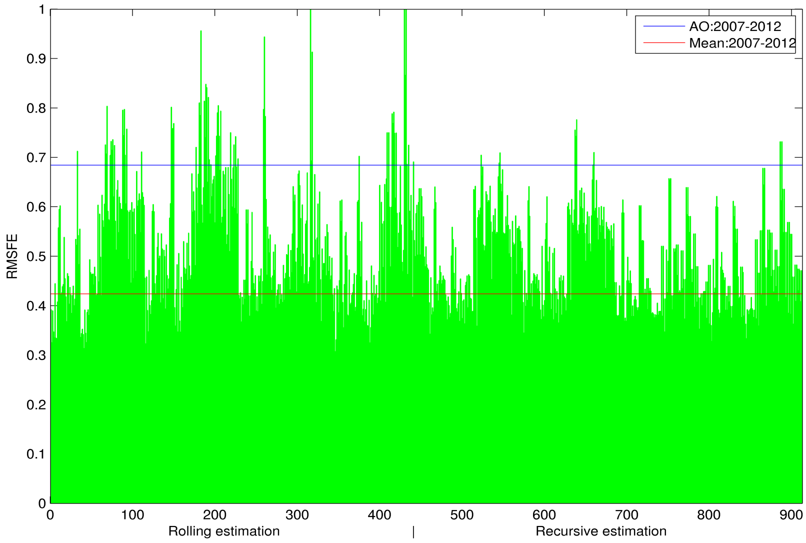


Figure 2.4: RMSFE - individual models - survey detrending



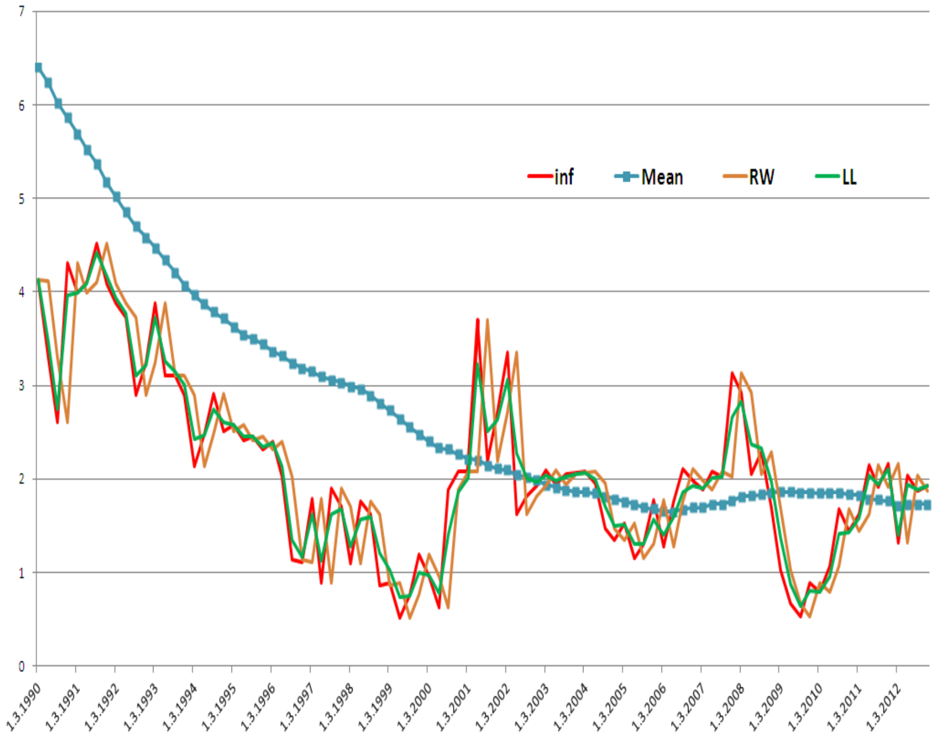


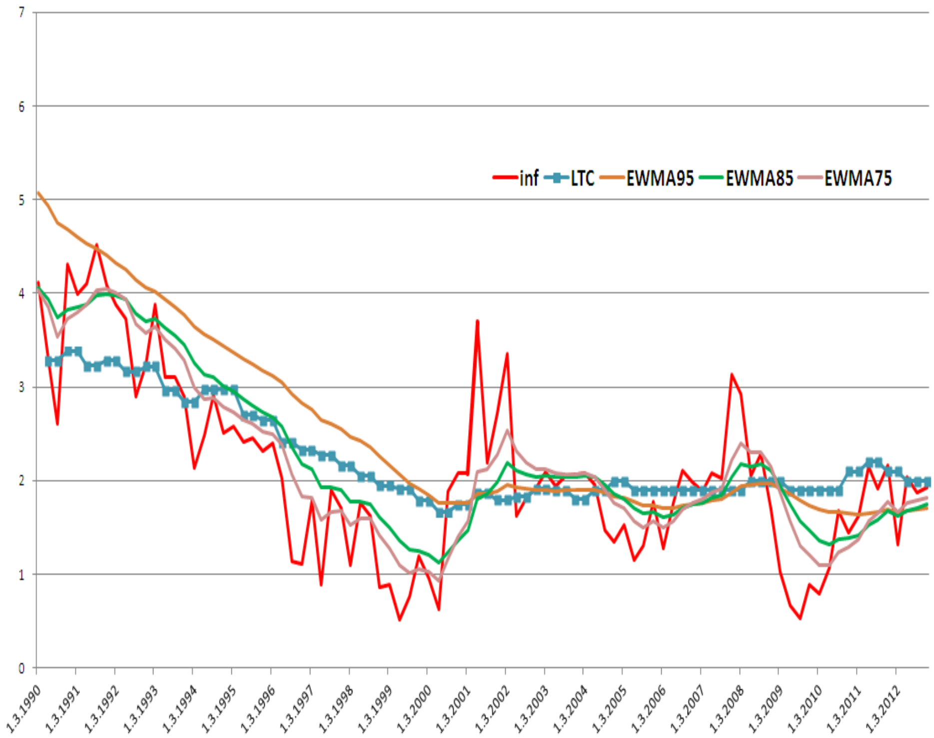
2.5.2 Different Detrending Techniques

In this section we discuss the results related to the different detrending techniques in more depth. Figure 2.5 plots the estimated trends at each point in time from 1990 - 2012. The upper graph shows the SW trend (lag of inflation) and the local level trend, which follow the inflation rate quite closely, with the SW trend (for obvious reasons) lagging behind. Further, it depicts the "constant trend", which at each point in time is just the mean inflation rate over the estimation sample up to this point. This average historical rate lies considerably above the inflation rate until 2000 after which it remains relatively constant around two percent. This offers one explanation why the models with constant inflation mean yield such a high RMSFE over the first evaluation sample: The estimated mean is significantly larger than the unobserved true trend of inflation and, thus, we forecast a series that is constantly below its trend rate.

The lower graph depicts the outcomes for the other methods to model the trend, namely the survey-based trend (the bias-corrected version is omitted in this graph), and the three different EWMA methods with a fixed coefficient on the moving average component. Until 1999 all trends move downwards as does the inflation rate, while the EWMA with the highest coefficient lies considerably above the other trends as it puts a relatively higher weight on (the higher) lags of inflation rates. As of 2000 the trend based on this latter method moves quite closely to the (relatively constant) survey-based trend, while the other two EWMA methods exhibit somewhat higher volatility.

Figure 2.5: Estimated trends





In the following we aim to answer the question whether detrending inflation is helpful for our forecasting exercises and which method yields the lowest forecast errors. In Figure 2.6 we show the average RMSFE over all PC models for the various detrending methods relative to the average RMSFE from the constant mean specification for all three evaluation samples. Figure 2.7 provides relative RMSFE against the same benchmark for the last two periods only and includes results for the survey-based methods (LRSE and LRSEC). The results for 'All' gives the relative RMSFE of the average over all models and all model categories. A value below one indicates that a given model category yields on average lower forecast errors than the constant mean approach.

The results in the first figure once more highlight the relatively bad performance (in terms of RMSFE) for the constant mean approach in the first period, where the SW models perform relatively well, while in the last period the models with inflation in differences perform worse. These differences are significant as indicated by the DM test (see Table 2.5 in the Appendix). The models based on the local level trend yield RMSFE that are similar to the SW approach and the three EWMA methods exhibit somewhat smaller RMSFE with the smallest on average coming from the approach with the highest coefficient (EWMA95).

Over the shorter sample we find that the long-run expectations trends work best in that the RMSFE for the LRSE and the LRSEC detrending methods are among the lowest. The results for EWMA95 detrending are comparable albeit leading to somewhat higher forecasting errors in the earlier subsample and slightly lower errors in the last period. The constant mean specifications yield comparable results, while the other two EWMA models result in somewhat larger errors. The SW and local level trend models are associated with the largest forecast errors, while in general the differences among detrending classes are not dramatic for the two subsamples.

Interestingly, averaging over all models comes relatively close in terms of accuracy to the best detrending approaches ex post.

Figure 2.6: Detrending - averages - long sample

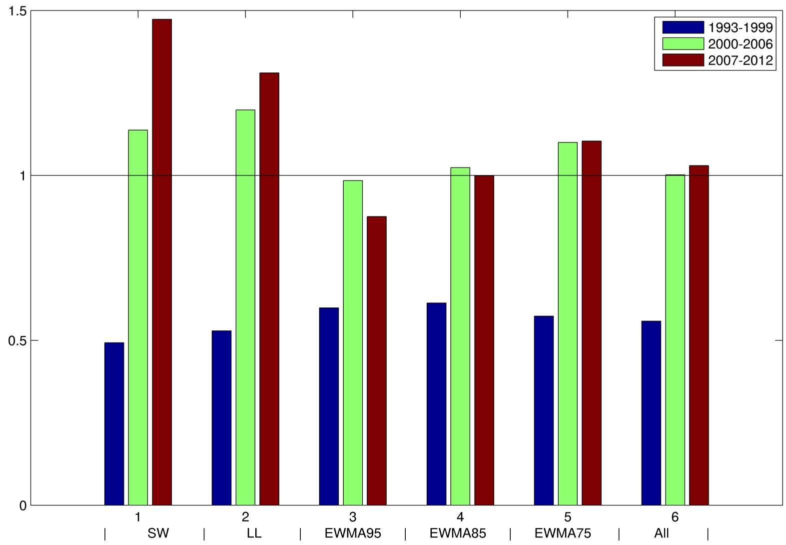
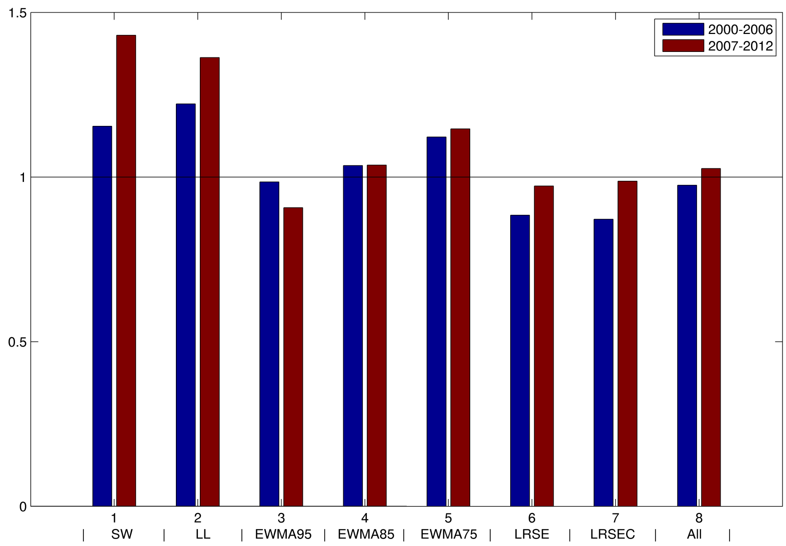


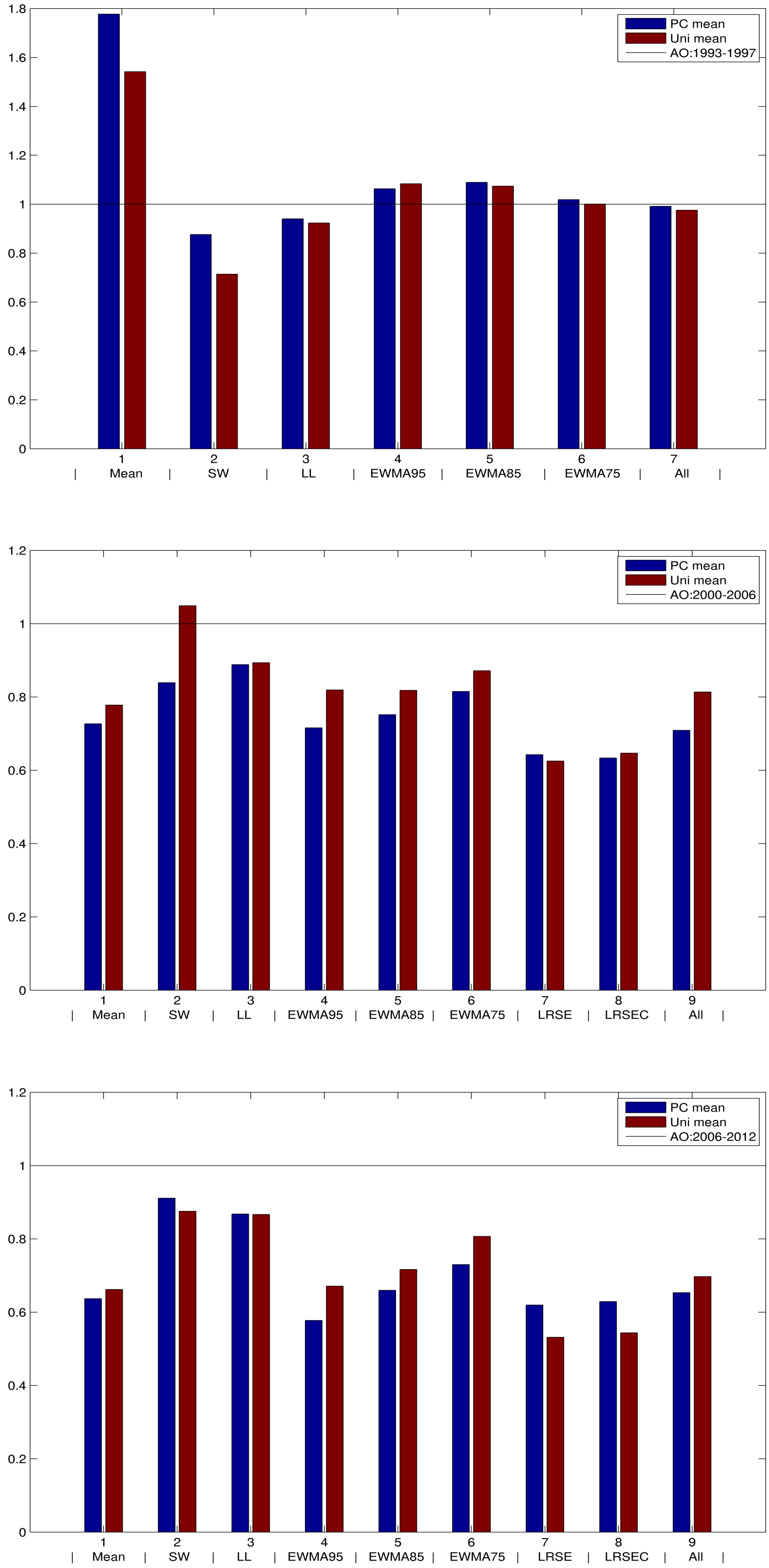
Figure 2.7: Detrending - averages - short sample



In the following we provide relative RMSFE against the random walk benchmark. We once more order the results by the averages over all individual models per detrending approach. In order to assess whether the RW model provides a good benchmark we also compare the forecasting results of our PC models to the predictions from the univariate versions of these models (see Section 2.3.3).

A few results stand out from Figure 2.8. With the exception of the models under the unit root assumption (SW) all detrending methods result in average RMSFE below the random walk benchmark for the two subsamples after 1999. Apart from the results for the local level category these differences are significant (see Table 2.6 in the Appendix). Even for the first period (1993-1999) only the constant mean models result in forecast errors that are on average significantly above the benchmark. For all periods and for each of the detrending methods it seems that the average RMSFE over all PC models is typically not very different from the average over all univariate models. While the univariate average forecast error over all models and model classes ('All') is slightly lower in the first period, the corresponding PC average yields lower errors after 2000. The success of PC models is episodic and gains with respect to univariate models are typically not large. For the last two periods the survey-based models exhibit the lowest errors with the univariate averages being even somewhat lower than those from the PC models. In this case, however, the models are not truly univariate as in the estimation we make use of long-run survey expectations.

