Forecasting Births Using Google

Francesco C. Billari
University of Oxford

Francesco D'Amuri
Bank of Italy
University of Essex
and
IZA

Juri Marcucci*
Bank of Italy

Draft prepared for presentation at PAA Annual Meeting, 2013, New Orleans
Session 155: Methods and Models in Fertility Research
April 6, 2013

Abstract

Monitoring fertility change is particularly important for policy and planning purposes. New data may help us in this monitoring. We propose a new leading indicator based on Google web-searches. We then test its predictive power using US data. In a deep out-of sample comparison we show that popular time series specifications augmented with web-search-related data improve their forecasting performance at forecast horizons of 6 to 24 months. The superior performance of these augmented models is confirmed by formal tests of equal forecast accuracy and superior predictive ability. Moreover, our results survive a *falsification test* and are confirmed also when a forecast horse race is conducted using different out-of-sample tests, and at the state rather than at the federal level. Conditioning on the same information set, the forecast error of our best model for predicting 2009 births is 35% lower than the Census bureau projections. Our findings indicate the potential use of Googe web-searches in monitoring fertility change and in informing fertility forecasts.

Keywords: fertility forecasting, US birth rates, Google econometrics, Forecast comparison, Keyword search.

*The views expressed are those of the authors and do not necessarily reflect those of the Bank of Italy. We wish to thank Joshua Goldstein, Emilio Zagheni and seminar participants at Bocconi University and the Max Planck Institute for Demographic Research for their very useful comments. Emails: francesco.billari@nuffield.ox.ac.uk (Francesco C. Billari), francesco.damuri@gmail.com (Francesco D'Amuri), juri.marcucci@gmail.com (Juri Marcucci, corresponding author). Correspondence address: Bank of Italy, Economic Research Department, Via Nazionale 91, 00184, Rome, Italy.

1 Introduction

In advanced societies (i.e., societies in which birth control is the default option), fertility change has challenged researchers for decades (e.g., Balbo et al., 2013). Recent shifts in general fertility trends (Myrskylä et al., 2009; Goldstein et al., 2009) and the effect of the economic recession (Sobotka et al., 2011) have outlined the relevance of both turning points and short- to medium-term fluctuations in fertility. One way to read this literature is that birth rates have come to be considered as one of the (multiple) indicators of the general well-being of societies. Monitoring birth rates, including its short-term (e.g., monthly) fluctuations becomes more important than it was earlier (Sobotka et al., 2005). Monthly fertility variation has therefore been proposed as the key building block of fertility monitoring (Sobotka et al., 2005).

As fertility choices are increasingly considered purposive behaviors, the role of intentions has become more central in the literature (Schoen et al., 1999; Balbo et al., 2013), despite earlier caveats on the predictive validity of fertility intentions, especially when measured referring to aggregated measures such as ideal family size (Ryder and Westoff, 1977; Quesnel-Vallée and Morgan, 2003). The proactive seeking of information, as it is now possible at low cost through the internet, and search engines in particular, can be seen as a signal of fertility intentions on a specific parity and over a time interval, i.e. the type of fertility intentions that are considered to be as more likely to be linked to actual behavior (Miller and Pasta, 1995; Schoen et al., 1999; Philipov, 2009). Such proactive search can therefore inform monitoring and forecasting fertility.

In general, forecasting fertility is not an easy exercise. This certainly true, and even well-known, for the medium and long-term, where demographic forecasting has put a great emphasis on the great uncertainty of fertility forecasting (Alkema et al., 2011) and on the large size of predictive intervals. An instructive example can be given by considering Lee's (1993) long-term fertility forecast for the US Total Fertility Rate in 2065. Lee's forecast has a 95 percent predictive interval ranging from .8 to 3.2 children per woman (see also Lee, 1998). In assessments of probabilistic population forecasts, fertility forecasting has been shown as the biggest source of error (Alho, 1992) and the fact that "Fertility forecasting is the weak point of stochastic population forecasts" has been explained by the fact that changing fertility trends account for large forecasting errors "even a few years ahead" (Ortega and Poncela, 2005. See also Booth, 2006).

For all these reasons, forecasting fertility for the shorter term (in a demographic sense) of one or two years, with a 2-year interval being often considered medium term in economic forecasts is particularly important. In this paper, we propose a set of indicators based on web searches (Google Indicators) that can be useful to forecast births over a 2-year horizon. Quantitative data on the volume and content of web-searches are becoming quickly available and are starting to be used extensively in social sciences. To the best of our knowledge this is the first study exploiting this rich and timely data source of Google web-searches to anticipate demographic trends.

Albeit in principle our approach starts from the purposive fertility decision-making of individuals and couple who seek information on fertility, pregnancy, parenthood and related issue, this approach does not need to rely on individual searches. Building on other pieces of research using Google Indicators, the monthly variation in fertility-related relative search intensity is sufficient to build the leading indicator that allows to contribute to the prediction of subsequent birth rates.

In our empirical analyses, we present a deep out-of-sample exercise where we compare basic time series specifications for forecasting monthly US birth rates (also including exogenous economic variables such as the GDP growth rate or the unemployment rate) with a set of otherwise identical models augmented with data capturing the intensity of web-searches broadly related to fertility decisions. In particular, we track the volume of Google web-searches for the keywords 'maternity', 'pregnancy' and 'ovulation'. In the forecasting exercise we also use the first principal component, which might be seen as a leading indicator of fertility sentiment. We find that models augmented with web-based search data significantly outperform standard ones at horizons up to two years. When forecasting alternatively at 6, 12, 18 and 24 months ahead, the best model in terms of lowest Mean Squared Error always includes a web-based indicator among the exogenous variables.

Our results hold for different out-of-sample intervals and are confirmed by formal tests of equal forecast accuracy such as the Diebold and Mariano's (1995) test and the Model Confidence Set (MCS) test by Hansen, Lunde and Nason (2011). Furthermore, our results survive a *falsification test* and are also confirmed when forecast horse races are conducted at the state rather than at the federal level. Standard time series specifications augmented with web-search related data can therefore be a useful tool for fertility forecasting, especially when the desire is to detecting turning points, traditionally one of the weakest points in forecasting birth rates (Booth, 2006). Conditioning on the same information set, the forecast error of our best model for 2009 births is 35% lower than the projection of the Census bureau.

