Predicting unemployment in short samples with internet job search query data

Francesco, D'Amuri

Bank of Italy - Research Department

30 October 2009

Predicting unemployment in short samples with internet job search query data

Francesco D'Amuri*

October 30, 2009

Abstract

This article tests the power of a novel indicator based on job search related web queries in predicting quarterly unemployment rates in short samples. Augmenting standard time series specifications with this indicator definitely improves out-of-sample forecasting performance at nearly all in-sample interval lengths and forecast horizons, both when compared with models estimated on the same or on a much longer time series interval.

Keywords: Google econometrics, Forecast comparison, Keyword search, Unemployment, Time series models.

JEL Classification: C22, C53, E27, E37, J60, J64.

*Bank of Italy - Research Department and University of Essex, Institute for Social and Economic Research. I wish to thank A. Brandolini, O. Jorda, J. Marcucci, G. Parigi, M. Porqueddu, A. Rosolia, P. Sestito, R. Torrini and F. Zollino for useful comments. Support from the Economic and Social Research Council is gratefully acknowledged. The opinions expressed in this paper do not necessarily reflect those of the Bank of Italy.

1 Introduction

In a recently published article, Ginsberg et al. (2009) developed a simple model forecasting physician visits due to influenza-like illness (ILI) in a particular region using as a single explanatory variable the ILI-related query fraction on total queries as recorded by Google search engine data, weekly available with short delay[1].

Following the high popularity of the internet as a job search mean (Stevenson, 2008), and the increasing need of up-to-date economic indicators in the current economic downturn, this approach has recently been extended to unemployment forecasting. In particular, the Google Index (GI), equal to the incidence of google job search related queries over total queries, proved to have predictive power in forecasting unemployment developments in Germany (Askitas and Zimmermann, 2009), Israel (Suhoy, 2009) and United States (Choi and Varian, 2009; D'Amuri and Marcucci, 2009), countries where unemployment data are monthly available. The scope of this article is to test the empirical relevance of this indicator in a country, Italy, where only quarterly unemployment data are available, and possible gains associated with improving forecast accuracy based on real time data are even greater. This comes to a cost, since the availability of quarterly data instead of monthly ones reduces the number of observations that can be used for estimation and for testing the forecast accuracy. Nevertheless, forecasting performance is assessed estimating a great number of alternative models, using different exogenous variables and different sample lengths, and then testing their out of sample forecast accuracy in a number of rolling and recursive estimates. Given the fact that the time series for which GI data are available is short, we also compare models estimated using the GI as an explanatory variable with otherwise identical models, but estimated on a much longer interval (starting with 1985:1), as well as on the same time interval. With really few exceptions, confined to cases in which seasonally unadjusted series were used for estimation on very small in-sample intervals, models including the Google Index perform better than the others, having lower Mean Squared Error at multiple estimation lengths and forecast horizons. This is true also when the comparison is made with respect to standard time series models augmented with exogenous variables perceived as leading indicators for labor market developments (in particular, the employment expectations taken from business surveys and the industrial production index).

This result suggests that, even in countries where only quarterly unemployment rate data are available and the GI time series is short, this variable should still routinely be included in models forecasting unemployment developments. This work is organized as follows: section 2 describes the data used and their limitations, while section 3 provides estimates and residuals' diagnostic tests.

[1] http://www.google.com/insights/search/.

Section 4 tests the predictive power of the GI by comparing models including it as an exogenous variable with a series of otherwise identical models, section 5 concludes.

2 Data

In this article, variables coming from different sources are combined. The dependent variable is defined as the quarterly unemployment rate as recorded by the Italian Labor Force Survey (ISTAT, 2009b). The novel explanatory variable used is the unemployment-related query fraction on total queries based on Google search data (the time series is available starting with the week ending the 10th of January, 2004), defined as the incidence of the queries for "job offers" (*"offerte di lavoro"*) on total queries. This is a weekly index, rendered quarterly by taking simple intra quarter averages. The index is normalized with a value equal to 100 indicating the week in which this incidence was the highest.

Note that a person is considered as unemployed if she is not currently employed and if she conducted at least a job search action in the preceding four weeks. Thus it is not correct to compare unemployment estimates for quarter t with Google Index values relative to exactly the same time span. Individuals interviewed at the beginning of the quarter and considered as unemployed might well have looked for a job through the internet in the preceding weeks (and thus in quarter t-1). Given that exact interview's week is not known, it is not possible to match exactly the two variables. Nevertheless, in order to minimize the resulting bias, Google Index data are rescaled two weeks ahead. While the official unemployment rate is available normally with a delay of about three months, Google search data are recorded weekly and released with short delay. In the empirical analysis, the Industrial Production Index monthly published by Istat (ISTAT, 2009a), and the results of the employment expectations survey conducted by the European Commission (European-Commission, 2009) will also be used as exogenous variables. This last indicator is equal in each sector to the balance between the number of those professionals who forecast an increase/decrease in employment in the next three months. A single, private sector-wide, index is obtained as a weighted average of the sector specific indicators using as weights the number of employed individuals taken from ISTAT (2009b) for each sector (industry without construction, construction, private services excluding retail, retail) in each relevant quarter. Indicators for each of these sectors are available only starting with 1985:1, the reason why the longest series used in estimation starts in that quarter. Abberger (2007) showed the predictive ability of a similar index for Germany.

To our knowledge this is the first study forecasting unemployment developments in Italy using this set of exogenous variables.