Figure 2.8: Detrending - PC vs univariate



2.5.3 Model Averaging

So far we have discussed only individual model results or the simple average over a certain group of models. We now aim to answer the question whether the simple average is a good model combination approach and which alternative – if any – leads to lower forecast errors. In Tables 2.1-2.3 we present relative RMSFE for the various averaging techniques discussed in Section 2.3.4, i.e., we divide forecast errors by the errors from the simple average benchmark. Thus, a value below one (above one) indicates that a particular averaging technique beats (performs worse than) the benchmark. We present results, where we average over the models subject to a constant mean, the SW models and over all models from all detrending approaches, respectively. We test for equal mean squared errors of a particular set of models relative to the benchmark by means of the Diebold-Mariano (1995) statistic.

The bad performance of the constant mean specifications for the first subsample discussed before provides a rationale for the finding that most averaging approaches result in significantly lower RMSFE than the benchmark 'mean' over the 1993-1999 period. The performance-based averages are able to put lower weights on the models that yield particularly high errors, while the benchmark method assigns all models the same weight. Even over the whole sample (1993-2012) this result is not reversed. For the last two periods significant improvements over the simple average are rare, exceptions being the trimmed means and the recursively-weighted averages without discounting past errors. The approaches using the recent best models and the information-theoretic averages yield particularly high errors in these cases. For the constant mean models it, thus, seems that no combination approach dominates the simple average.

The results for the models with inflation in differences (SW approach) are similar. In this case it is even more apparent that performance-based averaging does not offer improvements over the mean. Exceptions are the trimmed mean (90) approach in the first evaluation period and the AIC-based combination approach in the second sample yielding significantly lower forecast errors at the 10% level. The approaches using the recent best models again are associated with the highest errors. Once more no averaging approach dominates the simple mean and in case a method, such as, e.g., the information-theoretic approaches, provides better results for a particular period, differences are not large.

Finally, we provide averaging results for all models and all detrending approaches (5472 models over the long sample, i.e., excluding survey-based detrending).[13] Over the first and the last evaluation samples many of the performance-based combination methods (most recursively-weighted approaches and the trimmed means) perform significantly better than the benchmark. However, most of these same approaches yield significantly higher mean squared forecast errors in the intermediate period. Once again the simple mean provides a forecast that is hardly improved upon by any other averaging approach.

In summary, it seems that simple averaging is an adequate solution strategy for model heterogeneity and instability that is hard to beat by any other combination method. Further, averaging over all detrending methods makes the results robust to the exact specification for the unobserved trend of inflation.

[13]Note that for the last two periods the models where inflation is detrended by a survey variable are excluded for simplicity.

Table 2.1: Model averaging - constant mean

	Time period (in quarters)			
	1993:1-2012:4	1993:1-1999:4	2000:1-2006:4	2007:1-2012:4
S&W averaging				
Mean	1.00	1.00	1.00	1.00
Median	0.99	0.97**	1.01	1.04
RecBest4	0.98	0.82***	1.17	1.32**
RecBest8	0.93	0.76***	1.09	1.31***
RecBest16	0.95	0.72***	1.16	1.43*
RecBest24	0.94	0.76***	1.14	1.33
RWAND1	0.98***	0.97***	1.01	0.98**
RWAND2	0.95***	0.93***	1.02	0.96**
RWAD51	0.99	0.95***	1.04	1.08
RWAD52	0.96	0.88***	1.08	1.10
RWAD71	0.99	0.96***	1.03	1.07*
RWAD72	0.97	0.90***	1.06	1.10*
RWAD91	0.98***	0.97***	1.01	1.00
RWAD92	0.96***	0.92***	1.02	1.01
RWAD951	0.98***	0.97***	1.01	0.98
RWAD952	0.95***	0.93***	1.02	0.97
Trim90	1.00	1.00	1.01	0.98**
Trim50	0.97***	0.97***	1.03	0.92*
Information-theoretic comb				
AIC	1.01	0.97***	1.08**	1.07*
BIC	1.01	0.97***	1.08**	1.07*

The table presents relative RMSFE of the different averaging approaches discussed in Section 2.3.4 for the constant mean specification and against the simple average benchmark. A value above one shows that the averaging technique performs worse than the benchmark for a certain sample. ***, **, and * denote significantly different squared forecast errors at the 1, 5 and 10 percent level, respectively, as indicated by the Diebold-Mariano (1995) test. Note that for the performanced-based averages beginning with RW the sample starts only in 1994. The according abbreviations stand for: RWA=Recursively-weighted averages, (N)D=(Non)-discounted. A 1 in the end signals normal variances and a 2 squared variances (see the two λ in Section 2.3.4). The numbers before (5,7,9,95) stand for weights 0.5, 0.7, 0.9 and 0.95, respectively.

Table 2.2: Model averaging - SW approach

	Time period (in quarters)			
	1993:1-2012:4	1993:1-1999:4	2000:1-2006:4	2007:1-2012:4
S&W averaging				
Mean	1.00	1.00	1.00	1.00
Median	0.99	0.97	1.01	0.98
RecBest4	1.31**	1.30**	1.22**	1.38
RecBest8	1.22*	1.44**	1.14*	1.16
RecBest16	1.20***	1.46**	1.20**	1.06
RecBest24	1.08	1.42**	0.98	0.95
RWAND1	1.01	1.03	0.99	1.01
RWAND2	1.02**	1.10**	0.99	1.02
RWAD51	1.00	1.02	1.00	0.98
RWAD52	1.00	1.08	1.02	0.97
RWAD71	1.00	1.03	0.99	0.99
RWAD72	1.00	1.09*	0.99	0.98
RWAD91	1.00	1.03	1.00	1.00
RWAD92	1.01	1.10**	0.99	1.00
RWAD951	1.01	1.03	1.00	1.00
RWAD952	1.02	1.10**	0.99	1.01
Trim90	1.00	0.99*	1.00	1.01
Trim50	1.01	1.00	1.00	1.02
Information-theoretic comb				
AIC	0.99	1.00	0.96*	1.01
BIC	0.99	1.00	0.97	1.00

The table presents relative RMSFE of the different averaging approaches discussed in Section 2.3.4 for the SW class, i.e., with inflation in differences and against the simple average benchmark. A value above one shows that the averaging technique performs worse than the benchmark for a certain sample. ***, **, and * denote significantly different squared forecast errors at the 1, 5 and 10 percent level, respectively, as indicated by the Diebold-Mariano (1995) test. Note that for the performanced-based averages beginning with RW the sample starts only in 1994. The according abbreviations stand for: RWA=Recursively-weighted averages, (N)D=(Non)-discounted. A 1 in the end signals normal variances and a 2 squared variances (see the two λ in Section 2.3.4). The numbers before (5,7,9,95) stand for the weights 0.5, 0.7, 0.9 and 0.95, respectively.

Table 2.3: Model averaging - all models

	Time period (in quarters)			
	1993:1-2012:4	1993:1-1999:4	2000:1-2006:4	2007:1-2012:4
<u>S&W averaging</u>				
Mean	1.00	1.00	1.00	1.00
Median	1.02*	1.01	1.02	1.03
RecBest4	1.28***	1.08	1.50***	1.23
RecBest8	1.28***	1.08	1.33**	1.44**
RecBest16	1.20**	1.07	1.39**	1.11
RecBest24	1.16	1.03	1.12	1.34
RWAND1	0.98*	0.91***	1.03	0.98***
RWAND2	0.98	0.91***	1.04	0.97***
RWAD51	1.00	0.90***	1.08**	1.03
RWAD52	1.07*	0.93	1.18***	1.08
RWAD71	0.99	0.90***	1.07**	1.01
RWAD72	1.02	0.91**	1.12**	1.03
RWAD91	0.98	0.91***	1.04*	0.99**
RWAD92	0.98	0.91***	1.07*	0.97**
RWAD951	0.98*	0.91***	1.04*	0.98***
RWAD952	0.98	0.91***	1.06*	0.97***
Trim90	0.99*	0.96***	1.02	0.99*
Trim50	0.98	0.92***	1.06*	0.94***
<u>Information-theoretic comb</u>				
AIC	1.03*	1.11***	0.97	1.00
BIC	1.04*	1.12***	0.96	1.00

The table presents relative RMSFE of the different averaging approaches discussed in Section 2.3.4 for all models and trend specifications and against the simple average benchmark. A value above one shows that the averaging technique performs worse than the benchmark for a certain sample. ***, **, and * denote significantly different squared forecast errors at the 1, 5 and 10 percent level, respectively, as indicated by the Diebold-Mariano (1995) test. Note that for the performanced-based averages beginning with RW the sample starts only in 1994. The according abbreviations stand for: RWA=Recursively-weighted averages, (N)D=(Non)-discounted. A 1 in the end signals normal variances and a 2 squared variances (see the two λ in Section 2.3.4). The numbers before (5,7,9,95) stand for weights 0.5, 0.7, 0.9 and 0.95, respectively.

2.5.4 Different Specifications

In this section, we evaluate the performance of different groups of models that are ordered by various aspects discussed before. We analyse whether ADL or VAR models perform better, how well the rolling and recursive estimation schemes work relative to each other and which lag selection procedure yields the best results for the average RMSFE over all detrending methods. Results are shown in Figure 2.9.

Not surprisingly and in line with the literature (see Stock and Watson, 2009; Faust and Wright, 2012) these different aspects of the models and estimation techniques do not matter much, i.e., differences never exceed 10 percent. ADL and VAR models yield almost the same RMSFE for the average over all detrending methods in the first period, while VAR models perform better towards the end of the sample. Differences are not significant (see Table 2.7 in the Appendix).

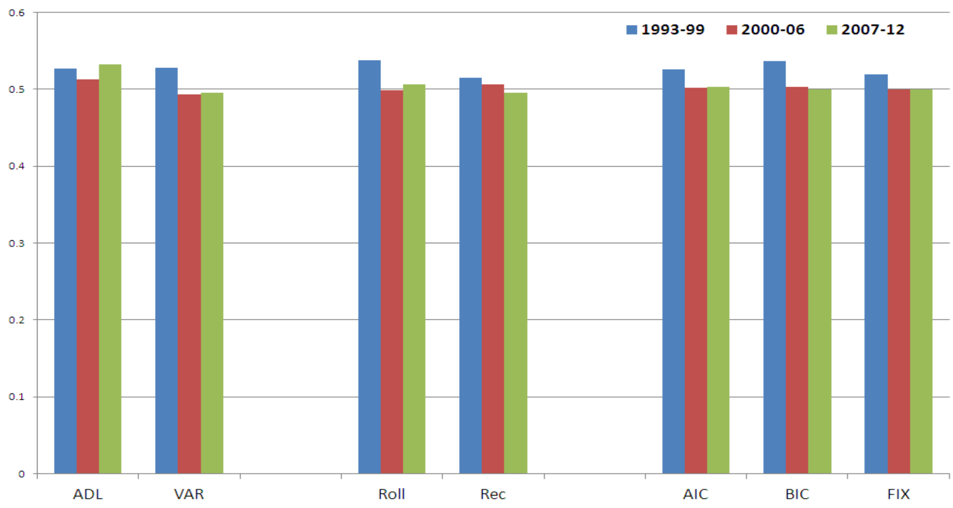
Similarly, using either a rolling or a recursive estimation window yields comparable results. Although for the first subsample RMSFE are significantly lower for the recursive approach differences are very small and become insignificant for the following two periods (see Table 2.7 in the Appendix).[14]

Finally, the lag selection approaches, namely using the AIC or the BIC criterion or enforcing a lag length of four, lead to roughly the same forecast errors. Only in the first sample the fixed lag length approach performs slightly better, though, not significantly so (see Table 2.7 in the Appendix).

The fact that PC models seem to outperform the benchmark random walk model, as well as other univariate models during some episodes, raises the question which predictor variables are the most important ones. We thus analyse which marginal cost measures result in the best model predictions and if the inclusion of supply shocks can improve these predictions. Figures 2.10 and 2.11 show the average RMSFE

[14] Generally, results may depend on the particular choice of window size. As suggested by Pesaran and Timmermann (2007), we also estimate all our PC models over different window sizes (30, 40 and 50 quarters) and the recursive window and provide the average over all these approaches. Results are reported in Table 2.8 in the Appendix; they indicate that our previous findings are relatively robust to the choice of window size: The different averaging approaches (apart from the recent best model class) yield significantly lower RMSFE than the benchmark AO model after 2000 and over the whole sample (see the results for 'All' PC models in Figure 2.8 for comparison). In the first period some averaging approaches can beat the benchmark AO model, while differences are not significant. In general, all averaging approaches lead to very similar relative RMSFE (apart from the recent best approaches).

Figure 2.9: RMSFE - ADL vs VAR, estimation window, lag selection



associated with all the models (and thus all detrending techniques) that include either of the potential marginal cost measures or supply shocks, respectively.

One remarkable result is that for each period there is at least one marginal cost measure associated with models that yield on average lower RMSFE than the univariate models (named 'None'). Also the average prediction over all PC models is as low as the univariate prediction over the first period and lower for the other two periods. The best predictor variable in terms of average RMSFE is either associated with output or with the unemployment rate. While the principal component models yield RMSFE that are around as large as the average over all PC models, the predictions associated with the unemployment recession gap (URXrec) have considerably larger RMSFE even for the last period which includes the financial crisis and the following recession.

The supply shock variable resulting in the best model predictions varies over time. While the imports of goods and services deflator (MTD) performs best in the first period, in the following period the exchange rate (EEN) leads to better results and in the final subsample the oil price (POE) performs best. On average an inclusion of a supply shock improves on the no supply shock specifications ('None') only in the intermediate period. In the first and the last period the models including the best performing supply shock variable yield forecast errors that are about as high as for the models without supply shock.