The innovative data source exploited in this paper has previously been used mainly in epidemiology and economics (Edelman, 2012). The first study to use it (Ginsberg et al. 2009), estimates the weekly flu spread in the US exploiting the pattern of influenza-related web queries. Choi and Varian (2012) used web search data to forecast consumer behaviour (Retail and Automotive sales, Travel destinations) and the initial claims for unemployment benefits in the US. Askitas and Zimmermann (2009), D'Amuri and Marcucci (2012) and Suhoy (2009) used Google-based data to forecast labour market dynamics respectively in Germany, the US and Israel. Also Central Banks are starting to publish reports on the suitability of these data for complementing more standard indicators of economic activity (Troy et al., 2012; McLaren and Shanbhogue, 2011).

The remainder of this paper is organized as follows: in Section 2 we describe the data used to predict US monthly birth rates, with a particular emphasis on the web-search indicators that are specifically developed for this study. In Section 3 we discuss forecasting models. In Section 4 we compare the out-of-sample performance of forecasting models in terms of Mean Squared Errors on two different out-of-sample intervals. In Section 5 we show the results of formal tests of forecast accuracy, while Section 6 provides additional robustness checks. Section 7 compares our best forecasts with the US Census Bureau fertility projections, while Section 8 concludes.

2 Data and the fertility 'Google Index'

The data used in this paper come from different sources[1] . Birth rates are constructed as the ratio of the monthly data on the number of births published by the US Department of Health and Human Services[2] and the monthly population aged 16 and more published by the Bureau of Labor Statistics (therefore, our birth rates exclude the population aged 0-15, which is usually included in the calculation of standard crude birth rates).[3] Birth data cover the time span from January 1990 (1990:1) to December 2009 (2009:12). In our forecasting exercise we also use the year-on-year growth rates of the quarterly GDP taken from the Federal Reserve Economic Data published by the St. Louis Fed[4] and of the monthly unemployment rate published by the Bureau of Labor Statistics.[5]

We create a specific fertility Google Index (GI) which summarizes fertility-related searches performed through the Google web search engine during a given time interval. Web search data are available almost in real time starting with the week ending on January 10, 2004 and report the incidence of queries using a specific keyword on total queries performed through Google in the relevant week. In this study, after some explorations, we decided to focus on three keywords: *'maternity'* (GI1), *'ovulation'* (GI2) and *'pregnancy'* (GI3), as indicators of fertility-related interests of web users. The values of the index, available free of charge,[6] are normalized with a value equal to 100 for the week with the highest incidence.

We chose the keywords *'maternity'* and *'pregnancy'* since we believe that they are popular and broadly used searches for people who want to get information on giving birth. We also used the word *'ovulation'*, as an indirect indicator of the intention to give birth as a proactive behavior. Under the assumption that the intention to give birth follows the same dynamics in

[1]All time series are seasonally adjusted using Tramo-Seats (Maravall, 2006).

[2]Data are available at http://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm The data used in this paper were downloaded on December 10th, 2010.

[3]Data are available at http://data.bls.gov/pdq/querytool.jsp?survey=In and were downloaded on January 13th, 2011.

[4]Data are available at http://research.stlouisfed.org/fred2/ and were downloaded on January, 13th, 2011.

[5]Data are available at http://data.bls.gov/pdq/querytool.jsp?survey=In and were downloaded on December 10th, 2010.

[6]Google data are available from www.google.com/insights/search/. The data used in this article were downloaded on January, 13th 2012.

fertile and infertile couples, movements in the utilization of this keyword could still be a leading indicator for the birth rate we want to forecast in this paper. In our analyses, we normalize each index in order for it to be comparable to the one for *'maternity'*; in other words, a value of 120 for *'pregnancy'* in a given month means that the volume of searches for this keyword was 20% higher than the all-time peak reached by the volume of searches for *'maternity'*. A single aggregate indicator that we also use is built using the first Principal Component of the three keywords–this single indicator could be defined as the fertility 'Google Index'.

Table 1 reports the descriptive statistics for the *birth rate*, the GDP year-on-year growth rate and the unemployment rate. The birth rate has been ranging between 14 and 19 per cent in the period 1990:1-2009:12, averaging around 16%. Among the GIs we propose, the keyword attracting the highest volume of searches is 'pregnancy' (i.e. GI3) that was on average almost 7 times more searched than 'maternity' (i.e. GI1) and 8 times more searched than 'ovulation' (i.e. GI2).

In Figure 1 we plot the birth rate and the two variables conveying the information related to the macroeconomic conditions (i.e. GDP and unemployment rate). Two facts are apparent from the figure: i) there is a decreasing trend in the birth rate; ii) this trend accelerates with the drop in GDP and the consequent increase in the unemployment rate taking place during the Great Recession (see Sobotka et al., 2011 for a general interpretation of the effect of unemployment on fertility with a particular emphasis on the Great Recession). As a first exploration, table 2 confirms the presence of a strong cross-sectional correlation of the expected sign between the birth rate and the GDP (0.70) and unemployment rate dynamics (-0.91). Moving to our Google search indexes, a strong cross-sectional correlation (0.79) is found between the *birth rate* and the GI for the keyword *'maternity'* (GI1), with smaller values found for *'pregnancy'* and *'ovulation'* (0.37 and -0.02, respectively). The four panels of Figure 2 show separately the evolution over time of the birth rate and each of the four Google-related indices (the three keywords plus their first principal component or the fertility "Google Index"- GIPC1) over the shorter time interval in which the indices are available (2004:1-2010:12). The keyword *'maternity'* (upper left panel) matches extremely well birth rate dynamics, confirming the highest cross-sectional correlation examined in Table 2, while for the other two keywords (*'ovulation'* and *'pregnancy'*) and the fertility 'Google Index', the similarity is less evident.

Before proceeding with testing the usefulness of the GIs in forecasting, we check for non-stationarity of the US birth rate by computing a robust univariate unit root test for the integration of the series. We perform the Augmented Dickey-Fuller test with GLS de-trending (ADF-GLS) suggested by Elliott et al. (1996). This test is similar to the more standard Dickey-Fuller t test but it applies GLS de-trending before the series is tested with the ADF test. Compared with the standard ADF test, ADF-GLS test has the best overall performance in terms of small-sample size and power. Table 3 reports the results of this unit root test both considering a constant (superscript μ) and a constant and trend (superscript τ) as exogenous regressors. We run these tests both for the full sample, i.e. 1990.1-2010.12, and for the short sample, i.e. 2004.1-2010.12.