The main limitation of the explanatory variable based on Google data is that it could be partly driven by on the job search, rather than unemployed job search activities which are the focus of this paper. In the 4th quarter of 2008, of the total of 2.4 million of individuals that declared to be engaged in job search activities in the previous four weeks, 640 thousand individuals (25 per cent) were employed[2]. This limitation is made more severe by the fact that, while unemployed job search is believed to follow the anti-cyclical variation of job separation rates, on the job-search is normally assumed to be cyclical. Another limitation is due to the fact that not all workers have access to the internet, and it is also presumable that workers using the internet for job search are not randomly selected among job-seekers. This should be a minor issue, given the increasing popularity of internet as a job search method and also due to the fact that a bias in the estimates would emerge only if shocks hit unemployed using/not using the internet for job search in a different way. As a final remark it has to be noted that the value taken by the Google index used here summarizes overall job search intensity of a population of individuals looking for a job via internet. It could virtually increase if internet job search intensity increases for a given pool of individuals. In this extreme case there would be no link between internet job search intensity and the number of people looking for a job. With these caveats in mind, we plot in Figure 1 the official unemployment rate figures and the Google job search index developments over the 2004:1-2009:1 interval. Visual inspection reveals strong similarities in the two series' dynamics, with the Google job search index seeming to be a leading indicator for the number of unemployed individuals. In the next sections we will assess its predictive power. This first impression is confirmed by the dynamic correlations (tab. 1): stronger between the unemployment rate and the GI when than between the EEI or the IPI and the GI.

3 Estimates and residuals' diagnostics

Before starting the empirical analysis we check for the presence of unit roots in the dependent variable by means of Augmented Dickey Fuller and Phillips-Perron tests, both without lags and with four lags. Since in all the four cases the formal tests fail to reject the null of the presence of a unit root, series in first differences are used for the remaining of the analysis. First differencing is common in the time series literature forecasting unemployment rates, see for example Montgomery et al. (1998) and Proietti (2003).

After having tested the properties of models with various lag structures following a general

[2]Author's calculations based on ISTAT (2009b).

to specific approach based on AIC and BIC minimization and residuals' diagnostics, a simple ARIMA(1,1,0) specification augmented with quarterly seasonal dummies is the preferred one. All main results hold true using seasonally adjusted series without seasonal dummies (see section 4). The ARIMA(1,1,0) is also the benchmark model in one of the most cited articles of the US literature on unemployment forecasting (Montgomery et al., 1998). The forecasting properties of this model will be compared across three dimensions: models estimated on the full sample (starting with 1985:1), models estimated on a shorter sample (starting with 2004:1) and otherwise identical models estimated on the shorter sample, but including the GI as an explanatory variable. For each group of models we alter the lag structure, the exogenous variables included and the month to which the explanatory variables refer to, for a total of 39 models. In table 2 we show estimation results for the basic ARIMA(1,1,0) model (model 1) and for five selected models including the lagged value of the Employment Expectations Index (EEI, model 2), the lagged value of the Industrial Production Index (IPI, model 3) and the current value of the Google Index (GI, model 4) or a combination of the GI and one of the other exogenous variables (models 5 and 6). It is interesting to note that the coefficient of the exogenous variables has the expected sign in each case. In particular, while the impact of lagged values of the EEI is negative but non significantly different from zero (Model 2), the coefficient estimate for IPI_{t-1} is negative and significant: an increase of one point in the IPI is associated with a decrease of 0.6 percentage points in the unemployment rate (Model 3). Also the GI coefficient estimate is positive as expected and significant, with a point increase in the GI associated with an increase of the unemployment rate of 0.44 percentage points. In terms of lower AIC and BIC the best performance is obtained by Model 5, adding the GI and the lagged EEI to the basic specification. Residuals diagnostics (see fig. 2 for residuals' plots), discussed for the main models here, but performed for each of the 39 estimated models, show no sign of missspecification, with mean zero approximately normal residuals, while Portmentau tests fail to reject the null of no autocorrelation up to the fourth lag.

4 Forecasting

Having assessed the explanatory power of the Google Index, its relevance in improving the models' forecasts is now tested. In particular, numerous out-of-sample forecast accuracy comparisons are conducted here. We compare the performance of three groups of models (for a total of 39 models) introduced before, using estimates obtained from rolling regressions performed on 7 different in-sample interval lengths: from 14 to 20 observations for the small sample models and from 90 to 96 for the full sample ones. Models estimated on the full sample always include the small sample interval at the end of the estimation period; the forecasting origin is consistent across specifications. Models' comparison is carried out by out-of-sample MSE. In table 3 one-step MSEs for these different groups of models are compared. When considering models estimated on the small sample, the inclusion of the *GI* dramatically reduces one step ahead MSE with few exceptions (the ARX1 model augmented with the lagged EEI performs better than the otherwise identical model including the GI when the in sample intervals are the smallest, equal to 14 and 15). The advantages of using the *GI* in forecast accuracy are confirmed when using longer forecast horizons (two or three steps ahead, see tables 4 and 5).

A more severe test is to compare the forecast accuracy of the GI models estimated on the short sample with models without the GI but estimated on the full sample, including 76 more observations at each in-sample estimation interval. Models estimated on this longer sample tend to perform definitely better than the ones using the short sample but not including the GI, probably due to a major precision in coefficient estimates underlying the out of sample projections. Nevertheless, the forecast performance of these models is comparable (or slightly superior) when compared with models estimated on the shorter sample but including the GI only when the latter are estimated using a really small number of in-sample observations. When at least 16 observations are used to estimated the GI models, their forecasting performance becomes strictly superior. Just to give an example, the best model using 17 in sample observations and the GI (a simple ARIMAX (1,1,0)) has a mean squared error of 0.08 when forecasting one step ahead, while the best model estimated on the longer series without GI (93 observations, ARIMAX (1,1,0) with the current value of the EEI) has a 0.12 MSE. The inclusion of the GI halves the best model's MSE when considering the same estimation sample (MSE is equal to 0.16 when excluding the GI).