Figure 2.10: RMSFE - real marginal cost measures

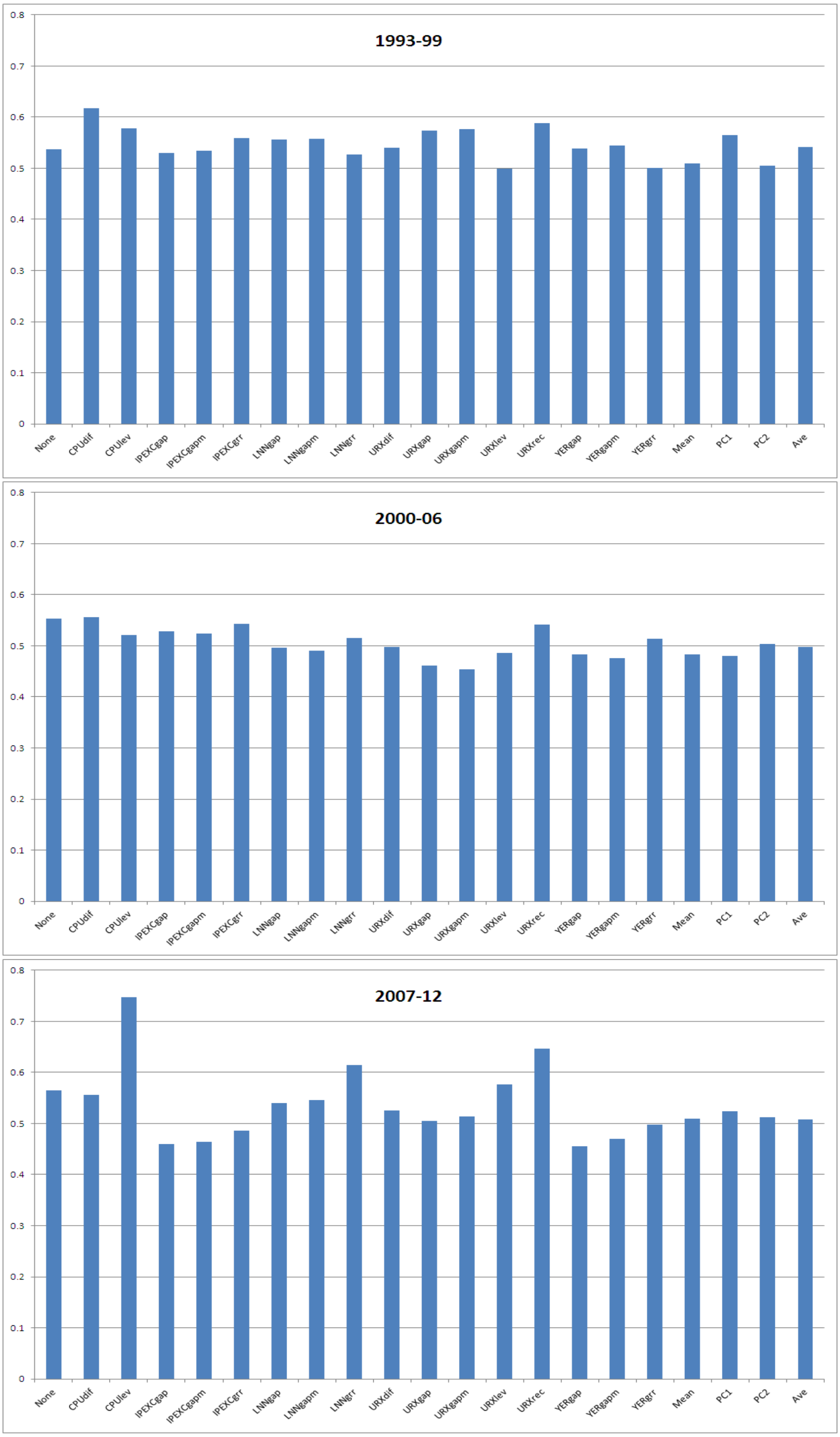
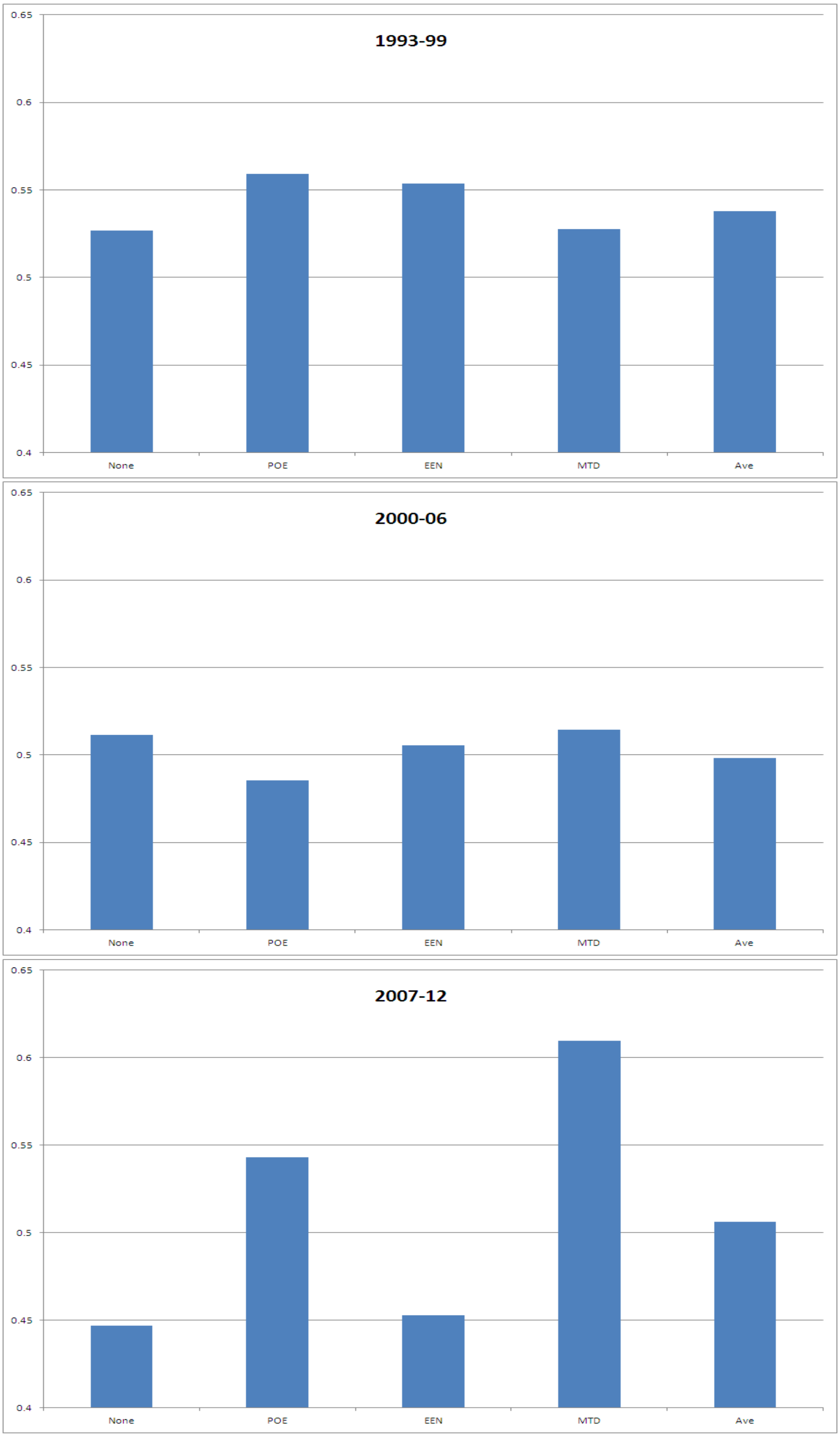


Figure 2.11: RMSFE - supply shocks



2.6 Conclusion

In this paper we evaluate the forecasting performance of a wide range of Phillips curve models and their combinations for euro area inflation over the period 1993-2012.

We find a significant degree of uncertainty around the best model specification - the range of the resulting point forecasts is very wide and the average individual model forecast error very heterogeneous. Further the relative performance of different models varies over time. In particular, the specifications with inflation in differences perform better in the first part of the sample, while they yield very high forecast errors towards the end of the sample. By contrast, the performance of ADL versus VAR specifications, different lag selection criteria and estimation windows appear in most cases comparable.

We compare different detrending techniques and demonstrate that detrending inflation by long-run survey inflation expectations from Consensus Economics prior to estimation yields the lowest forecast errors for most individual models, as well as on average. The simple average from such specifications outperforms the random walk benchmark. Thus, long-run survey inflation expectations seem a useful way to capture a time-varying mean in this context. The performance of the exponentially-weighted moving average trend with a low "forgetting" factor is about comparable and thus provides the best model-based alternative to detrending by survey expectations.

In line with results reported elsewhere, choosing the best model based on the past forecasting performance (almost) never improves upon some version of forecast combination. Regarding the latter, simple equally-weighted averaging appears to be an effective remedy against model uncertainty. Forecast combinations based on past performance and in particular the more sophisticated information-theoretic averaging offer improvements over simple averaging only in some cases. These findings underscore the usefulness of considering the average of a large set of candidate model predictions rather than relying on a single model.

Regarding the comparison with univariate benchmarks, averages of Phillips curve model forecasts typically improve upon the benchmark random walk model while it seems that other univariate model forecasts (combinations) are harder to beat and improvements (if any) are typically not large.

Regarding the predictor variables, it stands out that the unemployment rate or output growth (or their respective gaps) are often part of the best model. The inclusion of a supply shock only infrequently improves on the results and the performance of individual supply shocks is relatively volatile over time.

The focus of this work is on point forecasts from linear models with fixed coefficients estimated in the frequentist domain. Extending the analysis to consider time-varying parameter and/or non-linear models, Bayesian estimation methods and density forecasts is an interesting avenue for future research.

A2 Appendix to Chapter 2

A2.1 Data

This appendix describes the data that we use in the exercises, along with the respective sources. The data is quarterly and available for the period 1980:2011 for most series.

Data for HICP account for the changing composition of the euro area. Regarding back data for HICP, data prior to 1996 is estimated on the basis of the non-harmonised national consumer price indices. Data prior to 1991 exclude East Germany and country weights are calculated on the basis of PPP conversion rates before 1990. The back data has been seasonally adjusted using X12ARIMA.

The survey inflation expectations come from Consensus Economics. The aggregate series for the euro area is available as of 2003 and from 1990 to 2003 it is constructed on the basis of the forecasts for the largest euro area countries (see Castelnuovo, Nicoletti-Altimari, and Rodríguez-Palenzuela, 2003, for details).

The following abbreviations are used for the sources: ESA=ECB - ESA95 National Accounts, ICP=ECB - Indices of Consumer Prices, STS=ECB - Short-Term Statistics, FM=Bloomberg - Financial Market Data, EXR=ECB - Exchange Rates, MEI= OECD - Main Economic Indicators, SUR= EU Commission - Opinion Surveys, CONS=Consensus Economics.

Table 2.4: Data

Name	Description	Source
	Inflation series	
HICP	Overall HICP	ICP
YED	GDP deflator	ESA
HEX	HICP Excluding Energy	ICP
CLR	Consensus long-run inflation expectations	CONS
	Marginal cost measures	
URX	Unemployment rate (% of labour force)	STS
YER	Real GDP	ESA
LNN	Total employment (persons)	ESA
CPU	Capacity utilization	SUR
IPT	Industrial production index (total industry)	STS
	Supply shock variables	
POE	Oil price (UK Brent Crude Index in USD)	FM
EEN	Nominal effective exchange rate (EER12)	EXR
MTD	Imports of goods and services deflator	ESA
	Additional variables for principal component	
ITR	Gross Investment	ESA
XTR	Exports of Goods and Services (Real)	ESA
MTR	Imports of Goods and Services (Real)	ESA
IPEXC	EMU Production Index of Total Industry	MEI
RETSSTS	Total Turnover Index, Retail Trade excl. Fuel	STS

A2.2 Principal Components

The idea to use principal components generated from large macroeconomic data sets in (inflation) forecasting comes from Stock and Watson (1999). Principal component analysis relies on the assumption that a set of variables X_t is driven by a small number of factors and some idiosyncratic shocks which allows for the following representation:

$$X_t = \Lambda F_t + \nu_t, \tag{2.4}$$

where X_t is an $N \times 1$ vector of zero-mean, $I(0)$ variables, Λ is an $N \times k$ matrix of factor loadings, F_t is an $k \times 1$ vector of the factors and ν_t is an $N \times 1$ vector of idiosyncratic shocks, where N, the number of variables, is much larger than the number of factors k. Static factors can be estimated by minimizing the following objective function:

$$V_{N,T}(F, \Lambda) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \Lambda_i' F_t)^2, \tag{2.5}$$

where $F = (F_1, F_2, \ldots, F_T)'$, Λ_i' is the i-th row of Λ, X_{it} is the i-th component of X_t and T is the number of time periods.

We generate one principal component each from two different sets of variables. First, we use the variables that we consider as the standard marginal cost measures (see Section 4.2) plus the unemployment recession gap. Thus, the resulting principal component can be interpreted as a summary of the potential marginal cost measures. Second, we employ a set of variables that focus more on real activity: YER, PCR, URX, LNN, ITR, MTR, XTR, IPEXC, RETSSTS. Explanations for these variables can be found in Table 2.4.

A2.3 Other Results

Figure 2.12: RMSFE - individual models - SW approach - no constant

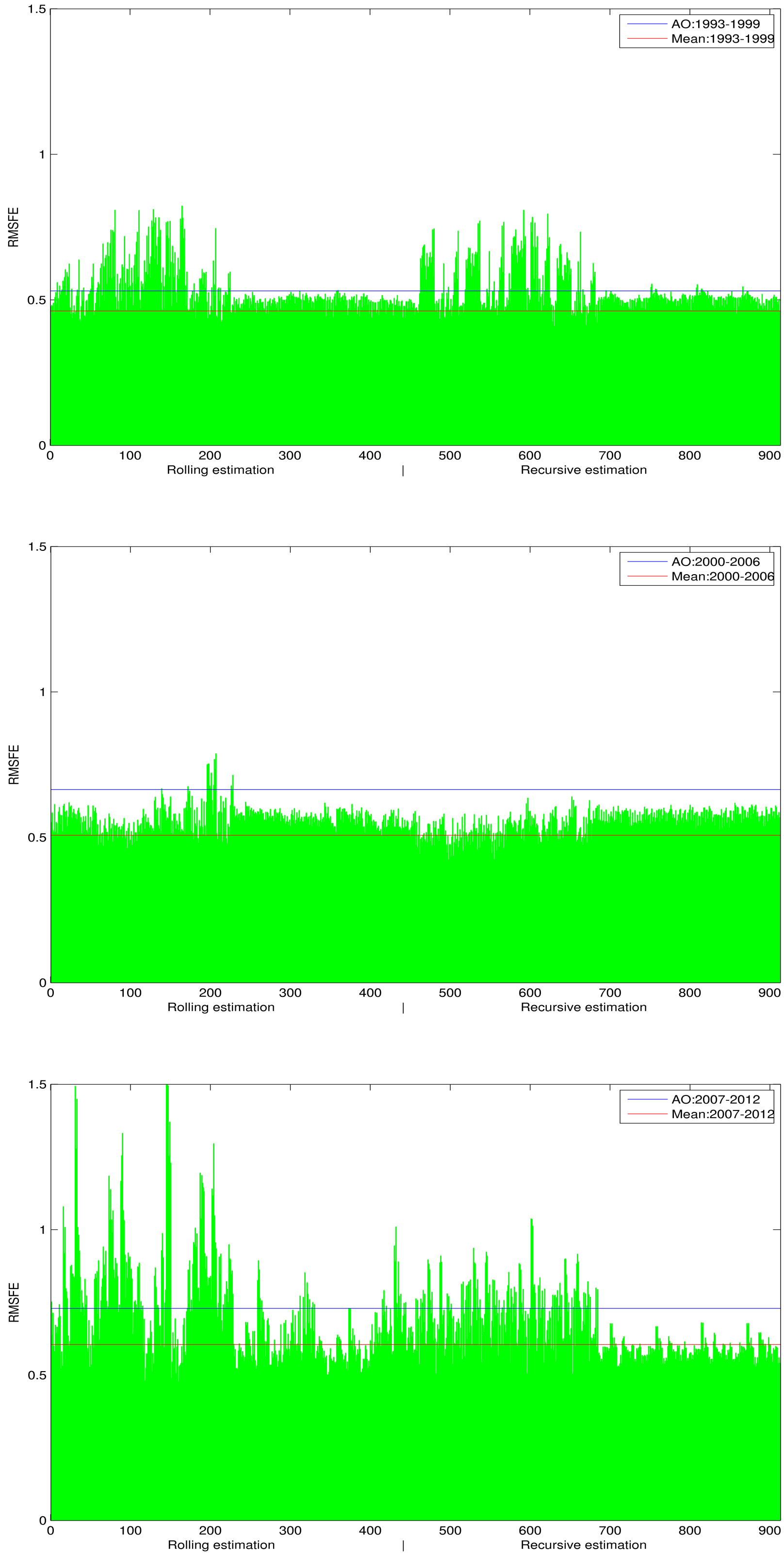


Table 2.5: DM test results for Figures 2.6 and 2.7

	Time period (in quarters)		
	1993:1-1999:4	2000:1-2006:4	2007:1-2012:4
SW	4.82***	-0.75	-1.97*
LL	8.28***	-1.36	-4.64***
EWMA95	7.09***	0.16	1.72*
EWMA85	7.76***	-0.18	0.01
EWMA75	8.64***	-0.65	-1.37
LRSE	-	1.93*	0.67
LRSEC	-	1.47	0.26
All	8.39***	0.29	-0.93

The table presents the Diebold-Mariano (1995) test statistic associated with the averages of the different detrending model classes discussed in Section 2.3.2 against the constant mean benchmark. The class 'All' includes the results for all model classes for the last two periods, while in the first period the survey-based methods are not included. ***, **, and * denote significantly different squared forecast errors at the 1, 5 and 10 percent level.

Table 2.6: DM test results for Figure 2.8

	Time period (in quarters)		
	1993:1-1999:4	2000:1-2006:4	2007:1-2012:4
PC models			
Mean	-8.02***	1.86*	3.30***
SW	0.83	2.19**	0.12
LL	0.90	1.24	1.87*
EWMA95	-0.87	2.96***	3.54***
EWMA85	-1.56	3.29***	3.37***
EWMA75	-0.46	3.61***	3.17***
LRSE	-	2.85***	3.83***
LRSEC	-	3.10***	3.64***
All	0.15	3.46***	3.41***
Univariate models			
Mean	-6.73***	2.23**	3.30***
SW	2.26**	-1.08	1.34
LL	1.22	1.34	2.00*
EWMA95	-1.49	2.45**	3.14***
EWMA85	-1.59	3.42***	3.07***
EWMA75	-0.03	3.45***	2.86***
LRSE	-	3.17***	3.79***
LRSEC	-	3.46***	3.60***
All	0.50	3.09***	3.31***

The table presents the Diebold-Mariano (1995) test statistic associated with the averages of the different detrending model classes discussed in Section 2.3.2 against the random walk benchmark for both PC and uniavriate model. The class 'All' includes the results for all model classes for the last two periods, while in the first period the survey-based methods are not included. ***, **, and * denote significantly different squared forecast errors at the 1, 5 and 10 percent level.