Looking at the birth rate (br_t), both ADF-GLS tests (with constant and with constant and trend) fail to reject the null of a unit root in the full and in the short sample. This result is not completely new: the time series of the birth rate is commonly considered as non-stationary (Booth, 2006; for an application, see Ermisch, 1988). According to the test, the time series of the unemployment rate is also non stationary, while the GDP growth rate is always stationary. As for GI data, the only series for which the tests reject the null of the presence of a unit root is the one for the keyword *'pregnancy'*. As a consequence, we conduct our forecasting exercise on the birth rate series in first differences. As for the economic variables and leading indicators, we use the levels. For GDP growth rate the choice can be justified through its stationarity, while for the unemployment rate the results of the unit root tests must be interpreted with caution. Actually, the unemployment rate is a bounded variable by definition (between 0 and 1 or 0 and 100) and nevertheless very often unit root tests cannot reject the null of a unit root or explosive behavior. For this reason, we decide to treat the unemployment rate as stationary. For the GIs, the sample is clearly too short to take the unit root test results at their face value. Even from Figure 2 we can notice that there is no explosive behavior for the Google indexes and, beyond that, they are also bounded between 0 and 100 by construction.

3 Forecasting fertility up to 24 months ahead

In our forecasting exercise we estimate a wide variety of linear models for $d(br_t) = br_t - br_{t-1}$, which denotes the first differences of the US birth rate. We compare the performance of these models in a horse race at four forecasting horizons: 6, 12, 18 and 24 months ahead. The out-of-sample prediction starts with 2007.1 and covers 36 months (see Section 4.1; as a robustness check, in Section 4.2 we also use a smaller evaluation sample starting with 2008.1). In all our forecasting exercises we use a rolling window.[7] In particular, as a starting point we estimate a set of AR and ARMA models that only use (the first differences) of birth rates as predictors, from the 1st to 12th lag for both the dependent variable and the moving average component.

In addition, we augment these basic specifications with macroeconomic indicators, estimating ARMAX(p, q) models, i.e. :

$$\phi(L)u_t = x_t'\beta + \theta(L)\varepsilon_t \tag{1}$$

where x_t' is a vector with a first column of ones and one or more columns of leading indicators. In particular, we add to the basic AR(MA) specifications either the unemployment rate or the GDP growth rate or both variables, with a lag structure that is identical to the one for the lagged dependent variable (from one to twelve).

Finally, to test the predictive power of fertility Google Indicators we also estimate each of these models adding as predictors one of the four GIs, each time selection only one lag (going from one to twelve) exploring all possible combinations of models.

In our pseudo-out-of-sample exercise we consider the situation that the real forecasters face when they produce their forecasts and the future values of the exogenous variables (x_t) need to be forecast. At any given date, we run our forecasting horse race using only the information that was actually available at that time. Therefore, we use simple AR(1) models to predict x_t, so that we could use such predictions as inputs in our forecasting models. As a robustness

[7]We also performed a forecasting horse race using a recursive scheme. The results are similar to those with a rolling scheme and are not reported for the sake of brevity, but they are available upon request.

check, we have also used ARMA(1,1) models, but since the main results were unchanged, we only present evidence obtained using a simple AR(1).

The inclusion of the fertility Google Indicators among the predictors necessarily restricts the length of the estimation interval, given that these indicators are available since 2004.1 only. An exercise comparing the forecasting performance of models estimated on samples starting in 2004.1 could be able to assess the predictive power of the GI, but it would be of little practical relevance if models estimated on the longer sample were better at predicting the birth rate dynamics. To tackle this issue we estimate the bunch of models just outlined above both on a 'Short Sample' (from 2004.1 onwards) for which GI data are available, and on a 'Long Sample' starting with 1990.1. When estimating GI-augmented models over this longer time span, we add a step dummy for the period in which web-search data are not available (i.e. 1990.1-2003.12). In the next section we present the results of the Out-of-Sample Forecasting comparison.

4 Out-of-Sample forecasting comparison

4.1 January 2007-December 2009

In Table 4 we present the results of our forecast comparison with forecast horizons at 6, 12, 18 and 24 months. For each forecast horizon the column labeled 'Rank' gives the rank of each model in terms of the lowest MSE, while the columns labeled 'Lag k' and 'Lag m' report the number of lags for the dependent variable and for the GI variable (if present), respectively.

We can notice that for all forecast horizons the best model (i.e. the model with the lowest MSE out-of-sample) always includes at least one GI as the exogenous variable at any forecast horizon. In particular, the $ARMA(10, 10)$, augmented with the GI for 'maternity' with three lags (i.e. $GI1_{t-3}$) and estimated on the Long sample is the best model when forecasting at 12 months ahead. The MSE of the out-of-sample forecasts obtained with this model is 14% lower than that obtained with the best model not employing GIs, an ARMAX(2,2) with the second lag of the GDP growth rate, ranking fourth in our forecasting horse race. Always at 12 steps ahead, the second best model not employing GIs (an ARMA(2,2) augmented with GDP growth rate and UR with 2 lags) ranks 16th in the horse race: compared to this model, the best model with a GI has a MSE that is lower by 44%.

When looking 24 months ahead, the best model is again Google-augmented with an indicator based on the searches for the keyword 'pregnancy' (an $ARMA(7,7)$ with the GI for 'pregnancy' with three lags ($GI3_{t-3}$)), and has a MSE that is 53% smaller than the best one without GIs, that is the ARMA(8,8) with no exogenous variable. In this case, the best model without GIs ranks 5th in the forecast comparison. The second best model not employing GIs is an ARMA(2,2), with GDP growth rate as additional exogenous variable, ranking 12th in the forecasting horse race; the benchmark model with GIs has a MSE that is 82% lower than this latter model. The horse-race forecast comparison yields similar results when forecasting alternatively at 6 or 18 steps ahead: the best model always include a GI and has a substantially lower MSE compared to models not employing web-based indicators.

4.2 January 2008-December 2009

As a first robustness test, we check the reliability of our results using a shorter out of sample interval covering the 24 months between January 2008 and December 2009 (Table 5). In this case we can exploit 12 more observations for the web-search related time series employed for the in-sample estimation but, as the evaluation sample shrinks accordingly, we can only forecast up to 18 months ahead. The specifications adopted in terms of lags of both the dependent and independent variables are the same as in the previous forecast comparison.

Results obtained in this case are even reinforced when compared to the ones described in the previous section, in which forecast where evaluated on the 2007:1-2009:12 evaluation sample. Also in this case at 6, 12 and 18 months ahead the best model always include a GI. Additionally, all the best 15 models in the forecast comparison at each forecast horizon are indeed based on web-search data. The best 'no GIs' model ranks 28th at 6 and 12 steps ahead and 49th at 18 steps ahead, while its MSE is higher by 221, 177 and 644%, respectively, when compared to the best model with a GI at the corresponding forecast horizon.