Results are reinforced when considering the seasonally adjusted series (tab. 6, 7 and 8). These models gain three degrees of freedom due to the fact that seasonal dummies need not to be included in the in-sample estimates. In this case, models including the GI have the lowest MSEs at all in-sample forecasting intervals and horizons. Only the models estimated on the long sample have similar MSEs when considering the shortest in sample forecasting intervals.

5 Conclusions

The aim of this article has been to test the empirical relevance of internet job search query data (Google index) in forecasting unemployment in a country, Italy, where only quarterly unemployment data are available and and its relevance has to be assessed on small samples. This real time indicator performs fairly well in estimating and forecasting the evolution of unemployment and it is superior to other widely accepted leading indicators of unemployment dynamics, such as employment expectations surveys and the industrial production index. More interesting, models estimated on small samples but including the google index perform better than otherwise identical models estimated on a much longer sample, even when augmented with other leading indicators. It is easy to guess that web search data will routinely be used for forecasting short term economic dynamics in the future, even in countries where only quarterly unemployment data are available.

References

Abberger, K. (2007). Qualitative business surveys and the assessment of employment - A case study for Germany. *International Journal of Forecasting* (23), 249–258.

Askitas, N. and K. F. Zimmermann (2009). Google Econometrics and Unemployment Forecasting. *IZA Discussion Paper* (4201).

Choi, H. and H. Varian (2009). Predicting Initial Claims for Unemployment Benefits. *Google technical report*.

D'Amuri, F. and J. Marcucci (2009). Google it! Forecasting the US unemployment rate with a Google job search index. *Bank of Italy, mimeo*.

European-Commission (2009). EU employment situation and social outlook. *Various months*.

Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant (2009). Detecting Influenza epidemics using Search Engine Query Data. *Nature* (457), 1012–1014.

ISTAT (2009a). Indice della produzione industriale. *Various months*.

ISTAT (2009b). Rilevazione sulle forze di lavoro. *Various quarters*.

Montgomery, A. L., V. Zarnowitz, R. S. Tsay, and G. C. Tiao (1998). Forecasting the U.S. Unemployment Rate. *Journal of the American Statistical Association* (93), 478–493.

Proietti, T. (2003). Forecasting the U.S. Unemployment Rate. *Computational Statistics & Data Analysis* (42), 451–476.

Stevenson, B. (2008). The Internet and Job Search. *NBER Working Paper* (13886).

Suhoy, T. (2009). Query Indices and a 2008 Downturn. *Bank of Israel Discussion Paper* (06).

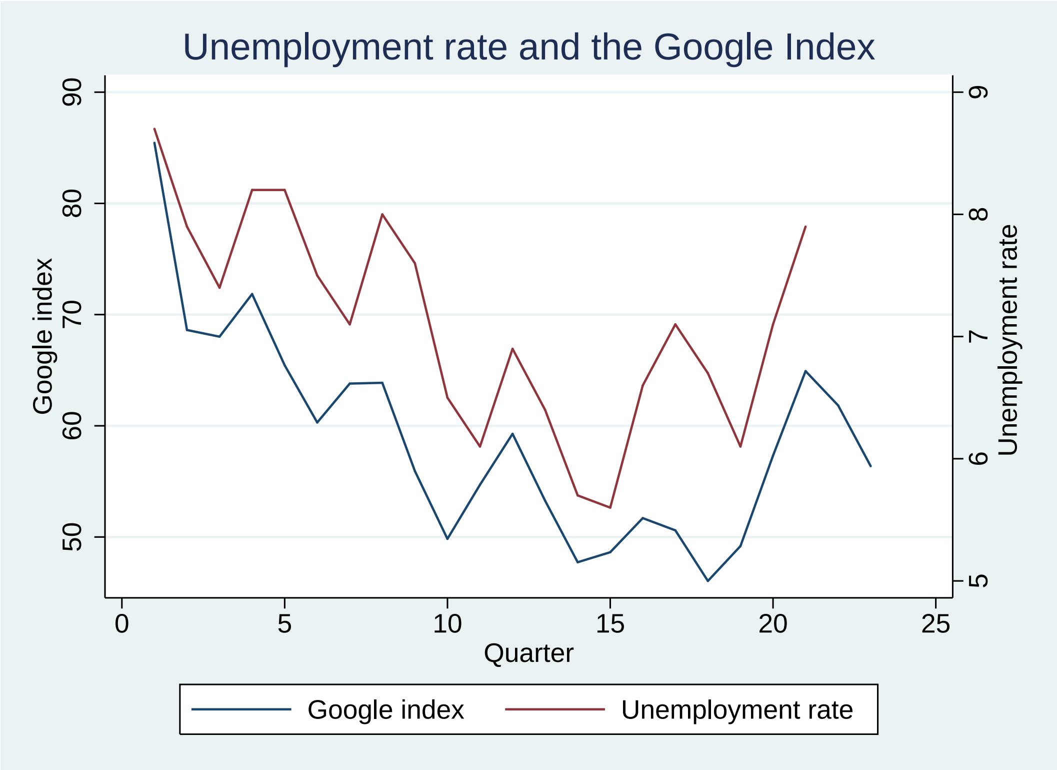

Figure 1: Unemployment rate and internet job search (Google index)