Table 2.7: DM test results for Figures 2.9

	Time period (in quarters)		
	1993:1-1999:4	2000:1-2006:4	2007:1-2012:4
ADL versus VAR			
ADL	-	-	-
VAR	-0.10	0.93	1.13
Estimation Window			
Roll	-	-	-
Rec	2.65**	-0.61	0.96
Lag Length Selection			
AIC	-	-	-
BIC	-1.53	-0.17	0.40
FIX	0.60	0.42	0.41

The table presents the Diebold-Mariano (1995) test statistic associated with the averages of the different specification shown in Figure 2.9. The benchmark for the three cases are the ADL average, the rolling window average and the AIC-based models, respectively. ***, **, and * denote significantly different squared forecast errors at the 1, 5 and 10 percent level.

Table 2.8: Model averaging - all models - different rolling windows

	Time period (in quarters)			
	1993:1-2012:4	1993:1-1999:4	2000:1-2006:4	2007:1-2012:4
S&W averaging				
Mean	0.78***	1.00	0.74***	0.65***
Median	0.80***	1.01	0.76***	0.68***
RecBest4	0.96	1.09	0.98	0.84
RecBest8	0.92	0.92	0.97	0.88
RecBest16	0.89	1.00	0.97	0.72
RecBest24	0.90	0.97	1.06	0.65**
RWAND1	0.77***	0.98	0.76***	0.64***
RWAND2	0.77***	0.97	0.78***	0.64***
RWAD51	0.78***	0.95	0.80***	0.66***
RWAD52	0.82***	0.94	0.85**	0.71***
RWAD71	0.78***	0.96	0.79***	0.65***
RWAD72	0.79***	0.94	0.82***	0.66***
RWAD91	0.77***	0.97	0.77***	0.64***
RWAD92	0.77***	0.95	0.79***	0.63***
RWAD951	0.77***	0.98	0.77***	0.64***
RWAD952	0.77***	0.96	0.78***	0.63***
Trim90	0.78***	0.98	0.75***	0.64***
Trim50	0.77***	0.92	0.79***	0.63***
Information-theoretic comb				
AIC	0.79***	1.07	0.72***	0.65***
BIC	0.79***	1.07	0.72***	0.65***

The table presents relative RMSFE of the different averaging approaches discussed in Section 2.3.4 for all models and trend specifications and against the random walk benchmark. Further, averages are constructed also over different rolling windows of 30, 40 and 50 quarters and the recursive approach. A value above one shows that the averaging technique performs worse than the benchmark for a certain sample. ***, **, and * denote significantly different squared forecast errors at the 1, 5 and 10 percent level, respectively, as indicated by the Diebold-Mariano (1995) test. Note that for the performanced-based averages beginning with RW the sample starts only in 1994. The according abbreviations stand for: RWA=Recursively-weighted averages, (N)D=(Non)-discounted. A 1 in the end signals normal variances and a 2 squared variances (see the two λ in Section 2.3.4). The numbers before (5,7,9,95) stand for weights 0.5, 0.7, 0.9 and 0.95, respectively.

Chapter 3

Inflation Expectation Dynamics: The Role of Past, Present and Forward-Looking Information

3.1 Introduction

Private expectations regarding future economic and policy developments influence current decisions about wages, savings and investments, and concurrently, economic policy decisions. In recent years there has been an increasing interest in explaining the private inflation expectations formation process by departing from the full information rational expectations hypothesis.

Within this literature, Mankiw and Reis (2002) propose a sticky-information model where private agents may form rational expectations, but only update their information set each period with a certain probability as they face costs of absorbing and processing information. Sims (2003) as well as Mackowiak and Wiederholt (2009) focus on partial and noisy information models: the observed inertial reaction of private agents arises from an inability to pay attention to all the noisy information available although people update continuously. It is an optimal choice for private agents - internalizing their information processing capacity constraints - to remain inattentive to some part of the available information because incorporating all signals is impossible (Moscarini, 2004). In both types of models, a fraction of the information set used by private agents is backward-looking, i.e. based on past information. Carroll (2003), Mankiw, Reis, and Wolfers (2003), Pesaran and Weale (2006), Branch (2007), Nunes (2009), Andrade and Le Bihan (2010), Coibion (2010b) and Coibion and Gorodnichenko (2010, 2012) provide empirical evidence based on survey data to characterize and distinguish these types of models.

Another strand of literature has focused on inflation dynamics and the role of private expectations estimating New Keynesian Phillips Curves (NKPC). Roberts (1995, 1997), Galí and Gertler (1999), Rudd and Whelan (2005), Nunes (2010) and Adam and Padula (2011), among others, assess the relative weights of forward- and backward-looking components of inflation. The latter may play a role due to "backward-looking" private agents, i.e. a share of firms that do not re-optimize their prices but set them according to a rule of thumb (see e.g. Steinsson, 2003) or index their prices completely to lagged inflation as in Galí and Gertler (1999) or Christiano, Eichenbaum, and Evans (2005).

By bridging these two strands of literature, this paper proposes to investigate the role of past, present and forward-looking information in inflation expectation dynamics. We aim at assessing whether and by how much private inflation expectations are driven by forward-looking information (i.e. further-ahead expectations), current information (i.e. the current output gap), or backward-looking information (i.e. past realizations of inflation).

To our knowledge, two papers have already opened this line of research. Lanne, Luoma, and Luoto (2009) find that inflation expectations are consistent with a sticky information model where a significant proportion of households base their inflation expectations on past inflation rather than the rational forward-looking forecast, while Pfajfar and Santoro (2010) show that private forecasts might be explained by three expectation formation processes: a static or highly auto-regressive region on the left hand side of the median, a nearly rational region around the median and forecasts on the right-hand side of the median formed with adaptive learning and sticky information.

Assuming that the hybrid NKPC is the true data generating process of inflation, our contribution to the literature is to propose an NKPC-based inflation expectations formation equation in order to evaluate the relative importance of past, present and forward-looking information in determining inflation expectation dynamics. Determining the respective weights of past, present and forward-looking information in inflation expectations is important because the real effects of monetary policy depend on the speed of price adjustments which in turn depend on the (in)completeness of information and/or the backward-lookingness of price expectations. Optimal monetary policy will therefore be determined by the degree of price stickiness (see e.g. Erceg, Henderson, and Levin (2000) or Steinsson (2003)) and by the expectations formation process, i.e. whether private agents use up-to-date information about the current state of the economy or continue using their previous plans and set prices based on outdated information (see e.g. Ball, Mankiw, and Reis (2005) or Reis (2009)). Policy recommendations thus depend on the degree of backward- and forward-lookingness of price setters and inflation forecasts.

We estimate our NKPC-based inflation expectations formation equation on US data, for which survey expectations from the Survey of Professional Forecasters are available on a fixed-horizon scheme and for a long time span: 1981Q3-2012Q3. We use both GDP deflator and CPI to measure inflation as well as various variables for marginal costs including a constructed measure of the output gap. In addition to our main question of interest, we also assess whether relative weights vary for different forecasting horizons and if expectations of consumers differ from those of professional forecasters.

Our results are threefold. First, professional forecasters put relatively more weight on forward-looking information, while past information is significant and the contribution of the marginal cost measure is small and often insignificantly different from zero. This result is found to be robust to the use of real-time data, to various measures of marginal costs, to the use of the mean of individual responses, to another estimation procedure namely GMM, and to the inclusion of potentially relevant additional variables. Second, the coefficients are similar to those found in the literature estimating the actual NKPC which suggests that professional forecasters indeed use this model to form their own inflation expectations. Consumers seem to differ from professionals in that their inflation forecasts do not follow the NKPC-based formation process. Third, we also find that the estimated parameters of this NKPC-based expectations formation model are relatively stable when the forecasting horizon varies or when we consider further-ahead horizons for forward-looking information.

While it might appear circular to explain the formation of expectations by further-ahead survey expectations, Ang, Bekaert, and Wei (2007) put forward that the information contained in median survey expectations may arise from a mechanism similar to Bayesian model averaging, or averaging across different individual forecasts that extracts common components. They also suggest that the satisfactory behavior of survey forecasts in contrast with econometric forecasts might be due to the ability of professional forecasters to identify structural change more quickly. In addition, Cecchetti et al. (2007) provide evidence that survey inflation expectations are correlated with future trend inflation and suggest that surveys have a good forecasting performance because survey respondents anticipate changes in trend inflation.

One interpretation why private agents may use further-ahead expectations - so information at horizons further-ahead than the forecasting horizon - to form their expectations is that further-ahead expectations might be seen as a representation of the long-run of the economy and therefore as a proxy for the expected equilibrium value of inflation, which would in turn depend, for instance, on the central bank credibility to achieve inflation stabilization. This is in line with the argument by Faust and Wright (2012) that inflation expectations for the quarters to come represent the way forecasters believe inflation takes from its current expected value (nowcast) towards the perceived trend or equilibrium inflation rate.

The two main implications of these results for policymakers are first that private expectations depend on past information, and second that anchoring medium- or long-term expectations enables anchoring short-term expectations. Besides the estimated parameters may serve for calibrating macroeconomic models in which private expectations are not solely forward-looking. Finally, another implication for future research is that professional forecasters appear to form their inflation expectations on the grounds of the hybrid NKPC.

The rest of the paper is organized as follows. Section 3.2 describes the methodology. Section 3.3 reports the empirical analysis, while sections 3.4 and 3.5 focus on deviations from the main model with an assessment of the effect of forecasting horizons and a comparison with consumers' forecasts respectively. Section 3.6

concludes.

3.2 Methodology

Galí and Gertler (1999) propose a hybrid New Keynesian Phillips Curve of the following form, where π_t is the inflation rate, $\mathbb{E}_t\pi_{t+1}$ expected future inflation, and mc_t a measure of marginal costs:

$$\pi_t = \lambda mc_t + \gamma_f \mathbb{E}_t\pi_{t+1} + \gamma_b \pi_{t-1}. \tag{3.1}$$

The coefficients γ_f and γ_b are the respective weights on the forward-looking and the backward-looking variable. The equation derives from a New Keynesian model with staggered price setting a la Calvo, where a fraction of firms set their prices using the lagged aggregate inflation rate.

Under the assumption of unbiased expectations it holds that $\pi_t = \mathbb{E}_t\pi_t + \epsilon_t$, where the error term ϵ_t has zero mean.[1] Combining these two equations yields the following NKPC-based inflation expectations formation equation:

$$\mathbb{E}_t\pi_t = \lambda mc_t + \gamma_f \mathbb{E}_t\pi_{t+1} + \gamma_b \pi_{t-1} - \epsilon_t \tag{3.2}$$

We use the output gap x_t as a proxy for marginal costs (as is common in the literature; see e.g. Fuhrer and Moore, 1995; Woodford, 2003) and we measure expected inflation by survey expectations as is recently done in the literature on Phillips curve estimations (see Nunes, 2009; Adam and Padula, 2011) or on monetary policy rules (see e.g. Orphanides, 2001). We thus estimate the following equation, where $\mathbb{S}_t$ represents inflation expectations collected from a survey of forecasters:

$$\mathbb{S}_t\pi_t = \delta x_t + \beta_f \mathbb{S}_t\pi_{t+1} + \beta_b \pi_{t-1} + \nu_t, \tag{3.3}$$

and where the error term $\nu_t = u_t - \epsilon_t$ has zero mean, while it is not restricted otherwise such as the estimation error u_t.

[1] We discuss a test of this assumption later on and analyze what a departure from it would imply for our estimations.

3.3 Empirical Analysis

3.3.1 Data

We focus on quarterly US data for which survey forecasts from the Survey of Professional Forecasters are available on a fixed-horizon scheme[2] and for a long time span: 1981Q3-2012Q3.[3] We use the median of individual responses as our baseline, and propose robustness tests with the mean. SPF inflation forecasts for both the GDP deflator and CPI inflation fulfill stationary requirements.[4] We also analyze how consumer expectations differ from those of professionals making use of the University of Michigan's Survey of Consumers.

Figures 3.1 and 3.2 plot SPF inflation expectations at the current horizon (nowcast) and the one-quarter ahead horizon for both the GDP deflator and CPI inflation. Consistent with US inflation history, inflation expectations followed the disinflation path until the end of the eighties while they have been anchored around two percent ever since. An exception to that is the considerable volatility in the nowcast of CPI inflation around the financial crisis.

As the output gap we employ the filtered version of real GDP growth. We use the nearly optimal one-sided Christiano-Fitzgerald (CF) filter under the common assumption of a business cycle duration of 6 up to 32 quarters (see Christiano and Fitzgerald, 2003). To check the robustness of the results we also use the output gap based on the Hodrick-Prescott filter.

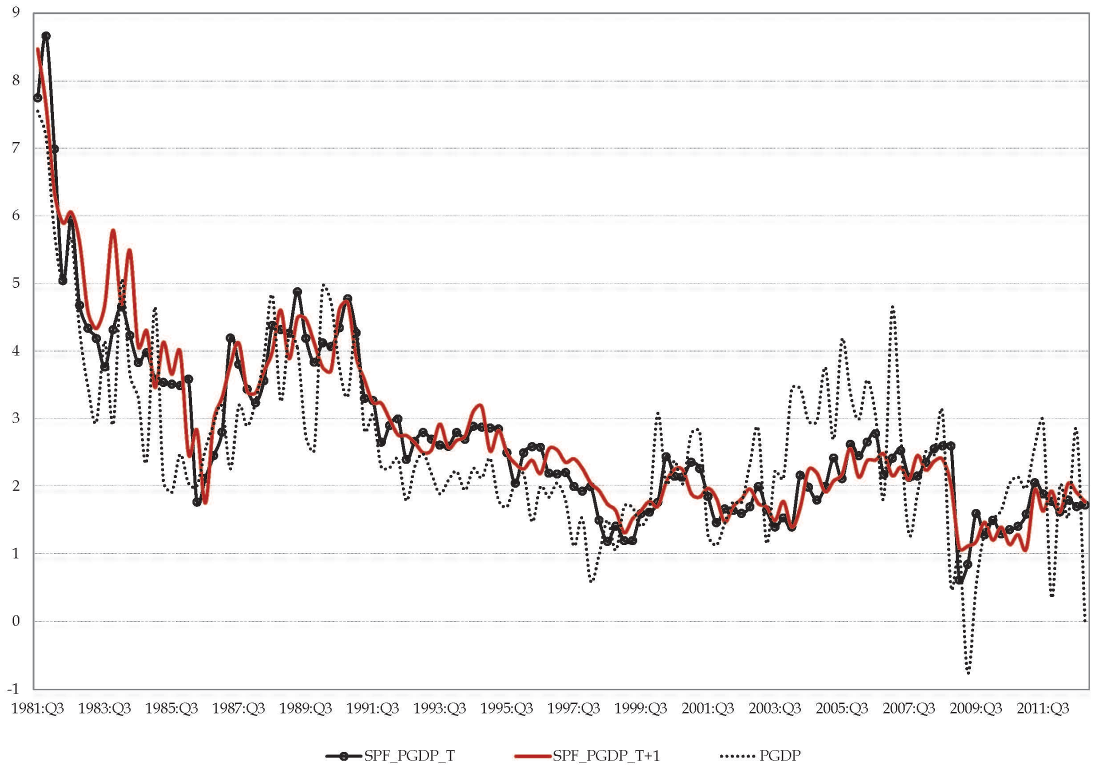
We also employ other marginal cost measures frequently considered in the literature namely unit labor costs, labor share, unemployment rate, inventories, industrial production index and capacity utilization. Further, we evaluate our models with real-time data to examine whether results are different with respect to the use of final revised data. The SPF survey and other real-time data come from the Federal

[2]An advantage of fixed-horizon forecasts compared to fixed-event forecasts is that the latter have a decreasing forecasting horizon in each calendar year. One might thus consider this variable as not being drawn from the same stochastic process which introduces heteroscedasticity in the estimation process.

[3]SPF expectations for the GDP deflator are actually available as of 1968, however, our sample starts at the above-mentioned date in order to fulfill stationarity requirements and to be consistent with respect to CPI inflation for which survey data does not exist before 1981.

[4]Stationarity tests are available from the authors upon request. We find that the null hypothesis of stationarity cannot be rejected for both the GDP deflator and CPI inflation survey variables at all horizons except for three-quarter-ahead expectations of the former inflation measure.

Figure 3.1: Survey PGDP inflation expectations and actual PGDP



Note: This figure shows SPF survey expectations for the GDP deflator (PGDP), as well as its realized values. The following abbreviations are used: spf_pgdp_t is the nowcast of the GDP deflator, spf_pgdp_t+1 is the one-quarter ahead forecast and pgdp is the actual GDP deflator measured with final data.