In this case, even GI-augmented models estimated on the short sample for which web-search related data are available (2004.1 onwards) have a superior forecast performance compared to models that do not use this kind of information and are alternatively estimated either on the Long or the Short sample.

A similar picture can be drawn from Figure 3 where we depict the forecast errors of 12-month-ahead forecasts for the longer out-of-sample (2007.1-2009.12, upper panel) and the shorter one (2008.1-2009.12, lower panel). In both charts we have depicted the best GI and no GI model for the forecasting exercise with a longer out-of-sample (i.e. models n. 207 and 14, respectively) and for that with the shorter one (i.e. models n. 3874 and 14, respectively). It is interesting to notice that the forecast errors from models using GIs are always closer to the zero line and also cross the latter more often. Models which do not employ GIs tend to over-predict the birth rate in both evaluation samples as shown by strictly negative forecast errors.

5 Comparing equal forecast accuracy

When evaluating the forecast accuracy of each model in the previous sections, we have only taken into consideration the ratios of the MSEs between a competitor model and a benchmark model. Nevertheless, after the work by Diebold and Mariano (1995) and West (1996), the community of forecasters has increasingly understood the importance of formally testing for out-of-sample equal forecast accuracy between two competing models. West (2006) provides a recent survey of the tests of equal forecast accuracy, while Busetti and Marcucci (2013) provide extensive Monte Carlo evidence on the best tests of equal forecast accuracy or forecast encompassing to be used in different frameworks. In this paper we employ a two-sided DM test for the null of equal forecast accuracy between the benchmark and the competitor.[8] We adopt the DM test even if it is usually undersized and has low power in small samples because if we can reject the null of equal forecast accuracy between any two competing models with

[8]The DM test is based on the loss differential between the benchmark (model 0) and the k-th competitor, i.e. $d_t = e^2_{0,t} - e^2_{k,t}$. To test the null of equal forecast accuracy $H_0 : E(d_t) = 0$, we employ the DM statistic $DM = P^{1/2}\bar{d}/\hat{\sigma}_{DM}$, where $\bar{d}$ is the average loss differential, P is the out-of-sample size, and $\hat{\sigma}_{DM}$ is the square-root of the long-run variance of d_t. The test is distributed as a Gaussian under the null of equal forecast accuracy between the two competing models.

the simple DM test, we can be almost sure that more powerful and also more complicated tests requiring bootstrapped or simulated critical values will certainly reject (see Busetti and Marcucci, 2013). We also use the two-sided version of this test because some models are nested and others are non-nested making the direction of the alternative hypothesis unknown. Using the two-sided version of the tests we can thus compare both nested and non-nested models, as is our case where the exogenous variable often differs from one model to another and only a small subset of models are really nested. We report the results of this test in both Tables 4 and 5 (second to last column).

When conducting our forecast comparison on the 2007.1-2009.12 evaluation sample at 12 steps ahead, the DM test fails to reject the null of equal forecast accuracy between the best web-search-based model (model n. 207, ARMA(3,3) augmented with the GI for 'maternity' at the third lag) and the best no Google model (model n. 14, ARMA(2,2) with 2 lags of the GDP growth rate). The DM test rejects the null at least at the 10% level for 13 of the 15 best models not employing Google data. We would like to emphasize again the fact that these results are extremely conservative and using a more powerful test would lead to even better results for models using Google data. A similar picture emerges at shorter (6-month-ahead) and longer (18- and 24-month-ahead) forecast horizons: the DM test rejects the null at 10% significance level in most no Google cases.

Results are even more clear-cut when the forecast comparison is done over the shorter out-of-sample (2008.1-2009.12) at 6-, 12- and 18-month-ahead. At any forecast length the DM test always rejects the null of equal forecast accuracy when comparing the Google-augmented benchmark model with the best 15 models which do not employ web-search-related data. Out of 45 models we count 44 rejections at the 1% level and one at the 5% level. This result is even more striking if we think that in this case we have a shorter evaluation sample, and the power of the test is thus limited.

Since we are comparing a large number of model-based forecasts (9408 models), it could be worthy to also obtain a joint confidence interval for all possible pairwise comparisons using a test based on multiple comparisons. We could have used the Reality Check test for data snooping by White (2000) or its improved version suggested by Hansen (2005). Both tests are based on superior predictive ability and allow for multiple comparisons against a pre-specified benchmark model. However we decided to compare the whole set of models jointly with the MCS test by Hansen, Lunde and Nason (2011) simply because we wanted to be agnostic about the choice of the benchmark model. The MCS is in fact defined as the set that contains the best models in terms of superior forecast accuracy without any assumptions about the true (benchmark) model. The MCS allows the researcher to identify, from a universe of model-based forecasts, a subset of models, equivalent in terms of superior ability, which outperform all the other competing models at a given confidence level α. The other thing to be noticed is that the MCS is a test of conditional predictive ability and therefore it allows a unified treatment of nested and non-nested models taking also into account estimation technique, parameter uncertainty, ratio of estimation and evaluation sample, and data heterogeneity.[9]

The MCS results for the longer (2007.1-2009.12) and the shorter (2008.1-2009.12) out-of-sample are reported in the last column of Tables 4 and 5.[10] Looking at Table 4 at 12-month-ahead forecast horizon, we can notice that the final MCS includes is largely dominated by Google-based models. We can also notice that at any forecast horizons the benchmark model we selected (i.e. the model with the lowest MSE) is always included in the final MCS. Furthermore, at any forecast horizons models for which we reject the null of equal forecast accuracy with the

[9] Let us denote the initial set of k-step-ahead forecasts $\mathcal{M}^0 : \{f_{i,t+k} \in \mathcal{M}^0 \quad \forall i = 1, \ldots, M\}$, where $t = 0, 1, \ldots, T-1$ and T is the forecast sample size and M is the number of models. The starting hypothesis is that all forecasts in the set $\mathcal{M}^0$ have equal forecasting performance, measured by a loss function $L_{i,t} = L(br_t, f_{i,t})$, where br_t is the birth rate and $f_{i,t}$ is the corresponding forecast at time t from model i. Let $d_{ij,t} = L_{i,t} - L_{j,t} \quad \forall i, j = 1, \ldots, M$ define the relative performance of forecast i and j. The null hypothesis for the MCS test takes the form $H_{0,\mathcal{M}^0} : E(d_{ij,t}) = 0 \; \forall i, j = 1, \ldots, M$. We use the 'range' statistic defined as $T_R = \max_{i,j \in \mathcal{M}} |t_{ij}|$ where $t_{ij} = \bar{d}_{ij} / \sqrt{\hat{var}(\bar{d}_{ij})}$ represents the standardized relative performance of forecast i with respect to forecast j, and $\bar{d}_{ij} = T^{-1} \sum_{t=1}^{T} d_{ij,t}$ is the sample average loss difference between forecast i and j. To obtain the distribution under H_0 a stationary bootstrap scheme is used. If H_0 is rejected, an elimination rule removes the forecast with the largest t_{ij}. This process is repeated until non-rejection of the null occurs, thus allowing the construction of $(1-\alpha)$-confidence set for the best forecasts in $\mathcal{M}^0$.