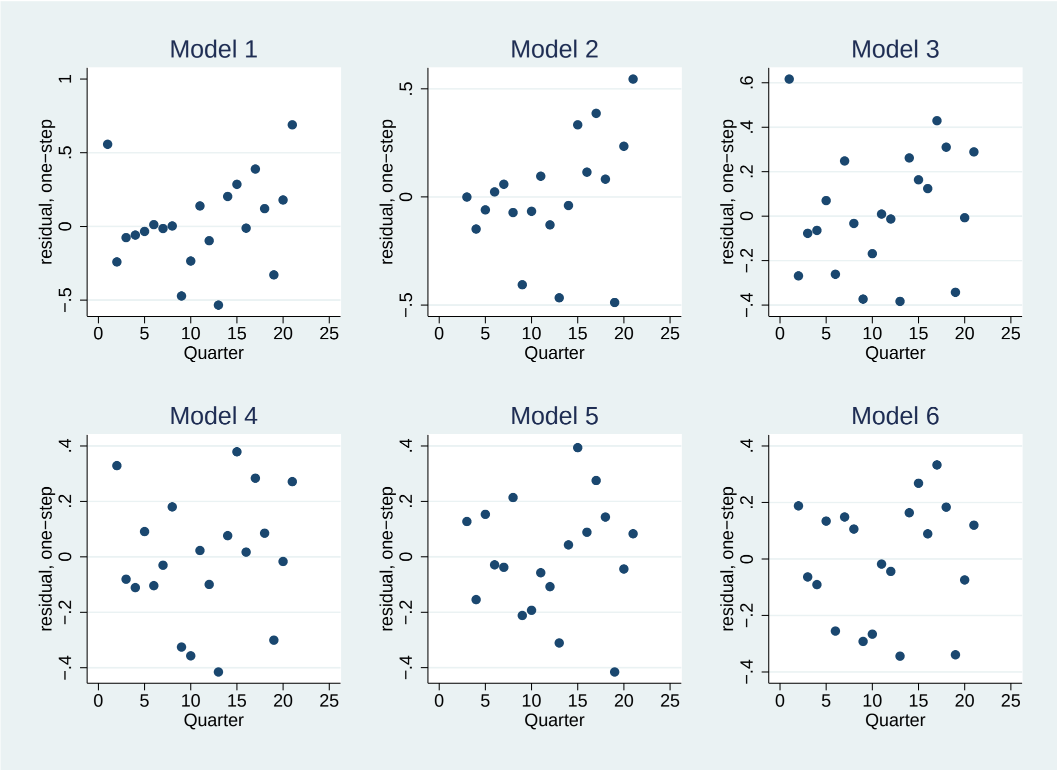

Figure 2: Selected models residuals' plot. See text for details.

	U_t	GI_t	GI_{t-1}	GI_{t-2}	EEI_t	EEI_{t-1}	EEI_{t-2}	IPI_t	IPI_{t-1}	IPI_{t-2}
U_t	1.000									
GI_t	0.821	1.000								
GI_{t-1}	0.775	0.786	1.000							
GI_{t-2}	0.395	0.566	0.809	1.000						
EEI_t	-0.408	-0.351	-0.098	0.108	1.000					
EEI_{t-1}	-0.488	-0.430	-0.243	-0.116	0.795	1.000				
EEI_{t-2}	-0.597	-0.557	-0.419	-0.380	0.457	0.678	1.000			
IPI_t	-0.149	-0.444	-0.028	-0.114	0.537	0.590	0.554	1.000		
IPI_{t-1}	-0.647	-0.620	-0.306	0.062	0.225	0.275	0.379	0.067	1.000	
IPI_{t-2}	-0.456	-0.263	-0.536	-0.206	0.140	0.074	0.154	-0.005	-0.062	1.000

Table 1: Dynamic correlations

Notes: U is the quarterly unemployment rate released by ISTAT (2009b); GI is the quarterly average of weekly Google Index data, summarizing internet job search intensity; IPI is the quarterly average of the monthly Industrial Production Index released by ISTAT (2009a); EEI is the quarterly average of the monthly economy-wide Employment Expectations obtained weighting the sector specific values released by European-Commission (2009) with sector specific number of employed. 2004:1-2009:1 interval.

Model	1	2	3	4	5	6
GI_t				0.044035	0.0589143	0.0328446
				(3.604541)	(2.269399)	(2.11691)
EEI_{t-1}		-0.0270814			-0.0129934	
		(-1.00688)			(-.5437773)	
IPI_{t-1}			-0.0614131			-0.0394722
			(-3.300874)			(-1.689716)
Dummy Q=2	-0.6626246	-0.6714958	-0.6062042	-0.4061492	-0.3995967	-0.4300818
	(-3.398511)	(-1.705016)	(-4.626498)	(-3.159385)	(-2.72638)	(-3.73056)
Dummy Q=3	-1.021122	-1.03037	-0.8340279	-0.9214937	-0.9280641	-0.8199131
	(-3.904108)	(-2.651042)	(-3.522234)	(-6.050438)	(-6.759543)	(-5.254275)
Dummy Q=4	-0.0951856	-0.0903765	-0.7110005	-0.2050636	-0.2458973	-0.5681298
	(-.8857289)	(-.4262475)	(-3.624205)	(-2.837706)	(-2.20283)	(-2.194642)
UR_{t-1}	-0.3827365	0.329652	0.2268644	0.3161321	0.3007895	0.2828104
	(-0.8575282)	(0.6352761)	(1.008321)	(0.8552952)	(0.8232047)	(1.027195)
AIC	17.4	17.4	12.3	10.7	8.5	9.0
BIC	23.3	24.0	19.3	17.7	16.1	17.0
Res. diagn.						
Mean	0.0022236	6.97E-06	0.0016974	-0.0051989	-0.0020127	-0.0026583
Skewness	0.2607763	-0.0509333	-0.0300112	-0.119582	-0.1173554	-0.3110008
Kurtosis	3.491314	2.817882	2.084041	2.223353	2.594147	1.904184
Jb norm. test (*Pvalue*)	.8074	.9829	.7039	.7595	.9167	.5161
Port. test (*Pvalue*)						
Lag1	0.6475	0.4862	0.8489	0.7985	0.7603	0.7798
Lag2	0.2425	0.0987	0.5937	0.2838	0.677	0.4537
Lag3	0.1639	0.1978	0.4326	0.1696	0.7885	0.3647
Lag4	0.2126	0.2867	0.5382	0.258	0.2985	0.4102

Table 2: Selected models' estimates. Values of the t statistic in parenthesis.