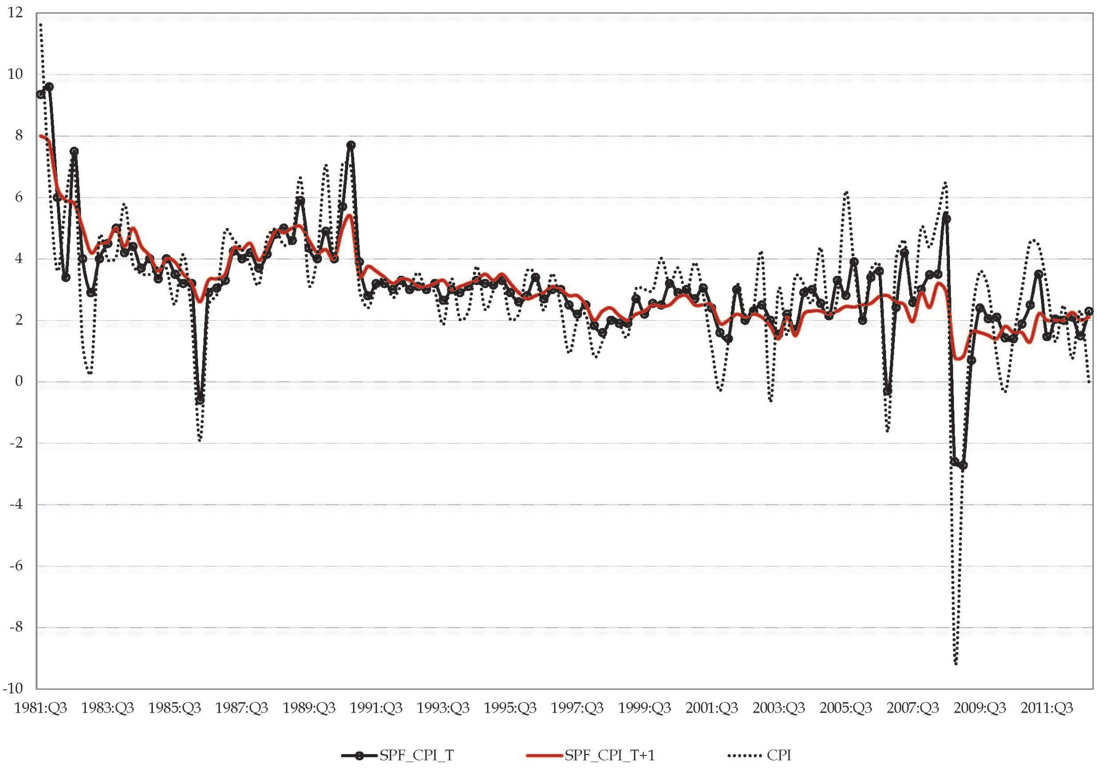
Reserve of Philadelphia, while final data and the University of Michigan's Survey of Consumers (UMSC) are from the FRED database. See the Data Appendix for more details.

3.3.2 Unbiasedness of Expectations

In this section we evaluate the assumption of unbiased expectations. To test for unbiasedness, we estimate a model: $\pi_{t+h} = \alpha_u + \beta_u \mathbb{S}_t \pi_{t+h} + \eta_t$, as is common in the literature (see e.g. Adam and Padula, 2011). Unbiasedness requires the constant α to be equal to zero and β_u to equal 1. If this is not the case a constant enters equation 3.2 and accordingly equation 3.3, and/or the coefficients are divided by a coefficient β_u which, however, would not require a different estimation technique.

The results of these tests are presented in Table 3.10 in the Appendix. Unbiasedness can be rejected for forecasts of GDP deflator at all horizons using final data,

Figure 3.2: Survey CPI inflation expectations and actual CPI



Note: This figure shows SPF survey expectations for CPI inflation, as well as its realized values. The following abbreviations are used: spf_cpi_t is the nowcast of CPI inflation, spf_cpi_t+1 is the one-quarter ahead forecast and cpi is the actual CPI inflation measured with final data.

however, once real-time data is employed it ceases to be the case for the nowcast. For forecasts of CPI inflation the nowcast is unbiased even for final data. To account for potential bias in expectations we estimate all models with a constant α verifying that it is insignificant.

3.3.3 Baseline Results

We present OLS estimates of equation 3.3 for both inflation measures in Table 3.1. We compute heteroskedasticity and autocorrelation robust Newey-West standard errors assuming that the autocorrelation dies out after four quarters. This choice corresponds to the Stock and Watson (2007b) rule of thumb where the Newey-West lag length is set equal to $0.75 \times T^{\frac{1}{3}}$ (rounded), T being the number of observations used in the regression.

The coefficients on the backward- and forward-looking element of the inflation

Table 3.1: NKPC-based inflation expectations formation model

	Baseline		Constrained	
	GDP deflator	CPI inflation	GDP deflator	CPI inflation
δ	-0.04^{*}	0.08	-0.03	0.07
	(0.02)	(0.05)	(0.02)	(0.07)
β_f	0.82^{***}	0.88^{***}	0.84^{***}	0.81^{***}
	(0.04)	(0.11)	(0.04)	(0.07)
β_b	0.14^{***}	0.19^{***}	0.16^{***}	0.19^{***}
	(0.04)	(0.04)	(0.04)	(0.07)
const	0.10	-0.28	-0.03	-0.05
	(0.12)	(0.29)	(0.03)	(0.08)
R^2	0.92	0.73	-	-
$\beta_f + \beta_b = 1$	0.31	0.41	-	-
Obs	124	124	124	124

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of equation 3.3 (including a constant), is conducted by OLS. Asymptotic Newey-West 4 lags standard errors are in parentheses in the 'Baseline' models. The 'Constrained' approach enforces the following condition: $\beta_f + \beta_b = 1$. In this case the variance estimates of the standard errors are the Huber/White/sandwich robust variance estimates. The data set comprises 1981Q3-2012Q3. In the last two rows the R^2 of the regression, as well as the p-value of an F test for the hypothesis that $\beta_f + \beta_b = 1$ are presented for the 'Baseline' estimations. The final row reports the number of observations. The output gap is derived by means of the CF filter.

expectations formation process are estimated to be $(0.82, 0.14)$ and $(0.88, 0.19)$ for the GDP deflator and CPI inflation, respectively. This is, forward-looking dynamics dominate the formation process for both inflation expectation measures, while the backward-looking part is still significant in either case. This outcome is consistent with the literature focusing on the expectations formation process which finds a role, small but significant, for backward-looking behavior as in Lanne, Luoma, and Luoto (2009) or Pfajfar and Santoro (2010). The resulting coefficients are also similar to those found in the literature on estimations of the actual New Keynesian Phillips Curve (see e.g. Galí and Gertler, 1999; Woodford, 2003; Nunes, 2010). It suggests that forecasters may form their forecasts on the grounds of the NKPC assuming that it properly captures inflation dynamics.

In line with the NKPC literature we evaluate the hypothesis that the weights on the backward- and the forward-looking element add up to one by means of a partial F test. For both inflation measures the null hypothesis cannot be rejected. This is exactly what other studies find in their evaluations of the actual NKPC (Galí and Gertler, 1999; Woodford, 2003).

As far as the marginal cost measure is concerned the results for the two inflation variables differ. Whereas the coefficient on the output gap is negative and marginally significant, i.e. at the 10 percent level, for the GDP deflator, it is positive and insignificant for CPI inflation. The negative sign on the output gap coefficient for the GDP deflator model might be a surprise on theoretical grounds, while it is well documented empirically in the NKPC literature (see Woodford, 2003; Nunes, 2010).

The high R^2 of 0.92 for the GDP deflator model among other things derives from the fact that survey expectations of the GDP deflator at different horizons are highly correlated. Given the high correlation among inflation variables and the survey measure we test for multicollinearity evaluating the uncentered variance inflation factors, and we reject it for the models we analyze in this paper and thus do not discuss this issue further. We also verify that including a constant does not improve the fit of the model, as the constant is statistically insignificant in both models.

As is common in the NKPC literature, we further evaluate a model where we constrain the sum of the coefficients β_f and β_b to one (see e.g. Galí and Gertler, 1999). In this case the variance estimates of the standard erros are the Huber-White/sandwich robust variance estimates. The results based on this approach are also presented in Table 3.1. For the GDP deflator the estimates are very similar, while the ouput gap is now completely insignificant. For CPI inflation the constrained approach yields similar coefficients. Given that the estimation of the constrained model involves a change in the dependent variable, no goodness-of-fit measure is provided as it would have a different interpretation.

3.3.4 Model Comparisons

In this section we compare our baseline model to two important alternative inflation expectations formation processes, namely a purely forward-looking ($\gamma_b = 0$ in equation 3.3) and a purely backward-looking model ($\gamma_f = 0$). The estimation results of the previous section already provide support for our NKPC-based expectations formation model, i.e. the fact that the coefficients on the forward- and backward-looking variables are signficantly different from zero can be interpreted as evidence in favour of our baseline model against the two alternative reduced models. To shed more light on this issue, we present parameter estimates for the alternative models

Table 3.2: Model comparisons

	Forward-looking model		Backward-looking model	
	GDP deflator	CPI inflation	GDP deflator	CPI inflation
$\delta_{(a)}$	-0.05*	0.08	-0.02	0.02
	(0.02)	(0.07)	(0.05)	(0.10)
β_f	0.92***	1.08***		
	(0.05)	(0.11)		
β_b			0.78***	0.46***
			(0.08)	(0.07)
const	0.16	-0.29	0.69***	1.62***
	(0.12)	(0.36	(0.19))	(0.16)
R^2	0.91	0.67	0.65	0.41
$\beta_f = 1$	0.11	0.48	-	-
$\beta_b = 1$	-	-	0.01***	0.00***
LR test	0.00***	0.00***	0.00***	0.00***
Obs	124	124	124	124

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of the forward-looking and the backward-looking model is conducted by OLS. Asymptotic Newey-West 4 lags standard errors are in parentheses. The data set comprises 1981Q3-2012Q3. In the rows below the parameter estimates the R^2 of the regression and the p-value of an F test for the hypothesis that the given parameter equals one are presented. Further, the p-value corresponding to an LR test of the alternative model relative to the baseline model and the number of observations are given.

and LR test results to provide evidence in favour or against the alternative models relative to our baseline.

The LR test clearly rejects the reduced models in favour of our baseline NKPC-based inflation expectations formation model for both the GDP deflator and CPI inflation. Note, however, that the LR test is based on the assumption of homoskedastic and non-autocorrelated errors. We thus ask the reader to interpret these results with caution. It stands out though that both LR test results and the t-statistics in Table 3.1 point in the same direction, i.e. our baseline model performs better than the alternatives.

Turning to the parameter estimates, it becomes obvious that the purely backward- and the purely forward-looking model perform very differently. The latter has an R^2 similar to the baseline case and the coefficient β_f is insignificantly different from one. The former model on the other hand has a significantly lower R^2 with the coefficient β_f being significantly smaller than one, while the constant is large and significant. We interpret these results as the purely forward-looking model approximating our baseline model reasonably well, while the backward-looking model is clearly inferior.

In either case though it seems that our baseline model performs better.[5]

3.3.5 Final versus Real-Time Data

We now present estimates based on real-time data since in our context the timing of information is paramount and calls for carefulness. Orphanides (2001) stresses that the use of final revised data in Taylor rule estimations may cause misleading results given that agents can only know the most recent publication of data rather than revisions that would be published in the future. Accordingly the determinants of inflation and hence inflation expectations should then depend on the information available to agents at that time. We thus also evaluate our models with real-time data stemming from the Real-Time Database from the Federal Reserve Bank of Philadelphia.

We replace both the inflation measure as well as the real GDP growth variable used to construct the output gap by their first vintage published. The results for both the GDP deflator and CPI inflation are presented in Table 3.3. The parameter estimates are qualitatively unchanged. While the forward-looking coefficient is somewhat lower and the backward-looking coefficient is somewhat higher than before in the GDP deflator model, both are higher in the CPI model. Note, however, that in the latter model the standard errors are larger which is related to the fact that real-time data for CPI inflation is not available before 1994Q1 and thus 52 observations less are used. Based on real-time data, the coefficient on the output gap becomes insignificant in the GDP deflator model, in the CPI model it is marginally significant.

One can also argue that even the first release of real GDP growth is not yet known at time t, as survey respondents have to provide their answers during a given quarter, while the first vintage of this given quarter will typically not be released before the following quarter. Therefore we replace the output gap measure based on this first release by the output gap measure based on the nowcast for real GDP growth from the SPF. The results are very similar to our baseline estimates as can

[5] We also analyzed an autoregressive model. Performing the two non-nested model tests suggested by Coibion (2010b), we find that both the baseline and the AR model cannot be rejected statistically, while our NKPC-based model is preferred over the alternative. Results are available upon request.

Table 3.3: Real-time data estimation

	First vintage		Nowcast	
	GDP deflator	CPI inflation	GDP deflator	CPI inflation
δ	-0.04	0.13*	-0.04	0.16*
	(0.03)	(0.08)	(0.03)	(0.08)
β_f	0.77***	1.02***	0.76***	1.00***
	(0.05)	(0.29)	(0.05)	(0.28)
β_b	0.17***	0.21***	0.18***	0.21***
	(0.04)	(0.03)	(0.04)	(0.03)
const	0.14	-0.52	0.15	-0.46
	(0.10)	(0.63)	(0.10)	(0.61)
R^2	0.93	0.70	0.93	0.70
$\beta_f + \beta_b = 1$	0.15	0.39	0.16	0.43
Obs	124	72	124	72

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of equation 3.3 (including a constant), is conducted by OLS. Asymptotic Newey-West 4 lags standard errors are in parentheses. The data set comprises 1981Q3-2012Q3 for the GDP deflator and 1994Q3-2012Q3 for the CPI model. In the last two rows the R^2 of the regression, as well as the p-value of an F test for the hypothesis that $\beta_f + \beta_b = 1$ are presented. The final row reports the number of observations. The results for 'First vintage' are based on the first release of both the inflation and the real GDP growth variable. The results for 'Nowcast' rely on the first release of the inflation variable and the nowcast of real GDP growth from the SPF. The output gap is derived by means of the CF filter.

be seen in Table 3.3.

3.3.6 Subsamples

One might ask whether the apparent fit of the NKPC model in explaining inflation expectation dynamics stems from the stability of inflation during the Great Moderation. In other words for a very high degree of autocorrelation in inflation and accordingly inflation a hybrid model, a forward-looking and a backward-looking model would all fit the data well. We have shown earlier that our NKPC-based model fits the data better than some important alternatives over the whole sample and we now want to examine whether our results are robust to the choice of the (sub)sample. Similar estimates would support the idea that the relative weights on past inflation and inflation expectations are not due to the particular process in inflation dynamics as present e.g. in the Great Moderation, but capture well a stable inflation expectation formation process independently of whether inflation itself is stable or decelerating.

Table 3.4: Subsample estimates

	GDP deflator		CPI inflation	
	Pre 1992Q3	Post 1992Q3	Pre 1992Q3	Post 1992Q3
δ	-0.06	-0.02	0.05	0.15*
	(0.04)	(0.02)	(0.06)	(0.07)
β_f	0.79***	0.83***	1.15***	1.07***
	(0.10)	(0.06)	(0.19)	(0.29)
β_b	0.12	0.14***	0.18*	0.16***
	(0.10)	(0.04)	(0.09)	(0.04)
const	0.31	0.06	-1.58***	-0.57
	(0.39)	(0.13)	(0.53)	(0.66)
R^2	0.82	0.79	0.75	0.56
$\beta_f + \beta_b = 1$	0.37	0.57	0.01**	0.39
Obs	43	81	43	81

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of equation 3.3 (including a constant), is conducted by OLS. Asymptotic Newey-West 4 lags standard errors are in parentheses. The data set comprises 1981Q3-2012Q3. In the last two rows the R^2 of the regression, as well as the p-value of an F test for the hypothesis that $\beta_f + \beta_b = 1$ are presented. The final row reports the number of observations. The output gap is derived by means of the CF filter. The break date corresponds to the date when inflation came back to the 2% inflation target.

Table 3.4 provides estimates of our NKPC-based model before and after 1992Q3 when inflation came back to the target range of typically around 2 percent. Further, although the starting date of the Great Moderation is typically set earlier, as of 1992 inflation followed an even more stable path (estimates are immune to the choice of this specific break date and are similar for all break dates tested between 1987 and 1995). Finally, setting the break date that late allows us to have a reasonably large first subsample (43 observations).