[10] We set the confidence level for the MCS to $\alpha = 0.05$, the block length to 10 and the number of bootstrap samples to 300. Such number might appear small but it is sufficient to identify the MCS. We did not choose a bigger number because using the range statistic we are comparing all possible pairwise combinations between model i and j and given the large number of models in our forecasting exercise a higher number of bootstrap samples would make the computation of the test more cumbersome.

DM test at 1% are never included in the final MCS. Such results are confirmed and corroborated with the shorter out-of-sample of Table 5.

6 Further robustness checks

The extensive out-of-sample comparison conducted in the previous sections showed that the best model (i.e. the model with the lowest MSE out-of-sample) always includes one of the fertility Google indicators as a predictor at four different forecast horizons and in two different out-of-samples.

In this section we provide two additional robustness checks for the forecast performance of web-search data augmented models; in particular: i) we provide a *falsification test*, testing the forecasting performance of an alternative Google-based indicator that shows the highest correlation with the birth rate in the in-sample, but captures the interest for a keyword that is unrelated to births; ii) we repeat the horse race of section 4.1 at the state level for each of the 51 US states (including Washington DC).

6.1 Falsification test

In this section we provide a falsification test for the main results of section 4.1. In particular, we test the forecast power of an alternative google-based indicator, that is chosen to be the one with the highest correlation with the time series of *birth rates* in the in-sample, but is not necessarily related to giving-birth. The identification of this keyword is made possible by the fact that Google developed a new application, called 'Google Correlate'[11] able to identify, within a specified time interval, the web searches for keywords that: i) show the highest correlation with a given keyword search ii) show the highest correlation with a given time series. In particular, we isolated the time series of the US birth rate and we used this application to find the keyword search that, among all searches conducted through the search engine, was mostly correlated with it in the in-sample (2004:1-2006:12). We found that this series was the GI for the keyword

[11] Available at www.google.com/trends/correlate/. See Mohebbi et al., 2011 for details on this application.

'KXMB', showing a correlation with the US birth rate of 0.85 in the relevant in-sample, but otherwise with no logical connection to birth-giving: KXMB is a local affiliate of CBS (one of the major US commercial broadcasting TVs) for central and western North Dakota. We use this alternative web-search indicator (labeled as GI5 in Table 4) for conducting a horse-race forecast comparison that is identical to the main one, whose results were presented in section 4.1. Looking at Table 4, we can see that models augmented with the fake GI indicator never rank among the best 15 models of the forecast comparison across all forecast horizons (6-, 12-, 18-, and 24-step-ahead), providing indirect evidence for the relevance of the web-search data when the underlying keywords have a direct link with fertility.

6.2 State level forecasts

As a further robustness check for the predictive properties of the Google Index, we estimated the same models introduced in section 4.1 for each of the 51 US states, assessing the percentage of states for which the best model in terms of lower MSE includes the GI. In this case we can only use two web-search related keywords, 'maternity' and 'pregnancy' (respectively GI1 and GI3): for the keyword 'ovulation' the indicator is not available at the state level since the underlying number of searches is not sufficiently high.[12] When we forecast the first-differenced series for the birth rate $d(b_t)$ at twelve step ahead over the 2007:1-2009:12 interval (the baseline in our forecast comparisons), the percentage of models adopting either 'maternity' or 'pregnancy' as a leading indicator among the best 5 models in terms of lowest MSE is equal to 89.6% (see Table 6). This percentage slightly goes down to 88.8% when considering the best 15 models. Even when we carry out a pairwise comparison, that is, when we compare all non-google models separately with models augmented either with GI1 or GI3 (panels B and C of Table 6), we find similar results.

[12]We are not aware of the threshold of minimum searches chosen by Google when deciding whether to publish the time series for a keyword.

7 Comparison with US Census Buread projections

The in-depth forecast comparison conducted until now has showed that basic time series specifications can significantly improve their forecast performance at short to medium prediction horizons when augmented with Google based data on the volume of web-searches related to birth-giving. Nevertheless, this result would be of little practical relevance if these forecasts did not improve on standard official population forecasts.

To check whether this is the case, we conduct an *ad hoc* forecast comparison. Conditioning on all the information available in December 2008 we compare our best forecasts obtained with GI-augmented models with the Census Bureau projection for 2009 released in 2008. In particular, we choose the best Google model according to its forecasting performance over the 2008.1-2008.12 out of sample - thus, not exploiting information available in 2009, the year of our comparison. Moreover, as in the rest of this paper, we do not use the actual realizations of the exogenous variables when forecasting out of sample, but we project them using a simple AR(1) process. In this way we are conditioning on the same information set for both forecasts. Finally, since our time series models predict the birth rate, while the Census projects the total number of births for year 2009, we forecast the total number of births by multiplying our predicted birth rates by the actual size of the population aged 16+ and we summing up the forecasts for each of the 12 months in order to obtain the total for 2009.

According to the US Department of Health and Human Services, the total number of births in 2009 was equal to 4.13 million. Our best forecast conditioned on 2008 information was equal to 4.07 million, while the Census projection was equal to 4.23 million. As a consequence, our best projection had a 35% lower forecast error compared to the Census one.

8 Conclusions

In this paper we introduced a set of fertility Google web-search indices that allow to monitor and contribute to forecasting monthly birth rates. This approach is consistent with current theories on fertility decision-making in advanced societies. These data, that track the volume of web-searches over time for a given keyword, are becoming increasingly popular in epidemiology and in economics; to the best of our knowledge this is the first study exploiting this rich and timely data source to anticipate demographic trends.