Notes: Models estimated on the 2004:1-2009:1 sample, estimates for the constant are not reported. GI is the quarterly average of weekly Google Index data, summarizing internet job search intensity; IPI is the quarterly average of the monthly Industrial Production Index released by ISTAT (2009a); EEI is the quarterly average of the monthly economy-wide Employment Expectations obtained weighting the sector specific values released by European-Commission (2009) with sector specific number of employed. See text for further details on variables' definition.

In sample obs.	90	14	14	91	15	15	92	16	16	93	17	17	94	18	18	95	19	19	96	20	20
Smpl length	L	S	S	L	S	S	L	S	S	L	S	S	L	S	S	L	S	S	L	S	S
Includes GI_t			Yes			Yes			Yes			Yes			Yes			Yes			Yes
AR1	0.25	0.26	0.19	0.29	0.29	0.17	0.24	0.33	0.20	0.25	0.26	0.08	0.33	0.34	0.09	0.45	0.40	0.06	0.52	0.75	0.33
$\text{ARX1}_{(IPI_t)}$	0.24	0.25	0.19	0.27	0.29	0.20	0.33	0.34	0.23	0.26	0.26	0.12	0.34	0.34	0.14	0.35	0.35	0.06	0.65	0.65	0.34
$\text{ARX1}_{(EEI_t)}$	0.17	0.41	0.28	0.19	0.37	0.23	0.14	0.36	0.25	0.12	0.25	0.10	0.17	0.30	0.10	0.19	0.34	0.05	0.16	0.62	0.29
$\text{ARX1}_{(IPI_t)}^{(EEIt)}$	0.25	0.43	0.31	0.29	0.39	0.26	0.34	0.40	0.29	0.26	0.27	0.15	0.34	0.34	0.17	0.32	0.32	0.06	0.58	0.58	0.28
$\text{ARX1}_{(IPI_{t-1})}$	0.29	0.31	0.21	0.32	0.34	0.22	0.36	0.37	0.22	0.34	0.35	0.15	0.44	0.44	0.20	0.35	0.35	0.03	0.61	0.61	0.07
$\text{ARX1}_{(EEI_{t-1})}$	0.32	0.16	0.23	0.37	0.17	0.18	0.34	0.26	0.20	0.36	0.19	0.10	0.46	0.24	0.10	0.65	0.22	0.02	0.85	0.43	0.13
$\text{ARX1}_{(IPI_{t-1})}^{(EEI_{t-1})}$	0.25	0.23	0.28	0.29	0.25	0.25	0.31	0.31	0.23	0.28	0.28	0.16	0.36	0.36	0.21	0.19	0.19	0.01	0.34	0.34	0.02
$\text{ARX1}_{(IPI_{M1}t)}$	0.27	0.34	0.29	0.30	0.36	0.25	0.36	0.40	0.27	0.32	0.33	0.18	0.42	0.43	0.25	0.47	0.48	0.26	0.90	0.90	0.71
$\text{ARX1}_{(EEI_{M1}t)}$	0.25	0.31	0.29	0.30	0.35	0.28	0.26	0.39	0.29	0.27	0.30	0.18	0.35	0.40	0.24	0.49	0.50	0.30	0.56	0.92	0.75
$\text{ARX1}_{(IPI_{M1}t)}^{(EEI_{M1}t)}$	0.33	0.48	0.37	0.37	0.52	0.35	0.44	0.52	0.34	0.42	0.44	0.24	0.53	0.54	0.32	0.51	0.53	0.31	0.96	0.96	0.78
$\text{ARX1}_{(IPI_{M1}t-1)}$	0.33	0.33	0.31	0.37	0.37	0.32	0.41	0.41	0.34	0.43	0.43	0.29	0.56	0.56	0.41	0.43	0.43	0.26	0.79	0.79	0.46
$\text{ARX1}_{(EEI_{M1}t-1)}$	0.22	0.22	0.26	0.25	0.25	0.25	0.29	0.29	0.23	0.22	0.22	0.14	0.28	0.28	0.18	0.31	0.31	0.19	0.63	0.63	0.43
$\text{ARX1}_{(IPI_{M1t-1})}^{(EEI_{M1t-1})}$	0.36	0.36	0.34	0.40	0.40	0.34	0.46	0.46	0.39	0.48	0.48	0.35	0.63	0.63	0.49	0.33	0.33	0.20	0.64	0.64	0.37

Table 3: MSE for one step ahead forecasts, rolling estimates.

Notes: GI is the quarterly average of weekly Google Index data, summarizing internet job search intensity; IPI is the quarterly average of the monthly Industrial Production Index released by ISTAT (2009a); EEI is the quarterly average of the monthly economy-wide Employment Expectations obtained weighting the sector specific values released by European-Commission (2009) with sector specific number of employed. Models estimated on in-samples of different lengths. Models estimated on the full sample (first column of each group; 90 or more observations) start the in-sample interval with 1985:1. Variables with subscript M1 are equal to the point value of the first month of the relevant quarter instead than to the quarter average.