For both the GDP deflator and CPI inflation, the coefficient on further-ahead expectations is similar before and after the break date and also corresponds to our estimate for the whole sample. Parameter estimates on past inflation are alike and significant for CPI inflation before and after 1992Q3, while they are similar for the GDP deflator but past inflation only becomes significant after the break date. This, however, can be explained by the relatively small sample size in the first subsample. These results provide evidence that our model indeed fits the data well and that our findings are not influenced by the choice of a particular sample, i.e. they are not driven primarily by the relatively stable inflation rates between 1992 and 2007.

3.3.7 Does More Information Matter?

We also examine whether the lack of some potentially important but omitted variables – the federal funds rate and oil prices – may bias the baseline estimates. Survey respondents might base their expectations on more information than is incorporated in equation 3.3 and one way to test whether forecasters form their expectations on the grounds of the NKPC is to add more variables to the regression to evaluate whether additional information changes our baseline estimates. We include a lag of the federal funds rate - denoted i - to represent the stance of monetary policy, as well as of the oil price growth rate - denoted oil - which can be interpreted as an external price shock, and analyze how these affect the results. Given the high autocorrelation in the interest rate (see e.g. Galí and Gertler, 1999; Mavroeidis, 2010), the previous stance of monetary policy might give an idea about the present and future stances. Similarly, in light of the fact that an external price shock takes some time to feed through the economy the shock history tells us something about future developments. The estimation results for equation 3.4 below (including a constant) are given in Table 3.5:

$$\mathbb{S}_t\pi_t = \delta x_t + \beta_f \mathbb{S}_t\pi_{t+1} + \beta_b\pi_{t-1} + \gamma_i\, i_{t-1} + \gamma_o\, oil_{t-1} + \eta_t. \tag{3.4}$$

The additional information does not seem to improve the fit of the GDP deflator model. The R^2 is almost the same as in the baseline case and the parameter estimates are essentially unchanged. The coefficient on the interest rate is insignificant, while the oil price coefficient is significant but very small. The conclusions from the baseline model remain unaltered and it seems that omitted variable bias in not an issue for the GDP deflator model.

The results for the CPI inflation model differ slightly. The coefficient on the oil price is insignificant, while the one on the interest rate is marginally significant, at the 10 percent level. γ_i is about -0.10, thus a 100 basis points increase in the lagged federal funds rate would - as expected - decrease the nowcast of CPI inflation by 0.1 percent above the indirect effect it has through expected inflation for the following period. At the same time the R^2 increases slightly from around 0.73 to around 0.75 relative to the baseline case. The output gap still has an insignificant

Table 3.5: Omitted variable bias

	GDP deflator	CPI inflation
δ	-0.04^{**}	0.05
	(0.02)	(0.05)
β_f	0.78^{***}	1.17^{***}
	(0.07)	(0.23)
β_b	0.11^{***}	0.17^{***}
	(0.04)	(0.05)
γ_i	0.03	-0.10^{*}
	(0.02)	(0.05)
γ_o	0.002^{**}	0.003
	(0.001)	(0.002)
const	0.11	-0.59
	(0.13)	(0.40)
R^2	0.92	0.75
$\beta_f + \beta_b = 1$	0.14	0.10*
Obs	124	124

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of equation 3.4 (including a constant), is conducted by OLS. Asymptotic Newey-West 4 lags standard errors are in parentheses. The data set comprises 1981Q3-2012Q3. In the last two rows the R^2 of the regression, as well as the p-value of an F test for the hypothesis that $\beta_f + \beta_b = 1$ are presented. The final row reports the number of observations. The output gap is derived by means of the CF filter.

coefficient. Finally, the coefficient on the forward-looking variable, γ_f, increases to 1.17. Given the relatively high standard error on the forward-looking variable, the hypothesis that the backward- and forward-looking coefficients add up to one cannot be rejected. It thus seems that in either case omitted variable bias is not present for our basline NKPC-based inflation expectations formation process.

3.3.8 Robustness

In the following we discuss various robustness checks. First, we examine other variables for marginal cost measures such as unit labor costs that are typically used in the NKPC literature. The ouput gap we use so far is constructed by means of the CF filter. Another filter that is commonly used in the literature is the Hodrick-Prescott (HP) filter (see e.g. Nunes, 2010). Therefore we show how our results change if we use this latter approach to construct the output gap. More importantly, many authors question the usefulness of the output gap to represent marginal costs in estimations of Phillips curves (among them Galí and Gertler, 1999; Sbordone, 2002; Galí, Gertler, and López-Salido, 2005). Other variables commonly suggested are unit labor costs, labor share, unemployment rate (as in the original Phillips curve), industrial production, capacity utilization or inventories. Estimation results for our models based on these marginal cost measures, as well as the different output gap are presented in the Appendix in Table 3.11. Given potential measurement error due to the use of surveys (for a discussion of this point see Adam and Padula, 2011) and potential endogeneity we also review our model results with the use of GMM. Finally, we analyze whether results differ for the mean versus the median of individual responses for expected inflation; see Table 3.12 and 3.13 for GMM based results and those based on the mean rather than the median, respectively. The main conclusions of Section 3.3.3 are robust to the different approaches presented in the Appendix.

3.4 The Effect of Forecasting Horizons

In this section, we depart from our baseline model in two ways. First, we increase the horizon of inflation expectations used by private agents to determine current inflation expectations. Second, we assess whether the formation process of inflation expectations for future quarters differs from the formation process of inflation expectations for the current quarter.

3.4.1 Near vs. Further-Ahead Forward-Looking Information

We aim at establishing the role of the horizon of forward-looking information in the expectations formation process, and more precisely whether private forecasters put relatively more weight on near or further-ahead forward-looking information. On the one hand one may expect that private agents have a better understanding of the closer economic outlook and thus put more weight on forward-looking information with a shorter horizon; on the other hand private agents might use forward-looking information as a representation of the long-run of the economy and of the equilibrium value of inflation and therefore put more emphasis on further-ahead forward-looking information.

Table 3.6: Near vs. further-ahead forward-looking information

	GDP deflator				CPI inflation				
	$\mathbb{S}_t\pi_t$	$\mathbb{S}_t\pi_t$	$\mathbb{S}_t\pi_t$	$\mathbb{S}_t\pi_t$	$\mathbb{S}_t\pi_t$	$\mathbb{S}_t\pi_t$	$\mathbb{S}_t\pi_t$	$\mathbb{S}_t\pi_t$	$\mathbb{S}_t\pi_t$
δ	-0.02	-0.04*	-0.05**	-0.04*	0.08	0.07	0.07	0.08	0.12
	(0.02)	(0.02)	(0.02)	(0.02)	(0.06)	(0.07)	(0.07)	(0.06)	(0.12)
$\beta_f(\mathbb{S}_t\pi_{t+2})$	0.74***				0.73***				
	(0.04)				(0.09)				
$\beta_f(\mathbb{S}_t\pi_{t+3})$		0.68***				0.68***			
		(0.04)				(0.08)			
$\beta_f(\mathbb{S}_t\pi_{t+4})$			0.68***				0.64***		
			(0.04)				(0.08)		
$\beta_f(\mathbb{S}_t\tilde{\pi}_{t+4})$				0.79***				0.75***	
				(0.04)				(0.09)	
$\beta_f(\mathbb{S}_t\pi_{t+10y})$									0.63***
									(0.17)
β_b	0.23***	0.24***	0.29***	0.18***	0.26***	0.28***	0.29***	0.25***	0.26***
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
const	-0.02	0.08	-0.03	-0.03	-0.07	-0.02	0.04	-0.13	0.07
	(0.11)	(0.13)	(0.12)	(0.12)	(0.26)	(0.24)	(0.26)	(0.27)	(0.48)
R^2	0.90	0.89	0.89	0.92	0.64	0.62	0.61	0.65	0.35
$\beta_f + \beta_b = 1$	0.46	0.12	0.45	0.48	0.91	0.57	0.36	1.00	0.48
Obs	124	124	124	124	124	124	124	124	84

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of equation (3.3), is conducted by OLS, where the horizon of the forward-looking component varies. $\mathbb{S}_t\tilde{\pi}_{t+4}$ represents the average expected inflation rate over the following four quarters. Asymptotic Newey-West 4 lags standard errors are in parentheses. The data set comprises 1981Q3-2012Q3, except for 10-year-ahead CPI expectations which start in 1991Q4. In the rows below the parameter estimates the R^2 of the regression, the p-value of an F test for the hypothesis that $\beta_f + \beta_b = 1$ and the number of observations are presented.

The results for both GDP deflator and CPI models have a similar pattern given in Table 3.6. The weight of forward-looking information decreases with the forecasting horizon, from 0.82 at the one-quarter-ahead horizon to 0.68 at the four-quarter-ahead horizon for the GDP deflator model and from 0.88 to 0.64 for the CPI model. Accordingly the weight on the backward-looking variable increases such that the sum of the forward- and backward-looking variable remains insignificantly different from one. The R-square decreases as the horizon increases, however not by much. It thus seems that private agents rely more on their assessment of the near economic outlook rather than on further-ahead perspectives, while the latter still has significant information for the nowcast.

Table 3.6 also features results on a model where the forward-looking component is the average expected inflation rate over the following four quarters ($\mathbb{S}_t\tilde{\pi}_{t+4}$). This model can be justified, as agents might find it easier to make predictions for an average over some quarters rather than for an individual quarter. They thus use this arguably more reliable average in their information set when forming their nowcast. The results indicate that this model works about as well as the benchmark for the GDP deflator, i.e. parameter estimates, an F-test on the sum of the two coefficients of interest and the R^2 are about the same. For the CPI model the R^2 is somewhat lower and the backward-looking variable receives a higher weight as in the benchmark case.

In addition, it is worth noting that for the CPI model, we also have 10-year-ahead expectations (on a smaller subsample starting in 1991Q4) and that the coefficient estimated is 0.63, very close to the 1-year-ahead estimate. Beyond this latter horizon, private forecasters give a similar weight to forward-looking information which suggests that these expectations capture the private agents' view on the long-run equilibrium value of inflation.

3.4.2 Different Expectation Pairs

We now assess whether the formation process of inflation expectations for future quarters differ from the formation process of inflation expectations for the current quarter. In this model, we continue to consider that forecasts at the horizon h are determined by forecasts at the horizon $h+1$ and we vary the value of h.

Table 3.7: The formation process of expectations at longer horizons

	GDP deflator			CPI inflation			
	$\mathbb{S}_t\pi_{t+1}$	$\mathbb{S}_t\pi_{t+2}$	$\mathbb{S}_t\pi_{t+3}$	$\mathbb{S}_t\pi_{t+1}$	$\mathbb{S}_t\pi_{t+2}$	$\mathbb{S}_t\pi_{t+3}$	$\mathbb{S}_t\pi_{t+4}$
δ	0.02	-0.03*	-0.002	0.01	0.001	-0.01	-0.03
	(0.02)	(0.01)	(0.02)	(0.02)	(0.00)	(0.01)	(0.02)
$\beta_f(\mathbb{S}_t\pi_{t+2})$	0.84***			0.95***			
	(0.04)			(0.02)			
$\beta_f(\mathbb{S}_t\pi_{t+3})$		0.87***			0.98***		
		(0.04)			(0.01)		
$\beta_f(\mathbb{S}_t\pi_{t+4})$			0.86***			0.95***	
			(0.06)			(0.01)	
$\beta_f(\mathbb{S}_t\pi_{t+10y})$							1.04***
							(0.07)
β_b	0.16***	0.06	0.17**	0.04***	0.01*	0.01	0.03***
	(0.03)	(0.04)	(0.07)	(0.01)	(0.01)	(0.01)	(0.01)
const	-0.09	0.18**	-0.01	-0.02	-0.06	0.06	-0.32*
	(0.07)	(0.07)	(0.12)	(0.06)	(0.04)	(0.05)	(0.19)
R^2	0.93	0.93	0.89	0.96	0.98	0.98	0.88
$\beta_f + \beta_b = 1$	0.92	0.01***	0.57	0.56	0.62	0.03**	0.33
Obs	124	124	124	124	124	124	84

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of equation (3.3), is conducted by OLS. Asymptotic Newey-West 4 lags standard errors are in parentheses. The data set comprises 1981Q3-2012Q3, except for 10-year-ahead CPI expectations which start in 1991Q4. In the rows below the parameter estimates the R^2 of the regression, the p-value of an F test for the hypothesis that $\beta_f + \beta_b = 1$ and the number of observations are presented.

For the GDP deflator model, the weight put on backward- and forward-looking information does not differ dramatically from the baseline model when h varies as can be seen in Table 3.7. One exception is the β_b coefficient for $h = 2$ which is insignificant while the constant is significant. For the CPI model, the coefficient on forward-looking information is sightly higher than in the baseline estimation when h varies, but most importantly the backward coefficient becomes null for $h = 2$ and 3. Finally, we estimate the effect of 10-year-ahead expectations on four-quarter-ahead expectations for the CPI model, and find an even larger and highly significant weight on forward-looking information.

These results suggest that the inflation expectations formation process and the relationship between inflation expectations and both backward- and forward-looking information are relatively stable across the horizons that private agents are typically considering.

3.5 Consumers vs. Professionals

Carroll (2003) compares professional and consumer forecasts and finds that household expectations are not rational and that professional forecasts, which may be considered rational, spread epidemiologically to the public. We here aim at shedding light on the potential discrepancy between the expectations formation process of professionals and consumers in order to assess whether consumers use the same relative weights on backward- and forward-looking information and whether the NKPC-based expectations model also fits their expectations.

We use the University of Michigan's Survey of Consumers to measure consumers' inflation expectations, available since 1991Q4 with a regular quarterly frequency. The survey collects forecasts at the 4-quarter horizon and at the 5-year horizon and we estimate the effect of the latter in setting the former. We compare this model to the closest available pairs of professional expectations, i.e. the effect of 10-year-ahead forecasts on 4-quarter-ahead forecasts for CPI inflation. Estimates are given in Table 3.8.

Table 3.8: Consumers vs. professionals

	SPF	UMSCI
	$\mathbb{S}_t\pi_{t+4}$	$\mathbb{S}_t\pi_{t+4}$
δ	-0.03	-0.05
	(0.02)	(0.03)
$\beta_f(\mathbb{S}_t\pi_{t+10y})$	1.04***	
	(0.07)	
$\beta_f(\mathbb{S}_t\pi_{t+5y})$		0.30**
		(0.14)
β_b	0.03***	0.12***
	(0.01)	(0.03)
const	-0.32*	1.78***
	(0.19)	(0.44)
R^2	0.88	0.35
$\beta_f + \beta_b = 1$	0.33	0.00***
Obs	84	90

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of equation 3.3 (including a constant), is conducted by OLS. Asymptotic Newey-West 4 lags standard errors are in parentheses. The data set comprises 1990Q2-2012Q3 for the University of Michigan's Survey of Consumers: Inflation, and 1991Q4-2012Q3 for the SPF CPI series. In the last two rows the R^2 of the regression, as well as the p-value of an F test for the hypothesis that $\beta_f + \beta_b = 1$ are presented. The final row reports the number of observations. The output gap is derived by means of the CF filter.

Estimates show a clear difference in the expectations formation process of professionals and consumers. The coefficient on forward-looking information for consumers is 0.30, very low compared to professionals: 1.04 while the coefficient on backward-looking information is higher: 0.12 compared to 0.03. The sum of the coefficients does not add up to one in the case of consumers and the constant is high 1.78 and strongly significant. As far as this specific expectation pair is concerned (10-year-ahead and 4-quarter-ahead expectations) we can thus reject the hypothesis that consumers form their forecasts on the grounds of the NKPC. Given a lack of consumer expectations in the Michigan survey for other horizons we leave the important question of how consumer expectations are formed at different horizons for future research.

3.6 Conclusion

In this paper, we aim at establishing the role of backward-, present and forward-looking information in the private inflation expectations formation process using an NKPC-based expectations formation model. We find that professional forecasters put relatively more weight on forward-looking information, while past information is significant and the contribution of the marginal cost measure is small and often insignificantly different from zero. These findings are robust to the use of real-time data, to various measures of marginal costs, to the use of the mean of individual responses, to another estimation procedure namely GMM, and to the inclusion of potentially relevant additional variables. The estimated coefficients are similar to those found in the literature estimating the actual NKPC suggesting that professional forecasters indeed use this model to form their own inflation expectations. This result also holds for two different subsamples where during one inflation decreases rapidly while during the other it is relatively stable. We also find that the estimated parameters of the NKPC-based expectations formation model are relatively stable when the forecasting horizon varies or when we consider further-ahead horizons for forward-looking information. Finally, consumers differ from professional forecasters in that their expectations formation process cannot be adequately modelled based on an NKPC.

