RATIONAL *VS.* PROFESSIONAL FORECASTS*

*João Valle e Azevedo***

ABSTRACT

We compare theoretical and empirical forecasts computed by rational agents living in a model economy to those produced by professional forecasters. We focus on the variance of the prediction errors as a function of the forecast horizon and analyze the speed with which it converges to a constant (which can be seen as a measure of the speed of convergence of the economy to the steady state). We look at a standard sticky-prices-wages model, concluding that it delivers a strong theoretical forecastability of the variables under scrutiny, at odds with the data (professional forecasts). The flexible prices-wages version delivers a forecastability closer to the data and performs relatively better empirically (with actual data), but mainly because forecasts deviate little from the unconditional mean. These results can be interpreted in at least two ways: first, actual deviations from the steady-state are not persistent, in which case the implications of the specific formulation of nominal rigidities for short-run dynamics are unrealistic; second, exogenous (or unmodelled) steady-state shifts attributable to, *e.g.*, changes in monetary-policy, taxation, regulation or in the growth of the technological frontier, occur in such a way as to strongly limit the performance of professional forecasters.

1.Introduction

Despite tremendous efforts over the past decades, macroeconomic forecasting seems as difficult as before. For most variables forecast accuracy is low, naive models prove hard to beat and sophisticated statistical methods provide marginal (if any) improvements at long horizons. This degree of uncertainty should perhaps be considered a feature of the economy, as the same difficulty characterizes professional forecasts (say, from the Philadelphia Survey of Professional Forecasters, henceforth Phil-SPF, the Federal Reserve Green Book, Fed Green-Book, or the European Central Bank Survey of Professional Forecasters, ECB-SPF). Still, there is clear evidence they rank very well in comparison with various statistical methods, being less prone to the structural instabilities of macroeconomic time series (*e.g.*, Faust and Wright, 2007 and Bernanke and Boivin, 2003 for evidence on Green-Book forecasts or Ang, Bakeart and Wei, 2007 for Phil-SPF inflation forecasts). Moreover, they can be seen as a fortunate aggregation of various individual forecasts that probably adapt fast to changes in the economy, each using different data, different methods and even some judgment. The question we address is whether the behavior of these forecasts shares characteristics of theoretical and empirical forecasts produced by typical dynamic stochastic general equilibrium (DSGE) models. We will call the first set of forecasts Professional" and the latter Rational". We view Professional forecasts as the best publicly available proxy for the forecasts produced by well

* This article is a summary of research done in collaboration with João Tovar Jalles, University of Cambridge. We would like to thank the comments of Nuno Alves, Mário Centeno and Ana Cristina Leal. The opinions hereby expressed are those of the author and do not reflect the vision of Banco de Portugal or the Eurosystem. Any errors and omissions are the sole responsibility of the author.

** Banco de Portugal, Economics and Research Department.

informed agents in the economy, providing a natural benchmark against which to confront the forecasts produced by rational agents living in a model economy that is taken seriously. We assess the fit of models to data by analyzing the differences in the relative performance of Rational vs Professional forecasts, paying special attention to the variance of the forecast errors as a function of the forecast horizon as well as to the speed with which it converges to a constant (the speed with which the forecasts converge to the unconditional mean of the variables under scrutiny). We view this exercise as useful to understand in which dimensions (or for which variables) theoretical models provide a reasonable description of the process (or speed) of convergence of the economy to the steady-state. We will interpret small differences in forecast accuracy (as a function of the forecast horizon) in the two worlds (Rational and Professional) as sign that the model is able to replicate an important dimension of actual data.

Our benchmark model economy will be the one discussed in Smets and Wouters (2007) in its New-Keynesian (NK) and Real Business Cycle (RBC) versions (*i.e.*, with and without nominal rigidities). The comparison of the two paradigms is instructive. In fact, nominal rigidities (along with real features adding persistence such as habit formation and investment adjustment costs) are often incorporated with the justification that they enable the models to replicate the persistence of the response to various shocks identified with vector autoregressions, *i.e.*, impulse response functions that are still alive" after two or even three years horizons. This translates into theoretical forecastability of the corresponding variables at very long horizons. However, we conclude that for the growth of real variables such as consumption, investment or output, Professional forecasters don't do better than the mean at horizons greater than 3 or 4 quarters. The notable exception is the unemployment rate (which we take as a proxy for hours worked in the theoretical model). For nominal interest rates both forecasters and agents that know the economy (in a sticky price model) can form forecasts that are superior to the mean at very long horizons (which simple time series models also can due to the very high persistence of nominal rates). For inflation, forecasters still add to the mean after 5 quarters, but little, whereas the standard sticky-price-wage model (under standard parametrizations) is still far from the mean of the process at long horizons, clearly at odds with the data. The RBC version is silent with respect to nominal rates and inflation but delivers forecasts of real variables that are closer (in terms of relative performance with respect to the unconditional mean and speed of convergence) to Professional forecasts. The notable exception relates to hours. Once we use these models to forecast actual data, the performance is extremely weak but less so with the RBC version (mainly because forecasts converge to the unconditional mean more rapidly). Again, the exception to this pattern is found with hours/unemployment.

The remainder of the article is organized as follows: Section 2 shows that, for a host of variables, the predictive power of Professional forecasts vanishes fast as the forecast horizon increases, *i.e.*, the gains (if any) that one obtains by using these forecasts instead of a (real-time) estimate of the unconditional mean of the variables are small. Section 3 confronts these facts with the theoretical and empirical performance of a standard DSGE model, whereas section 4 pays special attention to what occurs during recessions. Section 5 concludes.

2. Professional Forecasts: how much they deliver

Here we assess the predictive power of U.S. Professional forecasts, measuring simply their performance relative to an estimate of the unconditional mean of the variables analyzed.[1] In this way we investigate how informative these forecasts are and until when (in terms of forecast horizon) they provide relevant signal relative to what can be viewed as a steady-state forecast. It is still early to conduct a similar and conclusive analysis with euro area data due to sample size restrictions while for the purposes of our study the origin of the data is not relevant.

1 Analysis of the forecast performance of Phil-SPF is routinely conducted at the Federal Reserve Bank of Philadelphia, see Stark (2010) for a recent discussion.

2.1. Data

We analyze 15 macroeconomic indicators from the Phil-SPF,[2] namely: Nominal output measured by GNP/GDP (NOUTPUT), Real GNP/GDP (ROUTPUT), Industrial Production Index - Total (IPT), Real Personal Consumption Expenditures - Total (RCON), Net Exports (NETEXP), GDP deflator (GDPDEF), Consumer Price Index (CPI), Real Gross Private Domestic Investment -- Residential (RINVRESID), Real Gross Private Domestic Investment -- Nonresidential (RINVBF), Real Government Consumption and Gross Investment -- State and Local (RGLS), Real Government Consumption and Gross Investment -- Federal (RGF), Housing Starts (HSTARTS), unemployment (UNRATE), 3 -month T-bill rate (TB3MS) and 10 year T-bond rate (GS10). All data is firstly aggregated quarterly when necessary (to be consistent with the variables forecasted in the Phil-SPF) and except for unemployment and interest rates, all data is in growth rates. We look only at point forecasts and define these as the median forecast (across all respondents) in every release of the survey (results with the mean forecast are very similar). The individual respondent's point forecast are generally close to the central tendencies of their subjective distributions (*e.g.*, Engelberg *et al*. 2009) while there is clear evidence that this aggregation produces forecasts that are in general superior to individual forecasts. Obviously, a not so straightforward aggregation can result in forecast improvements, and this can be achieved even when there is (as in Phil-SPF) entry and exit of forecasters, see Capistrán and Timmerman (2009).

Fed Green-Book forecasts will not be analyzed here, please refer to the working paper version of this article, but we can refer that Phil-SPF forecasts represent best practice, or close to best practice, within professional forecasts. In any case, we should refer that Romer and Romer (2000), using data until 1991, have shown that Fed-Green Book forecasts of inflation and real GDP are statistically unbiased and dominate private sector forecasts (*i.e.*, suggesting that the Federal Reserve had considerably more information beyond what is known to the private sector). The period of Great Moderation"[3] between 1982-2007 has affected the time-series properties of many variables as well as the performance of Professional forecasts. In particular, D'Agostino and Whelan (2008) show that the superior forecast performance of the Fed-Green Book deteriorated considerably after 1991, with Phil-SPF forecasts taking the lead. We have reached similar conclusions. In the remainder of the paper we will thus focus exclusively on Phil-SPF forecasts, regarding them as a proxy of the best forecasts produced by well-informed agents in the economy.