In sample obs.	90	14	14	91	15	15	92	16	16	93	17	17
Smpl length	L	S	S	L	S	S	L	S	S	L	S	S
Includes GI_t			Yes			Yes			Yes			Yes
AR1	0.29	0.30	0.15	0.24	0.34	0.18	0.25	0.27	0.06	0.33	0.33	0.08
ARX1$_{(IPIt)}$	0.30	0.33	0.16	0.36	0.38	0.19	0.30	0.30	0.10	0.39	0.38	0.12
ARX1$_{(EEIt)}$	0.21	0.71	0.38	0.16	0.61	0.32	0.15	0.29	0.08	0.20	0.35	0.10
ARX1$^{(EEIt)}_{(IPIt)}$	0.32	0.76	0.42	0.39	0.65	0.34	0.31	0.32	0.11	0.40	0.40	0.15
ARX1$_{(IPIt-1)}$	0.37	0.38	0.24	0.42	0.43	0.25	0.38	0.39	0.17	0.51	0.51	0.15
ARX1$_{(EEIt-1)}$	0.39	0.18	0.23	0.36	0.28	0.17	0.38	0.21	0.07	0.49	0.25	0.10
ARX1$^{(EEIt-1)}_{(IPIt-1)}$	0.41	0.36	0.38	0.45	0.45	0.31	0.34	0.34	0.17	0.45	0.45	0.16
ARX1 $_{(IPIM1t)}$	0.30	0.37	0.24	0.36	0.41	0.27	0.32	0.33	0.19	0.41	0.42	0.18
ARX1 $_{(EEIM1t)}$	0.31	0.37	0.29	0.27	0.41	0.30	0.29	0.32	0.20	0.37	0.41	0.18
ARX1 $^{(EEIM1t)}_{(IPIM1t)}$	0.37	0.54	0.34	0.44	0.53	0.35	0.42	0.46	0.26	0.55	0.56	0.24
ARX1 $_{(IPIM1t-1)}$	0.41	0.41	0.30	0.46	0.46	0.33	0.47	0.47	0.34	0.62	0.62	0.30
ARX1 $_{(EEIM1t-1)}$	0.27	0.27	0.22	0.31	0.31	0.23	0.23	0.23	0.14	0.30	0.30	0.14
ARX1 $^{(EEIM1t-1)}_{(IPIM1t-1)}$	0.50	0.50	0.35	0.58	0.58	0.41	0.61	0.61	0.46	0.79	0.79	0.35

Table 4: MSE for two steps ahead forecasts, rolling estimates.

Notes: GI is the quarterly average of weekly Google Index data, summarizing internet job search intensity; IPI is the quarterly average of the monthly Industrial Production Index released by ISTAT (2009a); EEI is the quarterly average of the monthly economy-wide Employment Expectations obtained weighting the sector specific values released by European-Commission (2009) with sector specific number of employed. Models estimated on in-samples of different lengths. Models estimated on the full sample (first column of each group; 90 or more observations) start the in-sample interval with 1985:1. Variables with subscript M1 are equal to the point value of the first month of the relevant quarter instead than to the quarter average.

In sample obs.	90	14	14	91	15	15	92	16	16	93	17	17
Smpl length	L	S	S	L	S	S	L	S	S	L	S	S
Includes GI_t			Yes			Yes			Yes			Yes
AR1	0.24	0.33	0.19	0.25	0.26	0.08	0.33	0.30	0.09	0.45	0.39	0.07
ARX1$_{(IPIt)}$	0.38	0.40	0.23	0.32	0.33	0.13	0.41	0.40	0.16	0.48	0.48	0.14
ARX1$_{(EEIt)}$	0.15	0.45	0.28	0.14	0.25	0.09	0.19	0.28	0.09	0.23	0.33	0.01
ARX1$^{(EEIt)}_{(IPIt)}$	0.40	0.50	0.31	0.32	0.34	0.15	0.42	0.41	0.17	0.42	0.42	0.05
ARX1$_{(IPIt-1)}$	0.34	0.34	0.21	0.32	0.32	0.15	0.39	0.40	0.21	0.37	0.37	0.09
ARX1$_{(EEIt-1)}$	0.36	0.26	0.19	0.38	0.18	0.09	0.49	0.19	0.07	0.69	0.21	0.01
ARX1$^{(EEIt-1)}_{(IPIt-1)}$	0.51	0.50	0.40	0.41	0.42	0.30	0.32	0.32	0.17	0.26	0.26	0.03
ARX1$_{(IPIM1t)}$	0.37	0.43	0.33	0.32	0.34	0.20	0.40	0.40	0.26	0.51	0.51	0.29
ARX1$_{(EEIM1t)}$	0.27	0.33	0.25	0.28	0.24	0.13	0.37	0.30	0.19	0.52	0.41	0.23
ARX1$^{(EEIM1t)}_{(IPIM1t)}$	0.35	0.41	0.36	0.29	0.32	0.21	0.37	0.36	0.27	0.47	0.46	0.28
ARX1$_{(IPIM1t-1)}$	0.43	0.43	0.35	0.44	0.44	0.35	0.54	0.54	0.45	0.55	0.55	0.43
ARX1$_{(EEIM1t-1)}$	0.31	0.31	0.24	0.22	0.22	0.14	0.27	0.27	0.18	0.33	0.33	0.20
ARX1$^{(EEIM1t-1)}_{(IPIM1t-1)}$	0.51	0.51	0.38	0.57	0.57	0.44	0.70	0.70	0.58	0.72	0.72	0.59

Table 5: MSE for three steps ahead forecasts, rolling estimates.