We tested the predictive power of the new indices on monthly U.S. birth rates. In a deep out-of sample comparison we showed that popular time series specifications augmented with web-search-related data definitely improve their forecasting performance, both at short to medium term horizons (from 6 to 24 months ahead) and on different out-of-samples. The superior performance of web-search data augmented specification is broadly confirmed by formal tests of equal forecast accuracy such as the Diebold-Mariano pairwise test and the Model Confidence Set test which performs a multiple comparison without specifying any benchmark model. Our results also survived a *falsification test* and are confirmed also when the forecast horse race is conducted at the state rather than at the federal level. Standard time series specifications augmented with web-search-related data can be a useful tool for short- to medium-term forecasting, in particular in detecting turning points, traditionally one of the weakest points of birth rates forecasting. Conditioning on the same information set, the forecast error of our best model for 2009 births is 35% lower than the projection of the Census bureau.

References

Alho, Juha 1992. "The magnitude of error due to different vital processes in population forecasts", *International Journal of Forecasting*, 8: 301-314.

Alkema, Leontine, Raftery, Adrian E. , Gerland, Patrick, Clark, Samuel J., Pelletier, Francois, Buettner, Thomas and Heilig, Gerhard K. 2012. "Probabilistic Projections of the Total Fertility Rate for All Countries", *Demography*, 48: 815-839.

Askitas, Nikos and Zimmermann, Klaus F. 2009. "Google Econometrics and Unemployment Forecasting", *Applied Economics Quarterly*, 55: 107-120.

Balbo, Nicoletta, Billari, Francesco C. and Mills, Melinda 2013. "Fertility in Advanced Societies: A Review of Research", *European Journal of Population*, 29: 1-38.

Benitez-Silva, Hugo, Dwyer, Debra S., Gayle, Wayne-Roy and Muench, Thomas J. 2008. "Expectations in micro data: rationality revisited", *Empirical Economics*, 34: 381-146.

Booth, Heather. 2006. "Demographic forecasting: 1980 to 2005 in review", *International Journal of Forecasting*, 22: 547-581.

Busetti, Fabio, and Marcucci, Juri. 2013. "Comparing Forecast Accuracy: A Monte Carlo Investigation", *International Journal of Forecasting*, 29: 13-27.

Choi, Hyunyoung and Varian, Hal. 2012. "Predicting the Present with Google Trends in review", *Economic Record*, 88: 2-9.

D'Amuri, Francesco and Marcucci, Juri. 2012. "Google it! Forecasting the US unemployment rate with a Google job-search index ", Bank of Italy Working papers, *forthcoming.*

Diebold, Francis X. and Mariano, Roberto S. 1995. "Comparing Predictive Accuracy", *Journal of Business & Economic Statistics*, 13: 253-263.

Edelman, B. 2012. "Using Internet Data for Economic Research", *Journal of Economic Perspectives*, 26(2): 189-206.

Elliott, Graham, Rothenberg, Thomas J., and Stock, James H. 1996. "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64: 813-836.

Ermisch, John. 1988. "Econometric analysis of birth rate dynamics in Britain", *Journal of Human Resources*, 23: 563-576.

Ginsberg, Jeremy, Mohebbi, Matthew H., Patel, Rajan S., Brammer Lynnette, Smolinski, Mark S. and Brilliant Larry. 2009. "Detecting Influenza epidemics using Search Engine Query Data", *Nature*, 457: 1012-1014.

Goldstein, Joshua R., Sobotka, Tomas and Jasilioniene, Aiva 2011. "The End of "Lowest-Low" Fertility? ", *Population and Development Review*, 35: 663-699.

Hansen, Peter R. 2005. "A Test for Superior Predictive Ability", *Journal of Business and Economic Statistics*, 23: 365-380.

Hansen, Peter R., Lunde, Asger, and James M. Nason 2011. "The Model Confidence Set", *Econometrica*, 79(2): 453–497.

Lee, Ronald D. 1993. "Modeling and forecasting the time series of US fertility: Age patterns, range, and ultimate level", *International Journal of Forecasting*, 9: 187-202.

Lee, Ronald D. 1998. "Probabilistic Approaches to Population Forecasting", *Population and Development Review*, 24 (Supplement): 156-190.

Maravall, Agustin 2006. "An application of the TRAMO-SEATS automatic procedure; direct versus indirect adjustment", *Computational Statistics & Data Analysis*, 50: 2167-2190.

McLaren, N., and Shanbhorge, R. 2011. "Using Internet Search Data as Economic Indicators ", *Bank of England Quarterly Bulletin*, Second Quarter.

Miller, Warren B. and Pasta, David J. 1995. "Behavioural Intentions: Which Ones Predict Fertility Behaviour in Married Couples? ", *Journal of Applied Social Psychology*, 25: 530-555.

Mohebbi, Matt, Vanderkam, D. and J. Kodysh and R. Schonberger and H. Choi and S. Kumar. 2011. "Google Correlate Whitepaper", *Google technical report*, June.

Myrskylä, Mikko, Kohler, Hans-Peter and Billari, Francesco C. 2009. "Advances in development reverse fertility declines", *Nature*, 460: 741-743.

Ortega, Jose Antonio and Poncela, Pilar 2005. "Joint forecasts of Southern European fertility rates with non-stationary dynamic factor models", *International Journal of Forecasting*, 21: 539-550.

Philipov, Dimiter 2009. "Fertility intentions and outcomes: The role of policies to close the gap ", *European Journal of Population*, 25: 355-361.

Quesnel-Vallée, Amélie and Morgan, S. Philip 2003. "Missing the target? Correspondence of fertility intentions and behavior in the U.S. ", *Population Research and Policy Review*, 22: 497-525.

Schoen, Robert, Astone, Nan M., Kim, Y.J., Nathanson, Constance A. and Fields, Jason M. 1999. "Do fertility intentions affect fertility behavior? ", *Journal of Marriage and the Family*, 61: 760-799.

Sobotka, Tomáš, Winkler-Dworak, Maria, Testa, Maria Rita, Lutz, Wolfgang, Philipov, Dimiter, Engelhardt, Henriette and Gisser, Richard 2005. "Monthly Estimates of the Quantum of Fertility: Towards a Fertility Monitoring System in Austria ", *Vienna Yearbook of Population Research*, 3: 109-141.

Sobotka, Tomáš, Skirbekk, Vegard and Philipov, Dimiter 2011. "Economic Recession and Fertility in the Developed World ", *Population and Development Review*, 37: 267-306.