2.2. Methodology

We begin our analysis by taking the real-time vintage data from 1968q4 through 1981q3 - h quarters, for $h = 1, \dots, 5$.[4] We then estimate the unconditional mean of the variables under scrutiny by simply computing the average of each variable for this vintage, which is our benchmark forecast for 1981q3. As a reference, we also compute forecasts from an estimated Direct Autoregression (AR) using data through 1981q3 - h. We repeat this procedure using the vintage from 1968q4 through 1981q4 - h, $h = 1, \dots, 5$ and so forth until 2009q2. It should be noted that most variables are available with a delay of one quarter. Hence, to properly compare these benchmark forecasts with Phil-SPF's and Fed-Green Book's forecasts we re-label the forecast horizon so that the information sets with each method approximately coincide (so, the h step ahead, with $0 \leq h \leq 4$, Phil-SPF's and Green-Book's forecasts will be considered as $h+1$ step ahead forecasts since the latest observation of the variable to be forecasted does not in general

2 For complete information on the survey's background see http://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/spf-documentation.pdf as well as Zarnowitz (1969), Zarnowitz and Braun (1992) and Croushore (1993).

3 See, *e.g.*, McConnell and Perez-Quiros (2000), Stock and Watson (2003) and Giannone, Lenza and Reichlin (2008).

4 These series were retrieved from the Philadelphia Fed website. http://www.phil.frb.org/research-and-data/real--time-center/real-time-data/data-files/. See, *e.g.*, Croushore and Stark (2001, 2003) for a discussion of real-time data.

refer to the forecast moment,[5] which is approximately the middle of the quarter since at least 1990q3[6]). We then compare the forecast accuracy of the different surveys' predictions, AR and also real-time average by computing the ratio of Root Mean Square Forecast Error (RMSFE) of both the AR and Professional forecasts relative to the benchmark forecast (real-time average). It should be noted that the forecast error is defined as the difference between the forecast and the corresponding observation of the latest vintage of data available (results considering the h quarters ahead vintage alter little the results, at least in relative terms, see also Stark 2010 for a thorough analysis of this issue). Following Fair and Shiller (1989) we also run the following forecast encompassing regression:

$$y_{t+h} = \alpha + \beta_0 f_{t+h}^{\text{real}} + \beta_1 f_{t+h}^{x} + \varepsilon_{t+h} \tag{1}$$

where y_{t+h} is the observation of the variable forecasted, f^{real} is the forecast from the real-time average, f^{x} is the forecast from the candidate model x, in our case either the AR process or the Professional forecasts and ε_{t+h} is a regression error. Obviously, if $\beta_1 \neq 0$ then forecasts from candidate model x add information relative to the "real-time average". Both the RMSFE ratios and the β_1 coefficients are computed and presented for the full sample as well as for an aggregation of recession quarters as defined by the National Bureau of Economic Research (NBER).

It is important to note that a forecast performing as well as an estimate of the unconditional mean in terms of RMSFE (or encompassed by it, in which case $\beta_1 = 0$) may nonetheless be useful if it can more often accurately predict the direction of change in the actual series (Joutz and Stekler, 2000). With this in mind we will examine the sign forecast accuracy of the forecasts by constructing the following two-by-two contingency table in which the actual and forecast data for each quarter are classified (i) by whether the actual change in a given variable is positive (+) or negative/zero (-,0), and (ii) whether the forecast correctly or incorrectly predicted the sign:

Contingency Table	
$n_{11} \; : \; \Delta y_{t+h}(+), \Delta f_{t+h}^{x}(+)$	$n_{12} \; : \; \Delta y_{t+h}(-,0), \Delta f_{t+h}^{x}(+)$
$n_{21} \; : \; \Delta y_{t+h}(+), \Delta f_{t+h}^{x}(-,0)$	$n_{22} \; : \; \Delta y_{t+h}(-,0), \Delta f_{t+h}^{x}(-,0)$

where the actual change is $\Delta y_{t+h} = y_{t+h} - y_t$ and the predicted change is $\Delta f_{t+h}^{x} = f_{t+h}^{x} - y_t$. Note that y_t is the most recent (quarterly) value known at the time of the forecast. The main diagonal cells include numbers of correct sign forecasts and the the other cells include the numbers of incorrect sign forecasts. We then test the null hypothesis of no association between the frequency of actual and predicted changes (because correct predicted changes will always occur, what matters is whether their frequency is higher than what would be expected if actual and predicted changes were completely unrelated).

2.3. Results and Discussion

2.3.1. Forecast Accuracy

Our main empirical results regarding forecast accuracy of Phil-SPF forecasts are presented in Table 1, referring to the 15 macroeconomic variables defined before. It contains the ratio of the Root Mean Square Forecast Error (RMSFE) of both the AR and Professional forecasts relative to the benchmark forecast (real-time average) as well as the estimate of β_1 resulting from OLS estimation of Eq. (1) at different forecast horizons. Results are presented for the full sample and for an aggregation of recession periods.

5 This does not apply, *e.g.*, to interest rates, whose quarterly average is to be forecasted but are obviously available in the middle of the quarter, when the forecast is made.

6 The timing of the previous American Statistical Association/NBER survey (that was taken by the Philadelphia FED) is not known exactly but it is believed that it followed closely Phil-SPF's.

Table 1 (to be continued)

FORECAST PERFORMANCE OF THE PHILADELPHIA'S SURVEY OF PROFESSIONAL FORECASTERS

Espec.	Real Time Average (RMSFE) (1)	Rel. RMSFE AR (2)	β_1 AR OLS (3)	Rel. RMSFE SPF (4)	β_1 SPF OLS (5)	Rel. RMSFE AR (6)	β_1 AR OLS (7)	Rel. RMSFE SPF (8)	β_1 SPF-OLS (9)	Rel. RMSFE AR (10)	β_1 AR OLS (11)	Rel. RMSFE SPF (12)	β_1 SPF - OLS (13)	Rel. RMSFE AR (14)	β_1 AR OLS (15)	Rel. RMSFE SPF (16)	β_1 SPF - OLS (17)	Rel. RMSFE AR (18)	β_1 AR OLS (19)	Rel. RMSFE SPF (20)	β_1 SPF-OLS (21)
Time/H		1				2				3					4				5		
NOMINAL GNP/GDP (NOUTPUT)																					
Full	2.93	1.06	0.04	0.68	0.11**	1.13	0.01	0.81	0.11*	1.18	0.03	0.93	0.05	1.19	0.02	0.99	0.03	1.21	-0.004	1.07	0.02
Rec.	3.29	1.48	0.1	0.79	0.12	1.55	0.22	1.05	0.09	1.64	0.09	1.30	0.02	1.62	0.32	1.42	0.01	1.66	0.41*	1.56	0.02
PNB/PIB REAL (ROUTPUT)																					
Full	3.02	0.86	0.03	0.66	0.26***	0.89	0.07	0.78	0.21*	0.94	0.15	0.91	0.01	0.96	0.38	0.96	-0.09	1.01	0.13	0.98	-0.03
Rec.	3.05	1.30	-0.01	0.72	0.11	1.36	0.11	1.05	-0.15	1.45	0.09	1.33	-0.31	1.48	0.88*	1.46	-0.34*	1.52	0.58	1.49	0.31*
INDUSTRIAL PRODUCTION INDEX – TOTAL (IPT)																					
Full	4.92	0.76	0.75***	0.65	0.49***	0.94	1.85***	0.91	0.23	0.98	1.71***	1.07	-0.12	1.16	0.2	1.13	-0.23	1.18	-0.18	1.18	-0.27
Rec.	5.84	0.92	0.73***	0.76	0.52***	1.24	0.56***	1.19	0.13	1.30	0.52***	1.48	-0.2	1.59	0.36	1.59	-0.25	1.59	-0.24	1.66	-0.23
REAL PERSONAL CONSUMPTION EXPENDITURES – TOTAL (RCON)																					
Full	2.87	0.75	0.35*	0.68	0.42***	0.76	0.33	0.74	0.57**	0.77	0.35	0.78	0.35	0.8	0.69	0.84	-0.24	0.86	-0.15	0.83	0.34
Rec.	2.02	1.25	0.37	0.97	0.29*	1.28	0.38	1.10	0.45**	1.33	0.5	1.18	-0.19	1.49	0.91**	1.42	-0.45	1.58	-0.001	1.42	0.23
CONSUMER PRICE INDEX (CPI)																					
Full	2.13	1.41	-0.28	0.56	1.40***	1.44	-0.38	0.78	0.78***	1.46	-0.35	0.85	0.56***	1.50	-0.24	0.88	0.42***	1.56	-0.15	0.89	0.39**
Rec.	3.17	1.24	-0.06	0.51	0.89***	1.28	-0.22	0.74	0.52**	1.30	-0.11	0.81	0.46**	1.34	-0.07	0.85	0.43**	1.35	-0.07	0.86	0.41**
REAL GROSS PRIVATE DOMESTIC INVESTMENT – RESIDENTIAL (RINVRESID)																					
Full	19.76	0.6	0.19*	0.61	0.21***	0.73	0.41***	0.76	-0.02	0.81	0.59***	0.88	-0.05	0.86	0.11	0.95	-0.12	0.87	0.09	0.97	-0.17*
Rec.	20.27	0.55	0.09	0.42	0.04	0.79	0.29	0.76	-0.12	0.91	0.52	1.02	-0.03	0.95	0.06	1.17	-0.08	0.95	0.15	1.19	-0.07
REAL GROSS PRIVATE DOMESTIC INVESTMENT – NONRESIDENTIAL (RINVBF)																					
Full	10.20	0.83	0.20***	0.65	0.19*	0.84	0.23**	0.73	0.18	0.91	0.22	0.79	0.2	0.99	0.08	0.9	0.15	1.05	-0.18	0.94	0.16
Rec.	11.61	1.02	0.24*	0.71	-0.02	1.07	0.32	0.78	-0.26	1.16	0.54	0.88	-0.26	1.22	0.65	1.00	-0.24	1.30	0.11	1.08	0.17
REAL GOVERNMENT CONSUMPTION AND GROSS INVESTMENT – STATE AND LOCAL (RGLS)																					
Full	3.02	0.83	0.31*	0.82	0.19***	0.84	0.16	0.82	0.35***	0.82	0.11	0.83	0.24*	0.82	0.36**	0.85	0.48***	0.83	0.49**	1.01	0.06
Rec.	2.61	1.08	0.73***	1.08	0.001	1.05	0.41**	1.04	-0.06	1.04	0.14	1.03	-0.06	1.03	0.43**	0.98	0.19	1.05	0.56**	0.99	-0.01