Notes: GI is the quarterly average of weekly Google Index data, summarizing internet job search intensity; IPI is the quarterly average of the monthly Industrial Production Index released by ISTAT (2009a); EEI is the quarterly average of the monthly economy-wide Employment Expectations obtained weighting the sector specific values released by European-Commission (2009) with sector specific number of employed. Models estimated on in-samples of different lengths. Models estimated on the full sample (first column of each group; 90 or more observations) start the in-sample interval with 1985:1. Variables with subscript M1 are equal to the point value of the first month of the relevant quarter instead than to the quarter average.

In sample obs.	90	14	14	91	15	15	92	16	16	93	17	17	94	18	18	95	19	19	96	20	20
Smpl length	L	S	S	L	S	S	L	S	S	L	S	S	L	S	S	L	S	S	L	S	S
Includes GI_t			Yes			Yes			Yes			Yes			Yes			Yes			Yes
AR1	0.09	0.12	0.08	0.09	0.10	0.05	0.10	0.11	0.06	0.10	0.11	0.05	0.10	0.11	0.05	0.15	0.15	0.02	0.26	0.25	0.02
ARX1$_{(IPIt)}$	0.12	0.12	0.08	0.10	0.10	0.06	0.10	0.11	0.07	0.09	0.10	0.06	0.08	0.08	0.05	0.11	0.11	0.02	0.18	0.18	0.02
ARX1$_{(EEIt)}$	0.11	0.12	0.11	0.11	0.10	0.08	0.12	0.10	0.08	0.13	0.07	0.06	0.13	0.07	0.07	0.20	0.07	0.01	0.32	0.11	0.01
ARX1$^{(EEIt)}_{(IPIt)}$	0.12	0.14	0.12	0.11	0.11	0.09	0.11	0.11	0.10	0.07	0.08	0.07	0.07	0.07	0.08	0.06	0.06	0.01	0.10	0.10	0.01
ARX1$_{(IPIt-1)}$	0.14	0.14	0.12	0.14	0.14	0.12	0.14	0.14	0.11	0.14	0.15	0.11	0.17	0.17	0.14	0.09	0.09	0.02	0.13	0.13	0.02
ARX1$_{(EEIt-1)}$	0.09	0.10	0.09	0.09	0.09	0.07	0.10	0.11	0.08	0.10	0.09	0.07	0.10	0.06	0.06	0.14	0.05	0.01	0.25	0.09	0.00
ARX1$^{(EEIt-1)}_{(IPIt-1)}$	0.14	0.13	0.12	0.15	0.14	0.13	0.14	0.14	0.13	0.14	0.15	0.13	0.16	0.16	0.15	0.05	0.05	0.01	0.05	0.05	0.01
ARX1 $_{(IPIM1t)}$	0.14	0.14	0.09	0.12	0.12	0.07	0.13	0.13	0.08	0.13	0.13	0.09	0.12	0.12	0.10	0.16	0.16	0.04	0.26	0.26	0.05
ARX1 $_{(EEIM1t)}$	0.12	0.13	0.12	0.13	0.11	0.10	0.15	0.12	0.11	0.16	0.11	0.14	0.19	0.12	0.15	0.28	0.16	0.04	0.48	0.24	0.05
ARX1 $^{(EEIM1t)}_{(IPIM1t)}$	0.16	0.16	0.15	0.14	0.14	0.13	0.14	0.15	0.14	0.15	0.15	0.16	0.13	0.13	0.15	0.16	0.17	0.04	0.25	0.25	0.05
ARX1 $_{(IPIM1t-1)}$	0.17	0.17	0.12	0.18	0.18	0.13	0.16	0.16	0.12	0.19	0.19	0.15	0.20	0.20	0.16	0.16	0.16	0.06	0.28	0.28	0.10
ARX1 $_{(EEIM1t-1)}$	0.10	0.10	0.08	0.10	0.10	0.08	0.09	0.09	0.07	0.07	0.07	0.08	0.07	0.07	0.09	0.08	0.08	0.03	0.16	0.16	0.05
ARX1 $^{(EEIM1t-1)}_{(IPIM1t-1)}$	0.16	0.16	0.13	0.17	0.17	0.13	0.14	0.14	0.11	0.16	0.16	0.13	0.15	0.15	0.13	0.09	0.09	0.03	0.17	0.17	0.06

Table 6: MSE for one step ahead forecasts, rolling estimates, all series seasonally adjusted with TRAMO SEATS.

Notes: GI is the quarterly average of weekly Google Index data, summarizing internet job search intensity; IPI is the quarterly average of the monthly Industrial Production Index released by ISTAT (2009a); EEI is the quarterly average of the monthly economy-wide Employment Expectations obtained weighting the sector specific values released by European-Commission (2009) with sector specific number of employed. Models estimated on in-samples of different lengths. Models estimated on the full sample (first column of each group; 90 or more observations) start the in-sample interval with 1985:1. Variables with subscript M1 are equal to the point value of the first month of the relevant quarter instead than to the quarter average.