Suhoy, Tanya. 2009. "Query Indices and a 2008 Downturn ", *Bank of Israel Discussion Paper (2009.06).*

Troy, G. and Dilhan Perera and D. Sunner 2012. "Electronic Indicators of Economic Activity", *Australian Reserve Bank - Economic Bulletin*, June 1-12.

West, Kenneth D. 1996. "Asymptotic inference about predictive ability", *Econometrica*, 64: 1067-1084.

West, Kenneth D. 2006. "Forecast Evaluation", 100-134, in *Handbook of Economic Forecasting*, Vol. 1, G. Elliott, C.W.J. Granger and A. Timmerman (eds), Amsterdam: Elsevier.

Westoff, Charles and Ryder, Norman B. 1977. "The predictive validity of reproductive intentions ", *Demography*, 4: 431-453.

White, Halbert. 2000. "A Reality Check For Data Snooping", *Econometrica*, 68(5): 1097-1126.

Table 1: Descriptive statistics of birth rate (br_t) and exogenous variables (GI, GDP, UR).

	br_t	GDP	UR	GI1	GI2	GI3
Mean	0.16	2.51	5.84	71.60	59.12	479.96
Median	0.16	2.82	5.52	73.39	60.21	483.06
Maximum	0.19	5.32	10.07	80.94	63.24	522.74
Minimum	0.14	-4.16	3.85	57.70	50.48	428.91
Std.	0.0097	1.91	1.51	6.63	2.82	20.39
Skewness	0.88	-1.38	1.24	-0.68	-1.24	-0.61
Kurtosis	2.99	5.23	4.04	2.22	4.10	2.98
Observations	240	252	252	84	84	84

Notes: Br_t, GDP and UR are expressed in percentage points. GI1 is the monthly average of the google index for 'maternity', GI2 is the monthly average of the GI for 'ovulation', and GI3 is the monthly average of the GI for 'pregnancy'. The sample for br_t is 1990:1-2009:12. The GI index takes a value of a 100 in the week in which the ration between the number of searches for the keyword 'maternity' was the highest. We normalize each index in order for it to be comparable for the one for *'maternity'*; in other words, a value of 120 for *'pregnancy'* in a given months means that the volume of searches for this keyword was 20% higher than the all-time peak reached by the volume of searches for *'maternity'*. The sample for GDP YoY growth rate and UR is 1990:1-2009:12, while the sample for the google indices is 2004:1-2009:12.

Table 2: Correlation of birth rate (br_t) vs all exogenous variables (GI, GDP, UR).

	Short Sample: 2004:1-2010:12					
	br_t	GDP	UR	GI1	GI2	GI3
br_t	1.00					
GDP	0.70	1.00				
UR	-0.91	-0.76	1.00			
GI1	0.79	0.71	-0.86	1.00		
GI2	-0.02	-0.46	-0.05	-0.01	1.00	
GI3	0.37	0.18	-0.42	0.55	0.36	1.00
	Long Sample: 1990:1-2009:12					
	br_t	GDP	UR	GI1	GI2	GI3
br_t	1.00					
GDP	0.14	1.00				
UR	0.13	-0.59	1.00			

Notes: GI1 is the monthly average of the google index for 'maternity', GI2 is the monthly average of the GI for 'ovulation', and GI3 is the monthly average of the GI for 'pregnancy'. The sample for br_t is 1990:1-2009:12. The GI index takes a value of a 100 in the week in which the ration between the number of searches for the keyword 'maternity' was the highest. We normalize each index in order for it to be comparable for the one for *'maternity'*; in other words, a value of 120 for *'pregnancy'* in a given months means that the volume of searches for this keyword was 20% higher than the all-time peak reached by the volume of searches for *'maternity'*. The sample for GDP YoY growth rate and UR is 1990:1-2010:12, while the sample for the google indices is 2004:1-2010:12.

Table 3: Unit Root tests for birth rate (br_t) and all dependent variables (GI, GDP, UR).

Sample: 1990:1-2009:12			Sample: 2004:1-2009:12		
Variable	Test	Test stat.	Variable	Test	
br_t	$DF-GLS^{\mu}$	1.440	br_t	$DF-GLS^{\mu}$	0.003
	$DF-GLS^{\tau}$	-1.350		$DF-GLS^{\tau}$	-1.107
GDP_t	$DF-GLS^{\mu}$	-3.656***	GDP_t	$DF-GLS^{\mu}$	-2.432**
	$DF-GLS^{\tau}$	-3.867***		$DF-GLS^{\tau}$	-3.289**
ur_t	$DF-GLS^{\mu}$	-0.372	ur_t	$DF-GLS^{\mu}$	-0.601
	$DF-GLS^{\tau}$	-1.006		$DF-GLS^{\tau}$	-1.294
			$GI1_t$	$DF-GLS^{\mu}$	-0.417
				$DF-GLS^{\tau}$	-1.806
			$GI2_t$	$DF-GLS^{\mu}$	-0.503
				$DF-GLS^{\tau}$	-1.062
			$GI3_t$	$DF-GLS^{\mu}$	-2.563**
				$DF-GLS^{\tau}$	-3.042*
			$GIPC1_t$	$DF-GLS^{\mu}$	-1.470
				$DF-GLS^{\tau}$	-1.775

Notes: The $DF-GLS^{\mu}$ test indicates the test where a constant is included as the exogenous regressor, while $DF-GLS^{\tau}$ is the test with a constant and trend included. The critical values at 1, 5, and 10% for the $DF-GLS^{\mu}$ test are -2.574 (-2.593), -1.942 (-1.945) and -1.616 (-1.614), respectively, for the full sample 1990:1-2009:12 (short sample 2004:1-2009:12). Instead, the critical values at 1, 5, and 10% for the $DF-GLS^{\tau}$ test are -3.465 (-3.641), -2.921 (-3.081) and -2.624 (-2.788), respectively, for the full sample 1990:1-2009:12 (short sample 2004:1-2009:12). ***, ** and * indicate rejection at 1, 5 and 10%, respectively, of the null hypothesis of a unit root.
GI1 is the monthly average of the google index for 'maternity', GI2 is the monthly average of the GI for 'ovulation', and GI3 is the monthly average of the GI for 'pregnancy'. The sample for br_t is 1990:1-2009:12. The GI index takes a value of a 100 in the week in which the ratio between the number of searches for the keyword 'maternity' was the highest. We normalize each index in order for it to be comparable to the one for *'maternity'*; in other words, a value of 120 for *'pregnancy'* in a given month means that the volume of searches for this keyword was 20% higher than the all-time peak reached by the volume of searches for *'maternity'*. The sample for GDP YoY growth rate and UR is 1990:1-2009:12, while the sample for the google indices is 2004:1-2009:12.