Table 1 (continued)

FORECAST PERFORMANCE OF THE PHILADELPHIA'S SURVEY OF PROFESSIONAL FORECASTERS

Espec.	Real Time Average (RMSFE)	Rel. RMSFE AR	β_1 AR OLS	Rel. RMSFE SPF	β_1 SPF OLS	Rel. RMSFE AR	β_1 AR OLS	Rel. RMSFE SPF	β_1 SPF-OLS	Rel. RMSFE AR	β_1 AR OLS	Rel. RMSFE SPF	β_1 SPF - OLS	Rel. RMSFE AR	β_1 AR OLS	Rel. RMSFE SPF	β_1 SPF - OLS	Rel. RMSFE AR	β_1 AR OLS	Rel. RMSFE SPF	β_1 SPF-OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
Time/H		1				2				3				4				5			
REAL GOVERNMENT CONSUMPTION AND GROSS INVESTMENT – FEDERAL (RGF)																					
Full	7.22	1.03	-0.03	0.95	-0.11**	1.02	0.08	0.96	-0.01	1.08	0.06	0.98	0.08	1.06	0.01	1.01	0.11	1.01	0.06	0.96	0.06
Rec.	7.52	1	-0.002	0.89	0.17***	0.86	0.09	0.91	-0.13	1.02	-0.01	0.93	-0.04	1.04	0.02	0.98	-0.08	0.92	0.16	0.97	-0.07
HOUSING STARTS (HSTARTS)																					
Full	29.32	1.02	0.39	0.84	1.02***	1.04	0.04	1.03	0.72**	1.10	-0.92	1.11	0.42	1.07	0.03	1.18	-0.14	1.05	0.83	1.16	-0.29
Rec.	34.00	0.88	0.39	0.59	1.55***	0.96	-0.79	1.21	0.61	1.05	-7.44**	1.37	0.56	0.97	-0.19	1.46	0.4	0.94	0.77	1.37	0.75
GROSS DOMESTIC PRODUCT: IMPLICIT PRICE DEFLATOR (GDPDEF)																					
Full	1.92	1.43	0.22	0.67	0.49***	1.44	0.14	0.7	0.58***	1.44	0.08	0.72	0.63***	1.53	0.07	0.75	0.65***	1.59	0.07	0.77	0.53***
Rec.	2.15	1.56	0.04	0.72	0.44*	1.58	-0.26	0.73	0.52**	1.59	-0.14	0.75	0.67**	1.68	0.03	0.77	0.63**	1.67	-0.04	0.76	0.55**
CIVILIAN UNEMPLOYMENT RATE (UNRATE)																					
Full	1.71	0.21	-0.03	0.2	-0.06*	0.22	0.02	0.22	-0.09**	0.25	0.03	0.24	0.09	0.33	0.04	0.3	0.12*	0.41	0.04*	0.36	0.11**
Rec.	2.23	0.26	-0.01	0.24	-0.14*	0.28	0.11	0.25	-0.15**	0.31	0.14*	0.28	0.26*	0.37	0.13**	0.3	0.25***	0.39	0.12*	0.37	0.22***
NET EXPORTS OF GOODS AND SERVICES (NETEXP)																					
Full	55.59	0.89	-0.04	2.32	0.02***	0.46	-0.22	2.00	-0.02	1.21	-0.07	1.98	-0.01	0.99	-0.06	1.77	-0.02	1.19	-0.05	1.68	-0.02
Rec.	111.91	0.12	-0.02	0.29	0.42	0.07	-0.24	0.27	0.56	0.09	-0.32	0.28	0.47	0.1	-0.41	0.28	0.15	0.14	-0.49	0.29	0.01
10-YEAR TREASURY CONSTANT MATURITY RATE (GS10)																					
Full	1.98	0.14	0.89***	0.38	0.69***	0.16	0.81	0.33	1.04***	0.19	0.71***	0.32	1.08***	0.22	0.69***	0.32	1.08***	0.26	0.64***	0.33	1.06***
Rec.	2.58	0.07	0.73***	0.25	-0.08	0.11	0.72***	0.23	0.93***	0.16	0.56***	0.21	1.03***	0.19	0.49***	0.21	1.04***	0.24	0.44***	0.21	1.08***
3-MONTH TREASURY BILL (TB3MS)																					
Full	2.01	0.32	0.96***	0.09	0.99***	0.33	1.07***	0.22	0.99***	0.37	1.07***	0.25	1.02***	0.46	1.05***	0.28	1.08***	0.53	1.24***	0.33	1.15***
Rec.	2.83	0.36	0.87***	0.09	1.05***	0.33	1.00	0.28	1.02***	0.32	0.99***	0.28	1.06***	0.35	0.98***	0.26	1.14***	0.40	1.15***	0.27	1.24***

Source: Author's calculations.

Notas: The table comprises the forecasting performance results of 15 macroeconomic variables referred in the text in which the first column indicates whether one is taking the full-sample period (1981:3-2009:2) ("Full") or the aggregation of recession periods ("Rec") – simple combination of the 4 main recessions as identified by the NBER. Variables names are in accordance to the Philadelphia Fed's designation. The remaining of the table is organized as follows: row 2 identifies the specification; row 3 indicates the forecast horizon ranging from $h=1$ a $h=5$. Column 1 reports the Root Mean Squared Forecast Error (RMSFE) associated to using the real-time average as forecast for each vintage available at the forecast date (available at http://www.phil.frb.org/research-and-data/real-time-center/real-time-data/data-files/). Columns 2,6,10,14 and 18 report RMSFE obtained from an Autoregressive (AR) process relative to the real-time RMSFE – a value larger than one indicates that the AR's RMSE is larger than the real-time RMSFE and this is emphasized by a grey shade throughout the table's entries. Columns 4,8,12,16 and 20 report the relative RMSFE obtained from using the Survey of Professional Forecasters' (SPF) predictions to the Real Time RMSFE – a value larger than one indicates that the SPF's RMSFE is larger than the real-time RMSFE and this is emphasized by a grey shade throughout the table's entries. Columns 3,7,11,15 and 19 (5,9,13,17 and 21) report the coefficient β_1 resulting from OLS estimation of Equation (1) for forecast accuracy at different time horizons for the comparison of real-time and the AR process (SPF); *,**,*** denote significance at 10,5 and 1% levels, respectively (*i.e.*, rejection of the null hypothesis that the coefficient β_1 is zero).

The main conclusions follow:

- considering the full sample, Phil-SPF forecasts add signal relative to the benchmark (real-time average) only up to $h = 2$ when looking at the significance of the β_1 coefficients. The exceptions are Phil-SPF's CPI inflation, unemployment and interest rates predictions throughout the different forecast horizons and Phil-SPF's RGLS (State and Local Government Consumption and Gross investment growth) up to $h = 4$.

- considering the full sample, the relative (to the real-time average) RMSFE for Phil-SPF's forecasts is clearly less than one for all horizons only in the case of unemployment, interest rates and, in a lesser extent, inflation (CPI and GDP deflator). In the case of 10 year bond interest rates the AR outperforms Phil-SPF whereas for the 3-month T bill rate the opposite is true. For output (nominal and real) and specially industrial production, housing starts and net exports this ratio indicates mostly useless Phil-SPF forecasts at horizons greater than or equal to $h = 2$. For consumption, investment (residential and non-residential) and Government expenditures (federal and local) there is still some superiority on average (relative to the real-time average) at horizons $h = 3, 4, 5$. In these cases however, it would in general suffice to use a simple autoregression as the rel. RMSFE compares favorably with Phil-SPF's.