A3 Appendix to Chapter 3

A3.1 Data Appendix

Table 3.9: Data

Name	Description	Original frequency	Time period
	Real-time data first release		
rgdp_1st	Real GDP growth	Quarterly	1981Q3-2012Q3
pgdp_1st	GDP deflator	Quarterly	1981Q3-2012Q3
cpi_1st	Consumer price index	Quarterly	1994Q3-2012Q3
	Final data		
rgdp	Real GDP growth	Quarterly	1981Q3-2012Q3
pgdp	GDP deflator	Quarterly	1981Q3-2012Q3
cpi	Consumer price index	Quarterly	1981Q3-2012Q3
ulc	Unit labor costs	Quarterly	1981Q3-2012Q3
ls	Labor share	Quarterly	1981Q3-2012Q3
unemp	Unemployment rate	Quarterly	1981Q3-2012Q3
indpro	Industrial production index	Quarterly	1981Q3-2012Q3
cap_uti	Capacity utilization	Quarterly	1981Q3-2012Q3
invent	Inventories	Quarterly	1981Q3-2012Q3
	Survey data (x-quarters-ahead horizon)		
spf_pgdp_0	SPF pgdp expectations (0)	Quarterly	1981Q3-2012Q3
spf_pgdp_1	SPF pgdp expectations (1)	Quarterly	1981Q3-2012Q3
spf_pgdp_2	SPF pgdp expectations (2)	Quarterly	1981Q3-2012Q3
spf_pgdp_3	SPF pgdp expectations (3)	Quarterly	1981Q3-2012Q3
spf_pgdp_4	SPF pgdp expectations (4)	Quarterly	1981Q3-2012Q3
spf_cpi_0	SPF cpi expectations (0)	Quarterly	1981Q3-2012Q3
spf_cpi_1	SPF cpi expectations (1)	Quarterly	1981Q3-2012Q3
spf_cpi_2	SPF cpi expectations (2)	Quarterly	1981Q3-2012Q3
spf_cpi_3	SPF cpi expectations (3)	Quarterly	1981Q3-2012Q3
spf_cpi_4	SPF cpi expectations (4)	Quarterly	1981Q3-2012Q3
spf_cpi_10	SPF cpi expectations (10 years)	Quarterly	1991Q4-2012Q3
msi_1	UMSC cpi expectations (1 year)	Quarterly	1990Q2-2012Q3
spf_5	UMSC cpi expectations (5 years)	Quarterly	1990Q2-2012Q3

This appendix lists the data that we use in the estimation of our models, as well as the respective sources. We use quarterly frequency of the data series, where monthly series are converted to quarterly frequency by taking the three-month average. The following releases of the data are used: Final, first release and third release. The data series are available for the time periods as indicated in Table 3.9 below and come from the following sources: Real-time and SPF survey data from the website of the Federal Reserve of Philadelphia and final data and the University of Michigan's Survey of Consumers (UMSC) from the Federal Reserve of St. Louis FRED database. For all price series annualized quarter on quarter growth rates are calculated as: $\pi_t = ((\frac{p(t)}{p(t-1)})^4 - 1) \times 100$.

A3.2 Preliminary Tests

Table 3.10: Unbiasedness of survey inflation expectations

	Horizons (x quarters ahead)				
GDP deflator (final)	0	1	2	3	4
α	0.50**	0.80***	0.92***	1.18***	1.28***
	(0.22)	(0.26)	(0.30)	(0.30)	(0.29)
β_u	0.77***	0.63***	0.55***	0.44***	0.40***
	(0.06)	(0.07)	(0.08)	(0.08)	(0.07)
$\beta_u = 1$	0.00***	0.00***	0.00***	0.00***	0.00***
GDP deflator (1st release)	0	1	2	3	4
α	0.02	0.35*	0.43*	0.54**	0.71***
	(0.18)	(0.19)	(0.23)	(0.24)	(0.24)
β_u	0.92***	0.76***	0.69***	0.63***	0.56***
	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)
$\beta_u = 1$	0.17	0.00***	0.00***	0.00***	0.00***
CPI inflation (final)	0	1	2	3	4
α	−0.29	1.07**	1.43***	1.35**	1.55**
	(0.38)	(0.42)	(0.45)	(0.55)	(0.59)
β_u	1.09***	0.62***	0.49***	0.49***	0.42**
	(0.12)	(0.12)	(0.13)	(0.16)	(0.17)
$\beta_u = 1$	0.45	0.00***	0.00***	0.00***	0.00***
CPI inflation (1st release)	0	1	2	3	4
α	0.13	3.12***	2.50***	2.40**	2.36*
	(0.88)	(0.87)	(0.76)	(1.13)	(1.35)
β_u	1.00***	−0.26	−0.00	0.04	0.06
	(0.34)	(0.40)	(0.31)	(0.40)	(0.48)
$\beta_u = 1$	0.99	0.00***	0.00***	0.02**	0.05*

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of the equation $\mathbb{S}_t\pi_t = \alpha + \beta_u\pi_t + \eta_t$ is conducted with OLS for each pgdp and cpi inflation and with both real-time data (1st release) and final revised data. Asymptotic Newey-West 4 lags standard errors are in parentheses. The data set goes from 1981Q3-2012Q3, for the first three inflation measures, while it does not start before 1994Q3 for the first release of CPI inflation. Below the parameter estimates the p-value corresponding to a t test of $\beta_u = 1$ is presented.

A3.3 Robustness Tests

Other Marginal Cost Measures

Results with the HP-filtered output gap are very similar to the benchmark with the exception that the output gap is now significant at the 5 percent level in the GDP deflator model, though the coefficient is the same. The output gap measure remains insignificant for the CPI model. Thus, the results for the output gap coefficient are not sensitive to the choice of the filtering method.

Table 3.11: Other marginal cost measures

	Marginal cost measure						
GDP deflator	HP-GAP	ULC	LS	UNEMP	INDPRO	CAPUTI	INVENT
δ	-0.04^{**}	0.04^{*}	0.00	-0.01	-0.01	0.01	-0.00
	(0.02)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.00)
β_f	0.81^{***}	0.80^{***}	0.80^{***}	0.82^{***}	0.81^{***}	0.80^{***}	0.81^{***}
	(0.04)	(0.05)	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)
β_b	0.14^{***}	0.12^{***}	0.16^{***}	0.15^{***}	0.15^{***}	0.16^{***}	0.16^{***}
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
const	0.09	0.11	-0.40	0.13	0.08	-0.70	0.07
	(0.11)	(0.14)	(1.22)	(0.19)	(0.13)	(0.67)	(0.13)
R^2	0.92	0.92	0.91	0.91	0.91	0.91	0.91
$\beta_f + \beta_b = 1$	0.32	0.23	0.55	0.57	0.51	0.51	0.53
Obs	124	124	124	124	124	124	124
CPI inflation	HP-GAP	ULC	LS	UNEMP	INDPRO	CAPUTI	INVENT
δ	0.04	-0.06	-0.07^{*}	-0.06	0.02	0.01	0.00
	(0.06)	(0.05)	(0.04)	(0.05)	(0.02)	(0.02)	(0.00)
β_f	0.86^{***}	0.91^{***}	0.99^{***}	0.88^{***}	0.86^{***}	0.85^{***}	0.87^{***}
	(0.10)	(0.14)	(0.16)	(0.12)	(0.10)	(0.10)	(0.11)
β_b	0.20^{***}	0.20^{***}	0.18^{***}	0.19^{***}	0.19^{***}	0.19^{***}	0.18^{***}
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
const	-0.23	-0.28	6.45^{*}	0.11	-0.24	-1.35	-0.25
	(0.29)	(0.33)	(3.77)	(0.26)	(0.28)	(1.60)	(0.28)
R^2	0.73	0.72	0.73	0.72	0.73	0.72	0.73
$\beta_f + \beta_b = 1$	0.49	0.38	0.21	0.46	0.54	0.66	0.53
Obs	124	124	124	124	124	124	124

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of equation (3.3), is conducted by OLS. Asymptotic Newey-West 4 lags standard errors are in parentheses. The data set comprises 1981Q3-2012Q3. In the rows below the parameter estimates the R^2 of the regression, the p-value of an F test for the hypothesis that $\beta_f + \beta_b = 1$ and the number of observations are presented. The following abbreviations for the marginal cost measures are used: HP-GAP=HP filter-based output gap, ULC=Unit labor costs, LS=Labor share, UNEMP=Unemployment, INDPRO=Industrial production, CAPUTI= Capacity utilization, INVENT=Inventories.

Using unit labor costs as is common in many studies (e.g. Adam and Padula, 2011), we find a positive coefficient for the GDP deflator model as would be predicted by theory. The coefficient is only marginally significant, i.e. at the 10 percent level. For all other marginal cost measures the coefficient δ is very close to and statistically insignificantly different from zero in the GDP deflator case. The estimates for β_f and β_b are very similar to those presented in Table 3.1 and 3.3.

For the CPI inflation models all marginal cost measures result in an insignificant coefficient except for the labor share. For the latter we find a negative and marginally significant coefficient. In this model also the constant is marginally significant unlike in the other models, where it is always insignificant. The results for the backward- and the forward-looking coefficients are similar as before. The null hypothesis of the two coefficients adding up to one cannot be rejected in any case.

GMM

As argued by Adam and Padula (2011), analyses based on survey data might be subject to measurement errors, i.e. it is not clear that expectations are adequately measured nor that survey expectations respresent actual expectations. Further, it is not clear *ex ante* whether expectations of future inflation influence the nowcast or vice versa. Thus endogeneity issues might be present. For these reasons we estimate the model by GMM instrumenting the forward-looking variable; see GMM1 in Table 3.12. Given that the output gap is potentially unobserved, we also estimate a version, where the output gap is instrumented as well; see GMM2 in the same table.[6]

The results for the two different GMM estimation approaches do not differ. Thus, from here on we refer to both approaches just as the GMM results. For the GDP deflator the GMM approach yields a significant output gap coefficient with a similar value as before. However, compared to the benchmark model, the weights on the forward- and backward-looking variables change. While the former increases to around 0.87, the latter is smaller around 0.08. In any case the two remain significant and the hypothesis of these adding to one can still not be rejected at the conventional 5 percent level. The R^2 is almost not affected.

[6] We use the same instrument set as Nunes (2010), namely four lags of inflation and two lags each of unit labor costs, wage inflation, output gap and SPF expected inflation one-quarter ahead. This instrument set is based on Galí, Gertler, and López-Salido (2005), while the survey data has been added given that surveys are used as the endogenous variable rather than actual future inflation.

Table 3.12: GMM estimation

	GDP deflator		CPI inflation	
	GMM1	GMM2	GMM1	GMM2
δ	-0.03^{**}	-0.03^{**}	0.04	0.05
	(0.01)	(0.01)	(0.03)	(0.04)
β_f	0.87^{***}	0.87^{***}	0.79^{***}	0.79^{***}
	(0.03)	(0.03)	(0.03)	(0.03)
β_b	0.08^{**}	0.08^{**}	0.21^{***}	0.21^{***}
	(0.03)	(0.03)	(0.02)	(0.02)
$const$	0.11	0.11	-0.04	-0.03
	(0.07)	(0.07)	(0.10)	(0.10)
R^2	0.90	0.90	0.68	0.68
$\beta_f + \beta_b = 1$	0.06^{*}	0.06^{*}	0.88	0.86
$Hansen\ J$	0.72	0.66	0.87	0.80
$Kleibergen - Paap$	81.51	72.20	418.28	396.95
Obs	121	121	121	121

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of equation 3.3 (including a constant), is conducted by GMM, where the covariance matrix is corrected by the Newey-West approach with automatic bandwith selection. Standard errors are in parentheses. The instrument set consists of four lags of inflation, and two lags each of SPF expected inflation one-quarter ahead, unit labor costs, the output gap and wage inflation. Under GMM1 the results for the model where only the forward-looking variable is instrumented are given, while for GMM2 also the output gap is treated as endgenous. The output gap is derived by means of the CF filter. The data set comprises 1981Q3-2012Q3. Below the parameter estimates the R^2 of the regression, as well as the p-value of an F test for the hypothesis that $\beta_f + \beta_b = 1$ are presented. Further, the p-value corresponding to the Hansen J statistic, as well as the Kleibergen-Paap statistic are given. Maximal IV size critical values for the latter come from Stock and Yogo (2005) and are 40.09, 22.06 and 15.56 for GMM1 and 31.11, 17.06 and 12.25 for GMM2 at the 10, 15 and 20 percent level, respectively. The final row reports the number of observations.

For CPI inflation GMM yields results very similar to the benchmark: only γ_f is somewhat smaller around 0.79, while the rest of the results remain almost unchanged.

We perform some tests to examine the validity of the GMM approach. First, we present the p-value corresponding to the Hansen J statistic. The p-value, above 0.60 in all cases, shows that the null hypothesis of valid overidentifiying restrictions cannot be rejected. Second, we report the Kleibergen-Paap rank statistic that corresponds to the first-stage F statistic allowing for heteroskedastic and autocorrelated errors. As shown in Table 3.12 it exceeds the critical values by far and thus allows us to reject the null hypothesis of weak instruments.

Mean vs. Median Expectations

The Survey of Professional Forecasters also reports the mean of all respondents' expectations. Although the mean might be influenced by potential outliers, it seems

Table 3.13: SPF mean expectations

	GDP deflator	CPI inflation
δ	-0.01	0.07
	(0.02)	(0.05)
β_f	0.90***	0.85***
	(0.03)	(0.10)
β_b	0.06**	0.20***
	(0.03)	(0.03)
const	0.03	-0.19
	(0.06)	(0.27)
R^2	0.95	0.76
$\beta_f + \beta_b = 1$	0.07*	0.56
Obs	124	124

***, **, and * denote significance at the 1, 5 and 10 percent level, respectively. Estimation of equation 3.3 (including a constant), is conducted by OLS. Asymptotic Newey-West 4 lags standard errors are in parentheses. The data set comprises 1981Q3-2012Q3. In the last two rows the R^2 of the regression, as well as the p-value of an F test for the hypothesis that $\beta_f + \beta_b = 1$ are presented. The final row reports the number of observations. The output gap is derived by means of the CF filter.

worthwile to examine whether our conclusions so far hold for this expectation measure. Table 3.13 contains estimation results for SPF mean expectations.

The results for the GDP deflator are comparable to the benchmark, however, they differ in a few points. First, the output gap measure is statistically insignificant. Second, the forward-looking coefficient is somewhat larger around 0.90, while the backward-looking coefficient is below 0.10, both being significant in all cases. However, the hypothesis of these two adding up to one still cannot be rejected at the 5 percent level. Finally, the R^2 is slightly larger than before at around 0.95.

For CPI inflation the results are even closer to the benchmark. Apart from a slightly smaller forward-looking coefficient and a slightly higher R^2 no differences can be detected.

Bibliography

Adam, Klaus and Mario Padula. 2011. “Inflation Dynamics and Subjective Expectations in the United States.” *Economic Inquiry* 49 (1):13–25.

Ahn, Seung C. and Alex R. Horenstein. 2009. “Eigenvalue Ratio Test for the Number of Factors.” Mimeo.

Andrade, Philippe and Hervé Le Bihan. 2010. “Inattentive Professional Forecasters.” Document de travail No.307, Banque De France.

Andrews, Donald W. K., Marcelo J. Moreira, and James H. Stock. 2006. “Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression.” *Econometrica* 74 (3):715–752.

Andrews, Donald W.K. 1991. “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation.” *Econometrica* 59 (3):817–858.

Ang, Andrew, Geert Bekaert, and Min Wei. 2007. “Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?” *Journal of Monetary Economics* 54 (4):1163–1212.

Atkeson, Andrew and Lee E. Ohanian. 2001. “Are Phillips Curves Useful for Forecasting Inflation.” *Federal Reserve Bank of Minneapolis Quarterly Review* 25 (1):2–11.

Bai, Jushan and Serena Ng. 2002. “Determining the Number of Factors in Approximate Factor Models.” *Econometrica* 70 (1):191–221.

———. 2008. “Selecting Instrumental Variables in a Data Rich Environment.” *Journal of Time Series Econometrics* 1 (1):1–32.

———. 2010. "Instrumental Variable Estimation in a Data Rich Environment." *Econometric Theory* 26 (6):1577–1606.

Ball, Laurence, Gregory Mankiw, and Ricardo Reis. 2005. "Monetary Policy for Inattentive Economies." *Journal of Monetary Economics* 52 (4):703–725.

Bańbura, Marta, Domenico Giannone, Michèle Modugno, and Lucrezia Reichlin. 2012. "Now-Casting and the Real-Time Data Flow." Working Papers ECARES 2012-026, Universite Libre de Bruxelles.

Benkovskis, Konstantins, Michele Caivano, Antonello D'Agostino, Alistair Dieppe, Samuel Hurtado, Tohmas Karlsson, Eva Ortega, and Tímea Várnai. 2011. "Assessing the Sensitivity of Inflation to Economic Activity." Working Paper Series 1357, European Central Bank.