In sample obs.	90	14	14	91	15	15	92	16	16	93	17	17
Smpl length	L	S	S	L	S	S	L	S	S	L	S	S
Includes GI_t			Yes			Yes			Yes			Yes
AR1	0.09	0.13	0.06	0.11	0.13	0.07	0.10	0.11	0.05	0.10	0.10	0.05
$\text{ARX1}_{(IPIt)}$	0.13	0.13	0.07	0.13	0.13	0.08	0.11	0.12	0.06	0.09	0.09	0.05
$\text{ARX1}_{(EEIt)}$	0.13	0.13	0.11	0.14	0.12	0.10	0.15	0.08	0.06	0.16	0.07	0.06
$\text{ARX1}^{(EEIt)}_{(IPIt)}$	0.15	0.14	0.11	0.14	0.13	0.11	0.09	0.09	0.07	0.07	0.07	0.07
$\text{ARX1}_{(IPIt-1)}$	0.16	0.16	0.13	0.16	0.16	0.13	0.15	0.15	0.12	0.16	0.16	0.13
$\text{ARX1}_{(EEIt-1)}$	0.10	0.12	0.09	0.11	0.13	0.09	0.11	0.11	0.07	0.11	0.06	0.05
$\text{ARX1}^{(EEIt-1)}_{(IPIt-1)}$	0.17	0.17	0.15	0.17	0.18	0.16	0.16	0.16	0.13	0.15	0.15	0.13
$\text{ARX1}_{(IPIM1t)}$	0.13	0.13	0.07	0.13	0.14	0.07	0.12	0.12	0.09	0.11	0.11	0.09
$\text{ARX1}_{(EEIM1t)}$	0.14	0.15	0.10	0.16	0.14	0.11	0.17	0.13	0.14	0.20	0.12	0.15
$\text{ARX1}^{(EEIM1t)}_{(IPIM1t)}$	0.16	0.17	0.14	0.15	0.16	0.15	0.15	0.15	0.18	0.13	0.14	0.16
$\text{ARX1}_{(IPIM1t-1)}$	0.19	0.19	0.14	0.18	0.18	0.14	0.19	0.19	0.15	0.19	0.19	0.16
$\text{ARX1}_{(EEIM1t-1)}$	0.12	0.12	0.09	0.12	0.12	0.08	0.09	0.09	0.08	0.06	0.06	0.08
$\text{ARX1}^{(EEIM1t-1)}_{(IPIM1t-1)}$	0.20	0.20	0.16	0.18	0.18	0.15	0.18	0.18	0.15	0.17	0.17	0.15

Table 7: MSE for two steps ahead forecasts, rolling estimates, all series seasonally adjusted with TRAMO SEATS.

Notes: GI is the quarterly average of weekly Google Index data, summarizing internet job search intensity; IPI is the quarterly average of the monthly Industrial Production Index released by ISTAT (2009a); EEI is the quarterly average of the monthly economy-wide Employment Expectations obtained weighting the sector specific values released by European-Commission (2009) with sector specific number of employed. Models estimated on in-samples of different lengths. Models estimated on the full sample (first column of each group; 90 or more observations) start the in-sample interval with 1985:1. Variables with subscript M1 are equal to the point value of the first month of the relevant quarter instead than to the quarter average.

In sample obs.	90	14	14	91	15	15	92	16	16	93	17	17
Smpl length	L	S	S	L	S	S	L	S	S	L	S	S
Includes GI_t			Yes			Yes			Yes			Yes
AR1	0.11	0.15	0.07	0.10	0.12	0.05	0.10	0.09	0.02	0.15	0.14	0.01
ARX1$_{(IPIt)}$	0.16	0.16	0.08	0.13	0.13	0.06	0.10	0.10	0.03	0.15	0.15	0.02
ARX1$_{(EEIt)}$	0.14	0.14	0.13	0.15	0.09	0.08	0.16	0.04	0.06	0.24	0.05	0.02
ARX1$^{(EEIt)}_{(IPIt)}$	0.16	0.15	0.13	0.10	0.10	0.08	0.05	0.05	0.06	0.06	0.06	0.01
ARX1$_{(IPIt-1)}$	0.14	0.14	0.12	0.12	0.12	0.11	0.09	0.09	0.09	0.05	0.05	0.01
ARX1$_{(EEIt-1)}$	0.11	0.14	0.11	0.11	0.11	0.08	0.10	0.03	0.03	0.15	0.03	0.01
ARX1$^{(EEIt-1)}_{(IPIt-1)}$	0.16	0.16	0.16	0.14	0.14	0.15	0.07	0.07	0.08	0.01	0.01	0.01
ARX1 $_{(IPIM1t)}$	0.16	0.16	0.08	0.13	0.13	0.08	0.10	0.10	0.06	0.15	0.15	0.01
ARX1 $_{(EEIM1t)}$	0.16	0.16	0.11	0.17	0.13	0.12	0.20	0.09	0.11	0.30	0.14	0.00
ARX1 $^{(EEIM1t)}_{(IPIM1t)}$	0.17	0.18	0.19	0.15	0.15	0.21	0.09	0.08	0.16	0.12	0.12	0.00
ARX1 $_{(IPIM1t-1)}$	0.19	0.19	0.15	0.20	0.20	0.17	0.14	0.14	0.13	0.15	0.15	0.05
ARX1 $_{(EEIM1t-1)}$	0.14	0.14	0.09	0.11	0.11	0.07	0.05	0.05	0.05	0.08	0.08	0.02
ARX1 $^{(EEIM1t-1)}_{(IPIM1t-1)}$	0.20	0.20	0.17	0.21	0.21	0.19	0.14	0.14	0.14	0.14	0.14	0.06

Table 8: MSE for three steps ahead forecasts, rolling estimates, all series seasonally adjusted with TRAMO SEATS.

Notes: GI is the quarterly average of weekly Google Index data, summarizing internet job search intensity; IPI is the quarterly average of the monthly Industrial Production Index released by ISTAT (2009a); EEI is the quarterly average of the monthly economy-wide Employment Expectations obtained weighting the sector specific values released by European-Commission (2009) with sector specific number of employed. Models estimated on in-samples of different lengths. Models estimated on the full sample (first column of each group; 90 or more observations) start the in-sample interval with 1985:1. Variables with subscript M1 are equal to the point value of the first month of the relevant quarter instead than to the quarter average.