- for all variables except (again) interest rates, inflation and unemployment, Phil-SPF (and AR) forecasts that correspond to recession periods have a quite poor performance relative to the real-time average except at $h = 1$. Afterwards the rel. RMSFEs are higher than those obtained with the full sample and more frequently well above 1. This evidence is in line with *e.g.* Zarnowitz (1992) , Zarnowitz and Braun (1992), McNees (1992) and McNees and Ries (1983) who reported a number of systematic errors made by forecasters regarding recession periods. For $h = 2$, β_1 is nonetheless still significant for Phil-SPF in the case of consumption and for AR forecasts in the cases of state and local government expenditures, non-residential private investment and industrial production, despite the fact that rel. RMSFE is above 1.

Putting it simply, this exercise shows that for most variables a real-time estimate of the conditional mean is a hard to beat forecast even at short horizons. Regarding unemployment, nominal interest rates and inflation, Professional forecasts do contain relevant information beyond that of our crude benchmark forecast. In these cases, however, it is more clear that the distance between these forecasts and the real-time average forecast is surely overstated, in the sense that the latter is supposed to measure a steady-sate value that may be varying over time (*e.g.*, due to changes in monetary policy or labor market institutions). This is not damaging for our purposes as it allows us to refer to this distance as an upper bound on what a theoretical model (without regime shifts in monetary policy or labor market institutions) ought to deliver in terms of forecast accuracy relative to the steady-state forecast.

2.3.2. Sign forecast accuracy

Table 2 reports the cell counts for the contingency table described in section 2.2 and p-values for the null hypothesis of no association between actual and predicted changes for the Phil-SPF and real-time average forecasts. First, it is clear that most p-values for Phil-SPF forecasts are less than 0.1, or the null hypothesis of no association between actual and predicted changes is rejected, indicating that, in general, these forecasts accurately predict the direction of change in the actual series more often than what luck would determine. What is more interesting for our purposes is to compare the behavior of Professional forecasts to that of the benchmark (real-time average) forecast. First, we observe that for $h = 1, 2$ Phil-SPF's forecasts are, in general, clearly more informative than the real-time average (lower associated p-values), according to this criterion. Second, for real output Phil-SPF's forecasts do not clearly look more useful than the benchmark at $h = 3, 4, 5$. But the main result emerging from Table 2 is that at horizons greater than $h = 2$ and for all other variables except interest rates, CPI inflation, unemployment and to a lesser extent State and local Government spending, the null of no association (no valuable prediction of the direction of change) is either rejected for both Phil-SPF and real-time average forecasts or,

Table 2

SIGN FORECAST ACCURACY TESTS PHIL-SPF | REAL-TIME AVERAGE

H	Variable	p-value	
		Phil-SPF	Real-Time
1	NOUTPUT	0.00	0.03
2		0.00	0.06
3		0.17	0.15
4		0.16	0.11
5		0.08	0.14
1	IPT	0.15	0.09
2		0.08	0.16
3		0.07	0.16
4		0.17	0.14
5		0.17	0.16
1	HSTARTS	0.00	0.05
2		0.00	0.00
3		0.05	0.15
4		0.15	0.16
5		0.03	0.15
1	RCONS	0.00	0.13
2		0.00	0.07
3		0.00	0.15
4		0.02	0.03
5		0.00	0.05
1	RINVBF	0.00	0.08
2		0.00	0.07
3		0.01	0.07
4		0.01	0.06
5		0.00	0.07
1	RINVRESID	0.04	0.14
2		0.01	0.17
3		0.00	0.00
4		0.00	0.00
5		0.00	0.00
1	RGF	0.00	0.18
2		0.00	0.13
3		0.19	0.00
4		0.11	0.00
5		0.19	0.00
1	RGLS	0.00	0.16
2		0.00	0.17
3		0.09	0.08
4		0.05	0.15
5		0.08	0.16
1	UNRATE	0.00	0.15
2		0.00	0.16
3		0.00	0.16
4		0.00	0.16
5		0.03	0.15
1	NETEXP	0.11	0.02
2		0.11	0.03
3		0.15	0.03
4		0.15	0.02
5		0.16	0.02
1	CPI	0.00	0.01
2		0.00	0.00
3		0.06	0.14
4		0.01	0.11
5		0.02	0.12
1	TB3MS	0.00	0.09
2		0.00	0.09
3		0.00	0.09
4		0.00	0.17
5		0.00	0.16
1	GS10	0.00	0.19
2		0.00	0.16
3		0.00	0.16
4		0.00	0.15
5		0.00	0.17
1	GDPDEF	0.01	0.02
2		0.00	0.05
3		0.16	0.03
4		0.12	0.03
5		0.16	0.03
1	ROUTPUT	0.00	0.14
2		0.00	0.15
3		0.11	0.17
4		0.14	0.11
5		0.10	0.17

Sources: Author's calculations.

Notes: P-value (or Fisher's exact test) is for testing the null hypothesis of no association between the direction of change in the actual $\Delta Y_{t+h} = Y_{t+h} - Y_{t-1}$ and forecast $\Delta \hat{Y}_{t+h} = \hat{Y}_{t+h} - Y_{t-1}$ series.

when the null is not rejected for Phil-SPF forecasts, it is often rejected in the case of real-time average forecasts. All in all, the main message is that (with the exceptions mentioned) Professional forecasts certainly loose marginal informational content when compared to the benchmark after 2/3 quarters, in line with the previous subsection.

3.How does a Standard DSGE model forecast?

3.1.The model

We move now towards the core of the article, comparing the results above with the theoretical and empirical forecast performance of the medium-scale model analyzed and estimated in Smets and Wouters (2007) (henceforth SW07), based on Smets and Wouters (2003) and Christiano *et al*. (2005). The model has many of the features now popular in the growing so-called DSGE literature[7] including monopolistic competition in the goods and labour markets, ingredients aimed at improving the fit of the model to observables such as habit formation in consumption, investment adjustment costs, variable capacity utilization (all of these implying amplification of the effects of shocks) and crucially, nominal frictions such as sticky prices and wages along with partial backward-looking indexation. Monetary policy follows a Taylor rule and has real effects when nominal frictions are important. Seven shocks are included (total factor productivity, investment productivity, monetary policy, government spending, risk premium along with price and wage markup shocks) as well as seven observables: output, investment, consumption, wages (all in log differences, or growth rates) as well as inflation, nominal interest rates and (log of) hours. We use exactly the same data treatment as in SW07, implying that the match between the model's variables and Phil-SPF's counterparts is not perfect. Specifically, SW07 observables for output, consumption, investment and wages are expressed in per capita (working age population) terms and nominal interest rates are measured with the Federal funds rate (quite close to the 3-month T-bill rate from Phil-SPF nonetheless). The inflation measure in the model is GDP deflator inflation (*i.e.*, perfect match with Phil-SPF) whereas (minus) Phil-SPF's unemployment, while following closely hours, surely drifts somehow from the concept in the model.

We analyze the forecast performance of two versions of SW07: the original one featuring nominal rigidities, or New-Keynesian (NK) version, and another where we shut down these rigidities (RBC version, where we further reduce the observables by eliminating inflation and nominal interest rates). We use Smets and Wouters's estimated parameters (mode of the posterior distribution, obtained from combining the likelihood function with a set of independent priors for the 41 structural parameters included in the model) using data from 1984q1 through 2004q2. We choose this sample to avoid quibbles regarding the onset of the "Great Moderation" and likely changes in monetary policy within the period starting in 1966q1 (SW07's beginning of the sample). We arguably go against the RBC version by not re-estimating the model, *i.e.*, we keep fixed the structural parameters not related to nominal rigidities. Forecasts of the observables are just conditional expectations given the model and are obtained with the Kalman filter, which is also used to derive the theoretical covariances of the forecast errors for various horizons.[8]

We start with a theoretical analysis of the forecastability of the various variables implied by the model , *i.e.*, we assume the model is the economy and derive analytically the standard deviation of the forecast errors at various horizons. The (artificial) sample size is set at $T = 160$ (thinking in 40 year of post-war quarterly data). Chart 1 presents the theoretical relative (to the standard deviation of the variables) root