Bernanke, Ben and Jean Boivin. 2003. "Monetary policy in a data-rich environment." *Journal of Monetary Economics* 50 (3):525–546.

Beyer, Andreas, Roger E.A. Farmer, Jérôme Henry, and Massimiliano Marcellino. 2008. "Factor Analysis in a Model with Rational Expectations." *Econometrics Journal* 11 (2):271–286.

Boivin, Jean and Marc P. Giannoni. 2006. "Has Monetary Policy Become More Effective?" *The Review of Economics and Statistics* 88 (3):445–462.

Branch, William A. 2007. "Sticky Information and Model Uncertainty in Survey Data on Inflation Expectations." *Journal of Economic Dynamics and Control* 31 (1):245–276.

Buelens, Christian. 2012. "Inflation Forecasting and the Crisis: Assessing the Performance of Different Forecasting Models and Methods." *European Comission Economic Papers* (451).

Canova, Fabio. 2007. "G-7 Inflation Forecasts: Random Walk, Phillips Curve, or What Else?" *Macroeconomic Dynamics* 11 (1):1–30.

Canova, Fabio and Luca Sala. 2009. "Back to square one: Identification issues in DSGE models." *Journal of Monetary Economics* 56 (4):431–449.

Carroll, Christopher D. 2003. "Macroeconomic Expectations of Households and Professional Forecasters." *Quarterly Journal of Economics* 118 (1):269–298.

Castelnuovo, Efrem, Sergio Nicoletti-Altimari, and Diego Rodríguez-Palenzuela. 2003. "Definition of Price Stability, Range and Point Inflation Targets: The Anchoring of Long-Term Inflation Expectations." Working Paper Series 273, European Central Bank.

Cecchetti, Stephen, Peter Hooper, Bruce Kasman, Kermit Schoenholtz, and Mark Watson. 2007. *Understanding the Evolving Inflation Process*. U.S. Monetary Policy Forum 2007.

Christiano, Lawrence, Martin Eichenbaum, and Charles Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113 (1):1–45.

Christiano, Lawrence J. and Terry J. Fitzgerald. 2003. "The Bandpass Filter." *International Economic Review* 44 (2):435–465.

Clarida, Richard, Jordi Galí, and Mark Gertler. 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *The Quarterly Journal of Economics* 115 (1):147–180.

Clark, Todd E. and Taeyoung Doh. 2011. "A Bayesian Evaluation of Alternative Models of Trend Inflation." *Federal Reserve Bank of Cleveland, Working Paper no. 11-34* (11-34).

Clark, Todd E. and Michael W. McCracken. 2010. "Averaging Forecasts from VARs with Uncertain Instabilities." *Journal of Applied Econometrics* 25 (1):5–29.

Cochrane, John H. 2011. "Determinacy and Identification with Taylor Rules." *Journal of Political Economy* 119 (3):565–615.

Coibion, Olivier. 2010a. "Testing the Sticky Information Phillips Curve." *The Review of Economics and Statistics* 92 (1):87–101.

———. 2010b. "Testing the Sticky Information Phillips Curve." *The Review of Economics and Statistics* 92 (1):87–101.

Coibion, Olivier and Yuriy Gorodnichenko. 2010. “Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts.” Mimeo.

———. 2011. “Monetary Policy, Trend Inflation, and the Great Moderation: An Alternative Interpretation.” *American Economic Review* 101 (1):341–370.

———. 2012. “What Can Survey Forecasts Tell Us About Informational Rigidities?” *Journal of Political Economy* 120 (1):116–159.

Dale, Spencer. 2012. “Sticky Inflation.” URL http://www.bankofengland.co.uk/publications/Documents/speeches/2012/speech622.pdf. Speech given by Spencer Dale, Executive Director, Monetary Policy, and Chief Economist, Bank of England, 12.12.2012.

Davig, Troy and Eric M. Leeper. 2007. “Generalizing the Taylor Principle.” *American Economic Review* 97 (3):607–635.

Diebold, Francis X. and Robert S. Mariano. 1995. “Comparing Predictive Accuracy.” *Journal of Business & Economic Statistics* 13 (3):134–144.

Doepke, Joerg, Jonas Dovern, Ulrich Fritsche, and Jiri Slacalek. 2008. “Sticky Information Phillips Curves: European Evidence.” *Journal of Money, Credit and Banking* 40 (7):1513–1519.

Doornik, Jurgen A. 2007. *Object-Oriented Matrix Programming Using Ox*. 3rd ed. London: Timberlake Consultants.

Dotsey, Michael, Shigeru Fujita, and Tom Stark. 2011. “Do Phillips Curves Conditionally Help to Forecast Inflation?” Mimeo.

Erceg, Christopher, Dale Henderson, and Andrew Levin. 2000. “Optimal Monetary Policy with Staggered Wage and Price Contracts.” *Journal of Monetary Economics* 46 (2):281–313.

Fagan, Gabriel, Jérôme Hendry, and Ricardo Mestre. 2005. “An Area-Wide Model for the Euro Area.” *Economic Modelling* 22 (1):39–59.

Faust, Jon and Jonathan H. Wright. 2012. *Forecasting Inflation*, *Handbook of Economic Forecasting*, vol. 2. Elsevier. Forthcoming.

Favero, Carlo A., Massimiliano Marcellino, and Francesca Neglia. 2005. "Principal Components at Work: The Empirical Analysis of Monetary Policy with Large Data Sets." *Journal of Applied Econometrics* 20 (5):603–620.

Fisher, Jonas D. M., Chin Te Liu, and Ruilin Zhou. 2002. "When Can We Forecast Inflation?" *Economic Perspectives - Federal Reserve Bank of Chicago* 26 (1):30–42.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2000. "The Generalized Dynamic-Factor Model: Identification And Estimation." *The Review of Economics and Statistics* 82 (4):540–554.

Fuhrer, Jeff and George Moore. 1995. "Inflation Persistence." *Quarterly Journal of Economics* 110 (1):127–159.

Fuhrer, Jeffrey C. and Giovanni P. Olivei. 2010. "The Role of Expectations and Output in the Inflation Process: An Empirical Assessment." Public Policy Brief No. 10-2, Federal Reserve Bank of Boston.

Fuhrer, Jeffrey C., Giovanni P. Olivei, and Geoffrey M.B. Tootell. 2009. "Empirical Estimates of Changing Inflation Dynamics." Mimeo.

———. 2011. "Inflation Dynamics When Inflation is Near Zero." Mimeo.

Galí, Jordi and Mark Gertler. 1999. "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44 (2):195–222.

Galí, Jordi, Mark Gertler, and J. David López-Salido. 2001. "European Inflation Dynamics." *European Economic Review* 45 (7):1237–1270.

Galí, Jordi, Mark Gertler, and J. David López-Salido. 2005. "Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve." *Journal of Monetary Economics* 52 (6):1107–1118.

Garratt, Anthony, James Mitchell, and Shaun Vahey. 2010. "Measuring Output Gap Uncertainty." CEPR Discussion Papers 7742.

Giannone, Domenico, Michele Lenza, Daphne Momferatou, and Luca Onorante. 2010. "Short-Term Inflation Projections: A Bayesian Vector-Autoregressive Approach." *International Journal of Forecasting* Forthcoming.

Gordon, Robert J. 1982. "Inflation, Flexible Exchange Rates, and the Natural Rate of Unemployment." In *Workers, Jobs and Inflation*, edited by Martin N. Baily. Washington D.C.: The Brookings Institution, 89–158.

———. 1990. "U.S. Inflation, Labor's Share, and the Natural Rate of Unemployment." In *Economics of Wage Determination*, edited by Heinz Konig. Berlin: Springer Verlag.

Hubrich, Kirstin. 2005. "Forecasting Euro Area Inflation: Does Aggregating Forecasts by HICP Component Improve Forecast Accuracy?" *International Journal of Forecasting* 21 (1):119–136.

IMF. 2013. "The Dog That Didn't Bark: Has Inflation Been Muzzled or Was It Just Sleeping." In *IMF World Economic Outlook, April 2013*. IMF.

Inoue, Atsushi and Barbara Rossi. 2011. "Identifying the Sources of Instabilities in Macroeconomic Fluctuations." *The Review of Economics and Statistics* 93 (4):1186–1204.

Kapetanios, George, Lynda Khalaf, and Massimiliano Marcellino. 2011. "Factor based identification-robust inference in IV regressions." Mimeo.

Kapetanios, George, Vincent Labhard, and Simon Price. 2008. "Forecasting Using Bayesian and Information-Theoretic Model Averaging: An Application to U.K. Inflation." *Journal of Business & Economic Statistics* 26 (1):33–41.

Kapetanios, George and Massimiliano Marcellino. 2010. "Factor-GMM estimation withlarge sets of possibly weak instruments." *Computational Statistics and Data Analysis* 54 (11):2655–2675.

Kleibergen, Frank. 2005. "Testing Parameters in GMM Without Assuming That They are Identified." *Econometrica* 73 (4):1103–1123.

Kleibergen, Frank and Sophocles Mavroeidis. 2009. "Weak Instrument Robust Test in GMM and the New Keynesian Phillips Curve." *Journal of Business and Economic Statistics* 27 (3):293–311.

Koop, Gary and Dimitris Korobilis. 2012. "Forecasting Inflation Using Dynamic Model Averaging." *International Economic Review* 53 (3):867–886.

Lanne, Markku, Arto Luoma, and Jani Luoto. 2009. "A Naïve Sticky Information Model of Households' Inflation Expectations." *Journal of Economic Dynamics and Control* 33 (6):1332–1344.

Lubik, Thomas A. and Frank Schorfheide. 2004. "Testing for Indeterminacy: An Application to U.S. Monetary Policy." *American Economic Review* 94 (1):190–217.

Ma, Adrian. 2002. "GMM Estimation of the Phillips Curve." *Economics Letters* 76 (3):411–417.

Mackowiak, Bartosz and Mirko Wiederholt. 2009. "Optimal Sticky Prices Under Rational Inattention." *American Economic Review* 99 (3):769–803.

Mankiw, N. Gregory and Ricardo Reis. 2002. "Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Quarterly Journal of Economics* 117 (4):1295–1328.

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers. 2003. "Disagreement About Inflation Expectations." In *NBER Macroeconomics Annual*, edited by Mark Gertler and Kenneth Rogoff. Cambridge: Cambridge University Press.

Marcellino, Massimiliano and Alberto Musso. 2010. "The Forecasting Performance of Real Time Estimates of the Euro Area Output Gap." CEPR Discussion Papers 7763.

Mavroeidis, Sophocles. 2004. "Weak Identification of Forward-looking Models in Monetary Economics." *Oxford Bulletin of Economics and Statistics* 66 (1):609–635.

———. 2005. "Identification Issues in Forward-Looking Models Estimated by GMM with an Application to the Phillips Curve." *Journal of Money, Credit and Banking* 37 (3):421–449.

———. 2010. "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence." *American Economic Review* 100 (1):491–503.

Mirza, Harun and Lidia Storjohann. 2013. "Making Weak Instrument Sets Stronger: Factor-Based Estimation of Inflation Dynamics and a Monetary Policy Rule." *Journal of Money, Credit and Banking* Forthcoming.

Montoya, Lourdes Acedo and Björn Döhring. 2011. "The Improbable Renaissance of the Phillips Curve: The Crisis and Euro Area Inflation Dynamics." European Economy. Economic Papers 446, European Commission.

Moscarini, Giuseppe. 2004. "Limited Information Capacity as a Source of Inertia." *Journal of Economic Dynamics and Control* 28 (10):2003–2035.

Musso, Alberto, Livio Stracca, and Dick van Dijk. 2009. "Instability and Nonlinearity in the Euro Area Phillips Curve." *International Journal of Central Banking* 5 (2):181–212.

Newey, Whitney K. and Frank Windmeijer. 2009. "Generalized Method of Moments With Many Weak Moment Conditions." *Econometrica* 77 (3):687–719.

Nunes, Ricardo. 2009. "On the Epidemiological Microfoundations of Sticky Information." *Oxford Bulletin of Economics and Statistics* 71 (5):643–657.

———. 2010. "Inflation Dynamics: The Role of Expectations." *Journal of Money, Credit and Banking* 42 (6):1161–1172.

Onatski, Alexei. 2009. "Testing Hypotheses About the Number of Factors in Large Factor Models." *Econometrica* 77 (5):1447–1479.

O'Reilly, Gerard and Karl Whelan. 2005. "Has Euro Area Inflation Persistence Changed over Time." *Review of Economics and Statistics* 87 (4):709–720.

Orphanides, Athanasios. 2001. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review* 91 (4):964–985.

———. 2004. "Monetary Policy Rules, Macroeconomic Stability, and Inflation: A View from the Trenches." *Journal of Money, Credit and Banking* 36 (2):151–175.

Pagan, Adrian Rodney. 2009. "Comments on "Phillips Curve Inflation Forecasts" by James H. Stock and Mark W. Watson." In *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective*, edited by Jeffrey C. Fuhrer. The MIT Press, 187–193.

Paloviita, Maritta. 2008. "Comparing Alternative Phillips Curve Specifications: European Results With Survey-Based Expectations." *Applied Economics* 40 (17):2259–2270.

Pesaran, Hashem and Allan Timmermann. 2007. "Selection of Estimation Window in the Presence of Breaks." *Journal of Econometrics* 137 (1):134–161.

Pesaran, M. Hashem and Martin Weale. 2006. *Survey Expectations*, *Handbook of Economic Forecasting*, vol. 1. Amsterdam: Elsevier.

Pfajfar, Damjan and Emiliano Santoro. 2010. "Heterogeneity, Learning and Information Stickiness in Inflation Expectations." *Journal of Economic Behavior and Organization* 75 (3):426–444.

Reis, Ricardo. 2009. "Optimal Monetary Policy Rules in an Estimated Sticky-Information Model." *American Economic Journal: Macroeconomics* 1 (2):1–28.

Roberts, John M. 1995. "New Keynesian Economics and the Phillips Curve." *Journal of Money, Credit and Banking* 27 (4):975–984.

———. 1997. "Is Inflation Sticky?" *Journal of Monetary Economics* 39 (2):173–196.

Rudd, Jeremy and Karl Whelan. 2005. "New Tests of the New Keynesian Phillips Curve." *Journal of Monetary Economics* 52 (6):1167–1181.

Rünstler, Gerhard. 2002. "The Information Content of Real-Time Output Gap Estimates: An Application to the Euro Area." Working Paper Series 182, European Central Bank.

Sbordone, Argia M. 2002. "Prices and Unit Labor Costs: A New Test of Price Stickiness." *Journal of Monetary Economics* 49 (2):265–292.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3):665–690.

Steinsson, Jon. 2003. "Optimal Monetary Policy in an Economy with Inflation Persistence." *Journal of Monetary Economics* 50 (7):1425–1456.

Stock, James H. and Mark M. Watson. 1999. "Forecasting Inflation." *Journal of Monetary Economics* 44 (2):293–335.

———. 2007a. "Why has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit, and Banking* 39 (7):3–34.

———. 2009. "Phillips Curve Inflation Forecasts." In *Understanding Inflation and the Implications for Monetary Policy, a Phillips Curve Retrospective*, edited by J. Fuhrer, Y.K. Kodrzycki, J. Sneddon Little, and G.P. Olivei. The MIT Press, 101–186.

———. 2010. "Modelling Inflation After the Crisis." Mimeo.

Stock, James H. and Mark W. Watson. 2002. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association* 97 (460):1167–1179.

———. 2007b. *Introduction to Econometrics.* Boston, MA: Addison Wesley.

———. 2008. "Forecasting in Dynamic Factor Models Subject to Structural Instability." In *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*, edited by Jennifer Castle and Neil Shephard. Oxford: Oxford University Press.

Stock, James H. and Jonathan H. Wright. 2000. “GMM with Weak Identification.” *Econometrica* 68 (5):1055–1096.

Stock, James H. and Motohiro Yogo. 2005. “Testing for Weak Instruments in Linear IV Regression.” In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,*, edited by D.W.K. Andrews and J.H. Stock. Cambridge, UK: Cambridge University Press.

Taylor, John B. 1999. *Monetary Policy Rules.* Chicago: University of Chicago Press.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton, NJ: Princeton University Press.

Wright, Jonathan. 2009. “Forecasting US Inflation by Bayesian Model Averaging.” *Journal of Forecasting* 28 (2):131–144.

Wright, Jonathan H. 2012. “Evaluating Real-Time VAR Forecasts with an Informative Democratic Prior.” *Journal of Applied Econometrics* Forthcoming.