7 See, *e.g.*, Adolfson et. al. (2007, 2008) and Christiano *et al*. (2009) for further (and growing) models.

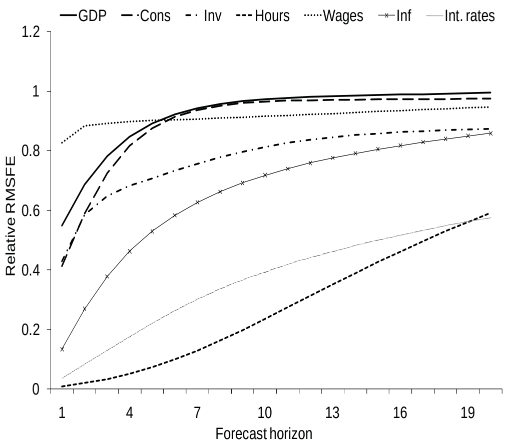
8 For the theoretical analysis this only implies that agents would be using a minimum mean square criterion if they were to pick this as a point forecast, *i.e.*, they know the parameters of the model and produce conditional expectations given the state space model. Regarding the empirical analysis in the paper, it is fair to say that Bayesian estimation of the models would make natural using as point forecasts the mean of the predictive density of future observations, see *e.g.*, Adolfson, Lindé, and Villani (2007).

mean squared forecast error (RMSFE) for output, consumption, investment, inflation, hours, nominal interest rate and wages of the original SW07 (NK version). As easily concluded, for nominal interest rates, inflation but also hours, there is a very strong predictability at short horizons, the relative RMSFE converges slowly and after 20 quarters this ratio is still around 0.4 for inflation and nominal interest rates and 0.7 for hours. For consumption, output and investment the initial level lies around 0.45-0.55 but convergence is fast except for investment. Wages is the least predictable variable, with a relative RMSFE starting around 0.8. All this means that a rational agent understanding this economy should be able to forecast in such a way as to beat clearly the unconditional mean in the case of hours, inflation, nominal interest rates even at very long horizons. For consumption, output and specially investment, he would clearly beat the mean even at 6 quarters ahead.

In the case of the RBC version (Chart 2) the conclusions are naturally quite different. The model becomes silent with respect to inflation and nominal interest rates but for the remaining variables the convergence of the RMSFE towards the standard deviation of the variables is much faster. For output, the relative RMSFE is around 0.8 for 1-step ahead forecasts and above 0.9 afterwards. For consumption and investment the speed of convergence is lower but clearly higher than that of the sticky prices/wages version. For wages, there is only significant predictability at 1-step ahead whereas for hours convergence of the RMSFE towards the standard deviation is slow but at a level clearly above that of the NK version. Now, it is important to note that this feature of the specific NK model analysed here is certainly common to any model featuring price and wage setting frictions along with an important indexation mechanism (to target or current inflation or a combination of the two) aimed at rationalizing the observed persistence of inflation, see *e.g.*, the models in Christiano *et al*. (2005), Adolfson *et al*. (2007), Ireland (2007) or Schorfheide (2005). This occurs because indexation generates high persistence in inflation and in other variables (and thus strong forecastability). In other way, any deviation of inflation from target in this kind of world represents a persistent (forecastable) deviation of the economy from its steady state.

Chart 1

RELATIVE RMSFE OF SW07 | NK VERSION

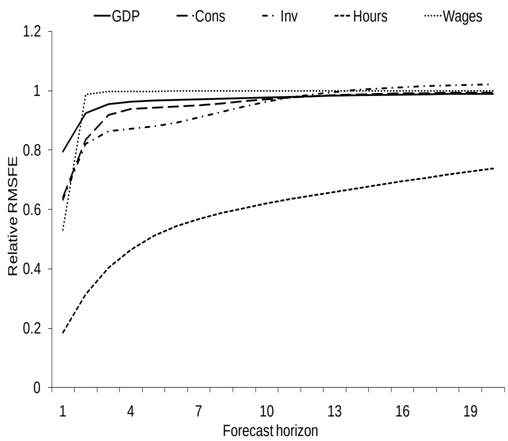


Source: Author's calculations.

Note: This figure presents the relative (to the standard deviation of the variables) RMSFE for each macro variable of interest at different forecast horizons (up to 20).

Chart 2

RELATIVE RMSFE OF SW07 | RBC VERSION



Source: Author's calculations.

Note: This figure presents the relative RMSFE for each macro variable of interest at different forecast horizons (up to 20).

3.2. Model *vs.* Data

Here we confront the results in section 2, regarding Phil-SPF's forecasts, with the theoretical and empirical forecast accuracy of the NK and RBC versions of SW07 analyzed above. To be clear, we view the relative (to the standard deviation) RMSFE of well-informed agents in the economy (Professional forecasters), as a statistic that should be matched by a realistic DSGE model, just as it should deliver steady-state ratios, volatilities and correlations that are close to what is observed in the data. *E.g.*, if this relative RMSFE for output growth at 1 quarter horizon is 0.3 in the model and 0.8 in the case of Professional forecasters (data), we view this as an indication that the model delivers a forecastability that is at odds with the data. And similarly if after 10 quarters the model is still able to clearly outperform the mean whereas Professional forecasters don't. Comparison of Professional and Rational (given the model) forecasts can thus inform theory or at least show the limitations of the theoretical models, even though the mapping from Rational to Professional forecasts may be considered loose.

If nothing else, we believe Professional forecasts allow us to measure how fast (from the perspective of the forecasters) the economy is moving towards the steady-state. Specifically, we can measure this convergence to the steady-state through the speed with which the RMSFE converges to the standard deviation of the variables. In fact, if after some time (horizon) the forecast is (on average) very close to the unconditional mean of the variable under scrutiny, this means the forecaster believes the economy (or at least that variable) takes as much time to reach the steady-state (in the absence of unpredictable shocks). With rational expectations this must be a characteristic of the process generating the data.[9]

Now, results in the previous section suggest that for most real variables (and in particular output, investment and consumption) Professional forecasts loose grip after 2 quarters, meaning that using as forecast an estimate of the unconditional mean of the variables does not imply loosing valuable information. Professional forecasts of unemployment and nominal interest rates are still clearly superior to the mean after one year whereas for inflation (CPI and GDP) there is forecastability but in a lesser extent. Notice further that we are using as benchmark a real-time estimate of the mean. If this mean is time-varying or shifts occasionally, *e.g.*, if the steady-state changes due to changes in taxation or in monetary policy (that changes for instance target inflation), the real-time average will not be efficient whereas professional forecasters are probably aware of these shifts. This is useful for our purposes as it allows us to interpret the relative RMSFE of Professional forecasts (which is thus deflated) as a **lower bound** on what a realistic theoretical model (without steady-state shifts) ought to deliver in terms of forecast accuracy relative to the steady-state forecast. Similarly then, in the mapping from Professional forecasts to the theoretical performance of SW07, we must see the model as corrected for regime changes, hence we cannot be as demanding when using the models in a pseudo out-of sample forecasting exercise with actual data.

Table 3 compares the results, for the theoretical and empirical (with actual data) relative (to the standard deviation) RMSFE of the NK and RBC versions of SW07, vis-a-vis that obtained with Phil-SPF forecasts. In the analysis of the empirical performance we focus on the sample 1981q3- 2009q2 (coinciding with the previous SPF's evaluation sample). If we take first the theoretical rel. RMSFE for output and investment, it is clear that the distance between the rel. RMSFE of Phil-SPF and that of the theoretical model is in general smaller for the RBC model, clearly so for all h in the case of investment and for $h=3,4,5$ in the case of output. At $h=1,2$ in the case of output, the RBC has a clearly lower forecastability. This result for output and investment contrasts to what obtains with the NK version, where the strong predictability at $h=1$ and even at long horizons is at odds with Phil-SPF. In the case of consumption the RBC is more successful at matching the data when $h=1,2$, whereas for $h=3,4,5$ the evidence favors the NK version (notice however that it may well be the case that the rel. RMSFE obtained with Phil-SPF may not be statistically different from 1). With respect to hours/unemployment (we recall that

9 We are certainly aware of the difficulty of characterizing as rational a consensus (mean or median) forecast, see *e.g.*, Bonham and Cohen (2001). Rationality should arguably be analysed at the individual level but exit and entry of forecasters in the surveys makes this a difficult task.

Phil-SPF forecasts unemployment, which explains nonetheless around 80% of the variation in hours), the RBC is closer to Phil-SPF at all horizons, although the rel. RMSFE is consistently above that of Phil-SPF for $h \geq 2$. This is in clear contrast with the strong predictability implied by the NK model. The RBC version is silent with respect to the nominal interest rate and inflation but for the NK model it is clear that while the behavior of the rel RMSFE function is very close to that of Phil-SPF in the case of nominal interest rates, for inflation the very high forecastability of the NK model does not match data from Phil-SPF forecasts. We notice also that even if the rational agent uses the forecasts produced by the univariate representation of inflation given the model (NK univariate, *i.e.*, using only past inflation to produce the forecast), the strong forecastability of inflation is almost unchanged. This seems a consequence of the degree of backward looking behavior (indexation) of inflation in the NK model. Once the rational agent observes current inflation and its history, information on other shocks is almost irrelevant to form the conditional expectation of inflation at some point in the future. If the model is realistic, this implies that a forecaster would only need to nail the univariate representation of inflation in order to obtain a close to efficient forecast.

Now, demanding from the models forecasts of actual data changes radically, in absolute terms, the picture above, with a clear deterioration of the empirical counterpart of the statistics above.[10] Nonetheless, Table 3 (bottom panel) shows that for real output and investment the RBC is close to Phil-SPF (and dramatically superior to the NK version). For wages (no data for Phil-SPF) the performance of both models is very similar whereas for consumption both the RBC and NK versions have a very weak performance (although the latter performs relatively better at horizons greater than 5 quarters, despite the fact that forecasts are close to the mean). For nominal interest rates , the NK model is close to Phil-SPF at $h = 1, 2$ but it drifts quite fast afterwards, becoming useless after 6 quarters (in clear contrast with the theoretical result). For inflation, the empirical performance of the NK model is beyond terrible, a qualification also deserved for the behavior of the RBC version with respect to hours (in this case the NK version is clearly more informative but not much compared to Phil-SPF at $h > 2$). As far as we are aware, only Rubaszek and Skrzypczynski (2008) compared forecasts from a (3 equations prototypical) DSGE model to SPF forecasts while using real-time data for estimation and forecasting (instead of the latest vintage of data and a fixed set of parameters, useful for our purposes). Their sample size is also larger than usual, spanning 1994:q1-2006:q2. The main conclusions are that while for a few horizons in the case of GDP growth the DSGE model seems to outperform SPF (not statistically significant difference in accuracy), in the case of inflation and short-term nominal interest rate SPF clearly outperforms the DSGE model.[11]

All in all, the results above suggest that the nominal rigidities apparatus of the NK model, which greatly amplifies the effects of shocks, tended to produce an excessively large theoretical forecastability, extending over long horizons. This seems clearly at odds with the data. The stripped down flexible prices version (RBC) delivers a forecastability resembling more that of the Phil-SPF while performing relatively better empirically (the important exception relates to hours/unemployment). This is due to the fact that deviations from the steady-state tend to be small, hence forecasts (conditional expectations) are closer to the mean of the variables. Thus, not taking risks (or not assuming a detailed knowledge of short-run dynamics) compensated in this context. The RBC model seemed more immune to misspecification (notice also that the RBC version was not even re-estimated, it keeps all the parameters from the estimated NK model). Next we repeat the analysis for recession periods.

10 Again, it is fair to recognize that the literature aknowledges the likely misspecification of DSGE models. *E.g.*, Del Negro *et al*. (2007) approximate a DSGE model by a vector autoregression (VAR) and then relax the implied cross-equation restrictions in order to improve fit. It is possible to optimally relax these restrictions and it is found that forecast accuracy improvements obtain.

11 Edge *et al*. (2010) do compare the forecast performance of an alternative DSGE model to Green-Book forecasts from 1996 through 2004, arguing for a positive contribution of the model in some instances (specially for output growth).

Table 3

RELATIVE RMSFE OF SPF'S FORECASTS *VIS-A-VIS* THE THEORETICAL AND EMPIRICAL NK AND RBC MODELS' PREDICTIONS

PANEL A - THEORETICAL

Variáveis	Modelo	Horizon												
		1	2	3	4	5	6	8	10	12	14	16	18	20
Real GDP growth	**SPF**	0.66	0.78	0.91	0.96	0.98	-	-	-	-	-	-	-	-
	RBC	0.79	0.92	0.96	0.96	0.97	0.97	0.97	0.98	0.98	0.98	0.99	0.99	0.99
	NK	0.55	0.69	0.78	0.85	0.89	0.92	0.96	0.97	0.98	0.98	0.99	0.99	1.00
Consumption	**SPF**	0.68	0.74	0.78	0.84	0.83	-	-	-	-	-	-	-	-
	RBC	0.64	0.83	0.92	0.94	0.94	0.95	0.96	0.97	0.98	0.99	0.99	0.99	0.99
	NK	0.41	0.59	0.72	0.82	0.88	0.91	0.95	0.97	0.97	0.97	0.97	0.97	0.98
RNR investment	**SPF**	0.65	0.73	0.79	0.90	0.94	-	-	-	-	-	-	-	-
	RBC	0.63	0.82	0.86	0.87	0.88	0.89	0.93	0.96	0.99	1.00	1.01	1.02	1.02
	NK	0.43	0.59	0.65	0.68	0.71	0.73	0.78	0.81	0.84	0.85	0.86	0.87	0.87
Hours	**SPF**	0.20	0.22	0.24	0.30	0.36	-	-	-	-	-	-	-	-
	RBC	0.18	0.31	0.40	0.47	0.51	0.54	0.59	0.62	0.65	0.67	0.69	0.72	0.74
	NK	0.01	0.02	0.03	0.05	0.07	0.10	0.16	0.24	0.31	0.39	0.46	0.53	0.59
Wages	**SPF**	-	-	-	-	-	-	-	-	-	-	-	-	-
	RBC	0.53	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NK	0.83	0.88	0.89	0.90	0.90	0.90	0.91	0.92	0.92	0.93	0.94	0.94	0.95
Inflation	**SPF**	0.67	0.70	0.72	0.75	0.77	-	-	-	-	-	-	-	-
	Univariate (NK-based)	0.14	0.29	0.40	0.49	0.55	0.60	0.68	0.73	0.77	0.80	0.82	0.85	0.86
	NK	0.13	0.27	0.38	0.46	0.53	0.58	0.66	0.72	0.76	0.79	0.82	0.84	0.86
Interest Rates	**SPF**	0.09	0.22	0.25	0.28	0.33	-	-	-	-	-	-	-	-
	Univariate (NK-based)	0.14	0.29	0.40	0.49	0.55	0.60	0.68	0.73	0.77	0.80	0.82	0.85	0.86
	NK	0.04	0.08	0.13	0.18	0.22	0.26	0.33	0.39	0.44	0.48	0.52	0.55	0.58

PANEL B - EMPIRICAL

Variáveis	Modelo	Horizon												
		1	2	3	4	5	6	8	10	12	14	16	18	20
Real GDP growth	**SPF**	0.66	0.78	0.91	0.96	0.98	-	-	-	-	-	-	-	-
	RBC	0.84	0.86	0.99	1.02	1.13	1.16	1.20	1.23	1.23	1.20	1.18	1.14	1.13
	NK	1.66	1.71	1.76	1.62	1.73	1.58	1.36	1.21	1.13	1.11	1.09	1.09	1.09
Consumption	**SPF**	0.68	0.74	0.78	0.84	0.83	-	-	-	-	-	-	-	-
	RBC	1.19	1.31	1.47	1.67	1.73	1.79	1.68	1.59	1.43	1.26	1.20	1.09	1.06
	NK	2.00	2.05	1.76	1.42	1.36	1.28	1.02	0.95	0.90	0.91	0.93	0.95	1.00
RNR investment	**SPF**	0.65	0.73	0.79	0.90	0.94	-	-	-	-	-	-	-	-
	RBC	0.66	0.93	1.07	1.10	1.06	1.02	0.95	0.96	0.99	1.03	1.10	1.13	1.15
	NK	1.23	2.15	2.75	2.96	2.91	2.76	2.22	1.84	1.50	1.31	1.23	1.18	1.17
Hours	**SPF**	0.20	0.22	0.24	0.30	0.36	-	-	-	-	-	-	-	-
	RBC	0.96	0.96	0.97	0.98	0.98	0.99	1.02	1.02	1.06	1.09	1.14	1.17	1.19
	NK	0.09	0.26	0.47	0.66	0.79	0.92	1.11	1.26	1.30	1.26	1.27	1.23	1.29
Wages	**SPF**	-	-	-	-	-	-	-	-	-	-	-	-	-
	RBC	0.98	0.99	1.01	1.02	1.02	1.03	1.04	1.02	1.05	1.06	1.09	1.12	1.14
	NK	1.08	0.97	0.97	0.98	0.99	1.00	1.02	1.03	1.07	1.09	1.13	1.16	1.18
Inflation	**SPF**	0.67	0.70	0.72	0.75	0.77	-	-	-	-	-	-	-	-
	NK	4.13	4.63	4.71	4.33	4.00	3.70	3.06	2.38	2.22	2.19	2.13	2.21	2.22
Interest Rates	**SPF**	0.09	0.22	0.25	0.28	0.33	-	-	-	-	-	-	-	-
	NK	0.06	0.19	0.39	0.64	0.90	1.11	1.48	1.68	1.85	2.00	2.09	2.15	2.21

Sources: Author's calculations

Notes: This table presents the relative (to the standard deviation of the variables) RMSFE at different forecast horizons, from h=1,…,20 for the SPF, theoretical and empirical NK and RBC models. Panel A also includes the univariate version of the NK SW07 model for both inflation and interest rates.

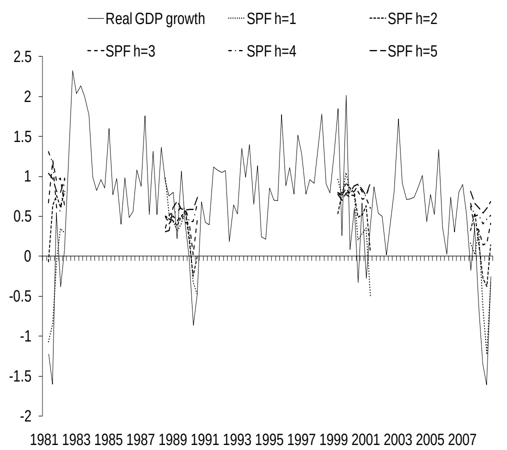
4. How do Rational and Professional Forecasts behave during Recessions?

There is clear evidence that macroeconomic forecasts fail to predict business-cycle's turning points and, moreover, forecasting the beginning of a recession one or two quarters in advance never occurred. In this aspect data (professional forecasts) are in line with standard models, where recessions must be seen as the result of large exogenous shocks (or at least unpredictable shocks in size and moment). Hence, one should not demand (or expect) accurate forecasts referred to the first period (quarter) of a recession. Afterwards, the theoretical mechanisms embodied in the models should be helpful in determining the path of observed variables.

Here we show that the conclusions above seem to carry over to recession[12] periods, and are certainly magnified. That is, the performance of the NK version of SW07 is quite poor compared to that of the RBC version. First, we recall that Professional forecasts (from the Fed-Green Book or SPF) have a poorer (relative to an estimate of the unconditional mean) performance during recession periods, specially at horizons greater or equal to 3 quarters. The exceptions occur with inflation and nominal interest rates as well as with housing market variables for short horizons (housing starts and residential investment). Despite this, they are clearly more accurate than model forecasts. To analyze this we simply plot the various forecast (Phil-SPF, NK and RBC) for $h = 1, \ldots, 5$ for real GDP growth, inflation and interest rates (Chart 3) Analysis of other real variables conveys a very similar message. As easily seen, Professional forecasts of real GDP have no clue about the beginning and dynamics of recessions with an anticipation of 2 or more quarters $(h \geq 3)$ whereas 1 quarter earlier they have some signal and for the current quarter they are accurate ($h = 1$, we recall that one step ahead forecasts in the case of Professional forecasts is really a nowcast). Now, although the RBC model performs poorly relative to professional forecasts, the characterization is very similar. The RBC obviously does not anticipate the recessions but provides signal about subsequent developments when $h = 1, 2$. The performance of the NK model is clearly very weak, specially during the last recession, where observed deflation and very low nominal interest rates contribute to forecasts that never consider consecutive negative growth (but instead a quick way out of the recession). This is clearly not the case in the 1991 and 2001 recessions. Again, the defensive" (or close to steady-state) dynamics implicit in the RBC version seem to at least produce forecasts that have some signal (although definitely close to the steady-state, or unconditional mean). For inflation and nominal interest rates we observe that professional forecasts are very accurate at short horizons and convey some signal at longer horizons. For nominal interest rates, the NK model does not produce out of bounds forecasts, but they are weak compared to those of surveys. For inflation, NK forecasts are very poor and do seem out of bounds, except during the last recession.

Chart 3 (to be continued)

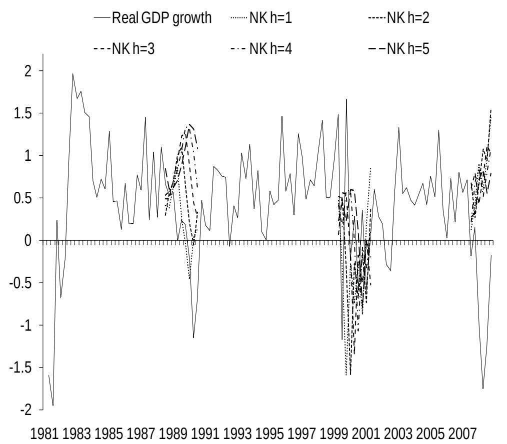
REAL GDP GROWTH AND SPF FORECASTS AT NBER RECESSIONS



Source: Author's calculations.

Note: This figure presents both the actual realization and SPF forecasts for real GDP growth between 1981 and 2009, the latter set plotting only observations at different horizons (1 to 5) for recession periods as identified by the NBER.

REAL GDP GROWTH AND NK FORECASTS AT NBER RECESSIONS



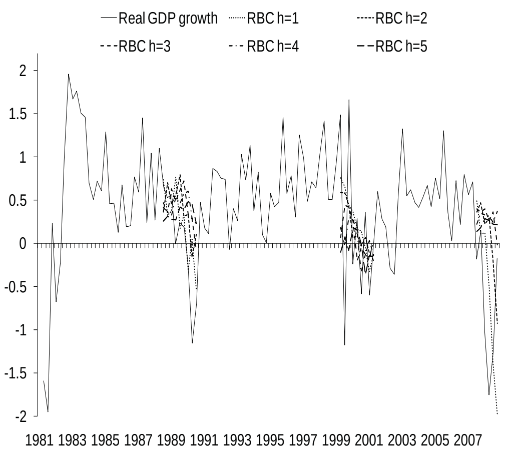
Source: Author's calculations.

Note: This figure presents both the actual realization and NK forecasts for real GDP growth between 1981 and 2009, the latter set plotting only observations at different horizons (1 to 5) for recession periods as identified by the NBER.

12 As identified by the NBER dates. For the purposes of this section we include a quarter before and a quarter after the recessions to capture turning points.

Chart 3 (to be continued)

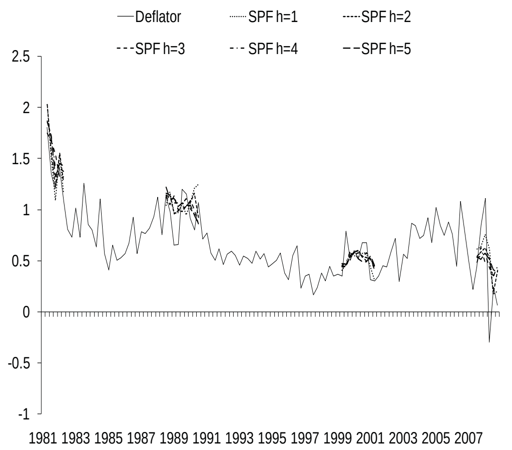
REAL GDP GROWTH AND RBC FORECASTS AT NBER RECESSIONS



Source: Author's calculations.

Note: This figure presents both the actual realization RBC forecasts for real GDP growth between 1981 and 2009, the latter set plotting only observations at different horizons (1 to 5) for recession periods as identified by the NBER.

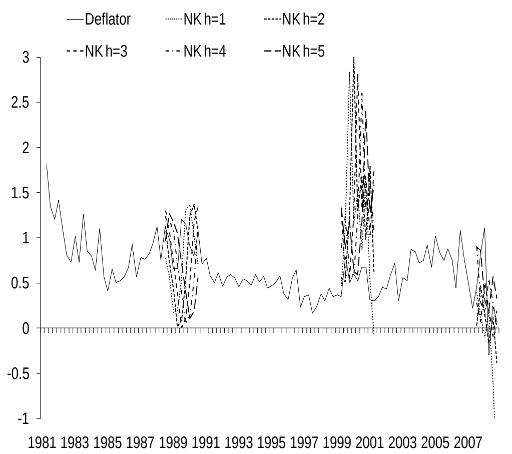
GDP DEFLATOR INFLATION AND SPF FORECASTS AT NBER RECESSIONS



Source: Author's calculations.

Note: This figure presents both the actual realization and predicted SPF values for inflation (GDP deflator) between 1981 and 2009, the latter set plotting only observations at different horizons (1 to 5) for recession periods as identified by the NBER.

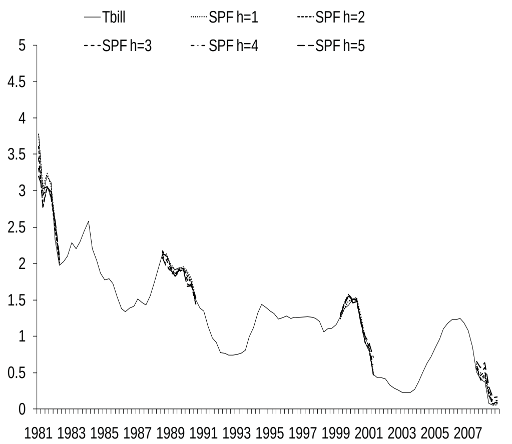
GDP DEFLATOR INFLATION AND NK FORECASTS AT NBER RECESSIONS



Source: Author's calculations.

Note: This figure presents both the actual realization and NK forecasts of GDP inflation between 1981 and 2009, the latter set plotting only observations at different horizons (1 to 5) for recession periods as identified by the NBER.

T-BILL RATE AND SPF FORECASTS AT NBER RECESSIONS

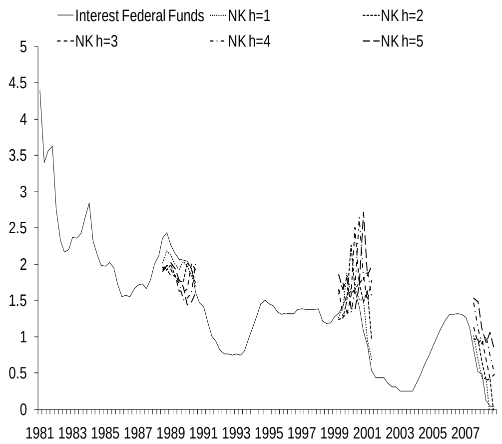


Source: Author's calculations.

Note: This figure presents both the actual realization and SPF forecasts for interest rates (T bill) between 1981 and 2009, the latter set plotting only observations at different horizons (1 to 5) for recession periods as identified by the NBER.

Chart 3 (continued)

FEDERAL FUNDS RATE AND NK FORECASTS AT NBER RECESSIONS



Source: Author's calculations.

Note: This figure presents both the actual realization and NK forecasts for interest rates (Federal Funds rate) between 1981 and 2009, the latter set plotting only observations at different horizons (1 to 5) for recession periods as identified by the NBER.

5.Concluding remarks

It seems unwise to expect too much from macroeconomic forecasts. For what really matters (real variables, but except for unemployment) best practice has little to say at horizons greater than 2, 3 quarters. If statistics derived from these facts inform general equilibrium modelling, in the sense that a rational agent understanding the economy should deduce similar statistics, they probably say the economy has not been deviating too much from the steady-state. In the theoretical models, this should translate into low forecastability (relative to a naive, or steady-state forecast and, again, except for unemployment - hours) of most variables. This occurs with the RBC version of the model analyzed here but clearly not with the NK version. Furthermore, even recognizing limitations in a model without nominal frictions and correspondingly limited departures from the steady-state, the fact is that empirical forecasts seem to indicate that the model less prone to misspecification is the RBC version. Forecasts are closer to naive (or to steady-state values) but provide some signal. The alternative (relying on a particular description of nominal rigidities) is not reliable. In our view, and given the effects of the inclusion of nominal frictions on forecast performance (theoretical and empirical), care should be taken at least on the way trend inflation (or varying central bank target) is modeled. In the model analyzed here and many others, the central bank target (steady-state inflation) is fixed, which implies that any deviation of inflation from target is necessarily interpreted by the model as a deviation from the steady-state (inflation gap). In order to improve fit the models must then include indexation mechanisms. In this respect we are persuaded by Cogley and Sbordone's (2008) analysis that once movements in trend inflation are taken into account, the (backward looking) indexation component of a general New-Keynesian Phillips curve is not needed to fit the data well. If indexation is incorrectly assumed, it implies a supposedly high theoretical forecastability of inflation (even if a rational agent only looks at past inflation) as we have shown. This is clearly at odds with the data (Professional forecasts) and does not survive a forecast evaluation with actual data. Another interpretation of the results rests on the observation that theoretical models used to fit several decades of data are likely missing relevant changes in monetary policy, product and labor market regulation, taxation or in the trend growth of technology. If these changes are reasonably unpredictable, there is potential compatibility between professional forecasters having a hard time and the NK model becoming seriously misspecified only along those dimensions, *i.e.*, nominal rigidities can play an important role which is hidden due to lack of control for what can be seen as steady-state shifts.

Bibliography

Adolfson, M., J. Lindé and M. Villani (2007), "Forecasting Performance of an Open Economy DSGE Model", *Econometric Reviews*, 26: 289-328.

Adolfson, M., S. Laséen, J. Lindé and M. Villani (2007), "Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through", *Journal of International Economics*, 72:481-511.

Adolfson, M., S. Laséen, J. Lindé and M. Villani (2008), "Evaluating an Estimated New Keynesian Small Open Economy Model", *Journal of Economic Dynamics and Control*, 32: 2690-2721.

Ang, A., Bekaer, G. and M. Wei. (2007), "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?", *Journal of Monetary Economics*, 54(4): 1163-1212.

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

Bonham, C. and R. H. Cohen (2001), "To Aggregate, Pool, or Neither: Testing the Rational Expectations Hypothesis Using Survey Forecasts," *Journal of Business and Economic Statistics*, 19(3): 278-91.

Capistrán, C. and A. Timmermann (2009), "Forecast Combination with Entry and Exit of Experts", *Journal of Business and Economic Statistics*, 27: 428-440.

Christiano, L., Eichenbaum, M. and C. Evans (2005), "Nominal rigidities and the dynamic effects of a shock to monetary policy", *Journal of Political Economy*, 113 (1): 1-45.

Christiano, L. J., Trabandt, M. and K. Walentin (2009), "Introducing Financial Frictions and Unemployment into a Small Open Economy Model", Sveriges Riksbank *Working Paper* Series No. 214.

Cogley, T. and Sbordone, A. M. (2008), "Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve", *American Economic Review*, 98(5): 2101-26.

Croushore, D. and T. Stark (2001), "A Real-Time Data Set for Macroeconomists", *Journal of Econometrics*, 105:111-130.

Croushore, D. and T. Stark (2003), "A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?", *Review of Economics and Statistics,* 85:605-617.

Croushore, D. (1993), "The Survey of Professional Forecasters", *Business Review*, Federal Reserve Bank of Philadelphia, November/December, 3-15.

D'Agostino, A. and K. Whelan (2008), "Federal Reserve information during the Great Moderation", *Journal of the European Economic Association*, 6(2-3): 609-620.

Del Negro, M., Schorfheide, F., Smets, F. and R. Wouters (2007). "On the Fit of New Keynesian Models", *Journal of Business and Economic Statistics*, 25: 123-143.

Edge, R. M., Kiley, M. T. and J. P. Laforte (2010), "A comparison of forecast performance between Federal Reserve staff forecasts, simple reduced-form models, and a DSGE model", *Journal of Applied Econometrics,* 25(4): 720-754.

Engelberg, J., Manski, C. F. and J. Williams (2009), "Comparing the Point Predictions and Subjetive Probability Distributions of Professional Forecasters", *Journal of Business and Economic Statistics*, 27: 30-41.

Fair, R. C. and R. J. Shiller (1989), "The Informational Context of Ex Ante Forecasts", *The Review of Economics and Statistics*, 71(2): 325-31.

Faust, J. and J. H. Wright (2007), "Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset", NBER *Working Papers* 13397.

Giannone, D., Lenza, M. and L. Reichlin (2008). "Explaining The Great Moderation: It Is Not The Shocks", *Journal of the European Economic Association*, 6(2-3): 621-633.

Ireland, Peter N. (2007), "Changes in the Federal Reserve's Inflation Target:Causes and Consequences", *Journal of Money, Credit and Banking*, 39(8): 1851-2110.

Joutz, F. and H. O. Stekler (2000), "An evaluation of the predictions of the Federal Reserve", *International Journal of Forecasting*, 16: 17-38.

McConnell, M. and G. Perez-Quiros (2000), "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?", *American Economic Review*, 90 (5): 1464-1476.

McNees, S.K., (1992), How large are economic forecast errors? *New England Economic Review*, 25-33.

McNees, S.K. and J. Ries (1983), "The track record of macroeconomic forecasts". *New England Economic Review,* 5-18.

Romer, C. and D. Romer (2000), "Federal Reserve information and the behaviour of interest rates", *American Economic Review*, 90: 429-457.

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

Schorfheide, F. (2005), "Learning and monetary policy shifts". *Review of Economic Dynamics*, 8(2): 392-419.

Smets, F. and R. Wouters (2003), "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area", *Journal of the European Economic Association*, 1(5): 1123-1175.

Smets, F. and R. Wouters. (2007) "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach", *American Economic Review*, 97(3): 586-606.

Stark, T. (2010), "Realistic evaluation of real-time forecasts in the Survey of Professional Forecasters", Federal Reserve Bank of Philadelphia, *Research Rap Special Report*, May 2010.

Stock, J.H. and M.W. Watson (2003), "Has the Business Cycle Changed and Why?," NBER Chapters, in: NBER *Macroeconomics Annual* 2002, Vol17, pp 159-230, National Bureau of Economic Research.

Zarnowitz, V. (1969), "The new ASA-NBER Survey of Forecasts by Economic Statisticians", *American Statistitican*, 23: 12-16.

Zarnowitz, V., (1992), "Business cycles: Theory, history, indicators and forecasting" in National Bureau of Economic *Research Studies in Business Cycles*, vol. 27. University of Chicago Press, Chicago.

Zarnowitz, V., Braun, P., (1992), "Twenty two years of the NBER-ASA Quarterly Outlook Surveys: Aspects and comparisons of forecasting performance", NBER *Working Paper* 3965, New York.