Table 4: Forecasting US birth rate in first differences ($d(br_t)$) - 2004-06 with AR(1) auxiliary model with 'false' Google index. Best 15 models at different forecast horizons (6, 12, 18, 24 months ahead) with and without GI.

Num	Model	Sample	Lag k	Lag m	GDP	UR	GI1	GI2	GI3	GI4	GI5	Google	MSE	rank	DM	MCS
							6 step - Overall - GI1									
3661	ARMA	Long	10	1						✓		✓	1.75E-06	1	0.00	✓
350	ARMA	Long	10	2	✓		✓					✓	1.75E-06	2	0.00	✓
3577	ARMA	Long	3	1						✓		✓	1.76E-06	3	0.02	✓
3625	ARMA	Long	7	1						✓		✓	1.89E-06	4	0.47	✓
349	ARMA	Long	10	1	✓		✓					✓	1.89E-06	5	0.19	✓
3562	ARMA	Long	1	10						✓		✓	2.02E-06	6	0.39	✓
14	ARMA	Long	2	0	✓								2.02E-06	7	0.44	✓
2473	ARMA	Long	7	1					✓			✓	2.04E-06	8	0.93	✓
205	ARMA	Long	10	1			✓					✓	2.09E-06	9	0.44	✓
206	ARMA	Long	10	2			✓					✓	2.13E-06	10	0.52	✓
207	ARMA	Long	10	3			✓					✓	2.15E-06	11	0.73	✓
3662	ARMA	Long	10	2						✓		✓	2.21E-06	12	1.49	✓
3626	ARMA	Long	7	2						✓		✓	2.25E-06	13	1.68*	✓
3580	ARMA	Long	3	4						✓		✓	2.32E-06	14	1.01	✓
214	ARMA	Long	10	10			✓					✓	2.40E-06	15	1.11	✓
							6 Step - No GI									
14	ARMA	Long	2	0	✓								2.02E-06	7	0.44	✓
38	ARMA	Long	2	0	✓	✓							2.46E-06	17	1.08	✓
26	ARMA	Long	2	0		✓							2.92E-06	42	1.55	✓
4707	ARMA	Short	3	0									4.04E-06	117	3.28***	
61	AR	Long	1	0	✓								4.17E-06	127	1.98**	
62	AR	Long	2	0	✓								4.34E-06	144	2.06**	
37	ARMA	Long	1	0	✓	✓							4.94E-06	193	1.18	
85	AR	Long	1	0	✓	✓							5.02E-06	204	2.15**	
13	ARMA	Long	1	0	✓								5.09E-06	213	1.52	
49	AR	Long	1	0									5.28E-06	245	2.64***	
86	AR	Long	2	0	✓	✓							5.29E-06	246	2.23**	
15	ARMA	Long	3	0	✓								5.32E-06	249	2.49**	
4719	ARMA	Short	3	0	✓								5.42E-06	266	3.62***	
50	AR	Long	2	0									5.43E-06	268	2.68***	
18	ARMA	Long	6	0	✓								5.50E-06	274	2.66***	
							12 Step - Overall									
207	ARMA	Long	10	3			✓					✓	2.60E-06	1	0.00	✓
3661	ARMA	Long	10	1						✓		✓	2.78E-06	2	0.23	✓
206	ARMA	Long	10	2			✓					✓	2.88E-06	3	0.50	✓
14	ARMA	Long	2	0	✓								3.03E-06	4	0.49	✓
2473	ARMA	Long	7	1					✓			✓	3.11E-06	5	0.54	✓
3662	ARMA	Long	10	2						✓		✓	3.53E-06	6	1.09	✓
3663	ARMA	Long	10	3						✓		✓	3.65E-06	7	1.09	✓
3665	ARMA	Long	10	5						✓		✓	3.67E-06	8	0.90	✓
3626	ARMA	Long	7	2						✓		✓	4.04E-06	9	1.19	✓
3664	ARMA	Long	10	4						✓		✓	4.26E-06	10	1.46	
205	ARMA	Long	10	1			✓					✓	4.34E-06	11	1.45	
326	ARMA	Long	8	2	✓		✓					✓	4.38E-06	12	1.81*	
350	ARMA	Long	10	2	✓		✓					✓	4.46E-06	13	2.31**	
1357	ARMA	Long	10	1				✓				✓	4.51E-06	14	1.75*	✓
3625	ARMA	Long	7	1						✓		✓	4.54E-06	15	1.34	✓
							12 Step - No GI									
14	ARMA	Long	2	0	✓								3.03E-06	4	0.49	✓
38	ARMA	Long	2	0	✓	✓							4.69E-06	16	2.06**	
26	ARMA	Long	2	0		✓							5.70E-06	24	2.85***	
2	ARMA	Long	2	0									8.22E-06	52	1.24	
61	AR	Long	1	0	✓								9.61E-06	78	2.86***	
62	AR	Long	2	0	✓								1.00E-05	85	2.93***	
13	ARMA	Long	1	0	✓								1.03E-05	88	1.74*	
37	ARMA	Long	1	0	✓	✓							1.12E-05	106	1.46	
49	AR	Long	1	0									1.21E-05	123	3.62***	
50	AR	Long	2	0									1.25E-05	135	3.68***	
15	ARMA	Long	3	0	✓								1.30E-05	151	3.29***	
1	ARMA	Long	1	0									1.34E-05	161	2.24**	
25	ARMA	Long	1	0		✓							1.34E-05	162	1.82*	
85	AR	Long	1	0	✓	✓							1.40E-05	174	3.35***	
65	AR	Long	5	0	✓								1.45E-05	189	3.31***	
							18 Step - Overall									
3665	ARMA	Long	10	5						✓		✓	4.00E-06	1	0.00	✓
3662	ARMA	Long	10	2						✓		✓	4.92E-06	2	1.16	✓
3664	ARMA	Long	10	4						✓		✓	5.09E-06	3	1.74*	✓
3663	ARMA	Long	10	3						✓		✓	5.15E-06	4	2.14**	✓
2510	ARMA	Long	10	2					✓			✓	5.35E-06	5	0.83	✓
2473	ARMA	Long	7	1					✓			✓	5.61E-06	6	0.89	✓
3661	ARMA	Long	10	1						✓		✓	5.68E-06	7	1.40	✓
207	ARMA	Long	10	3			✓					✓	6.76E-06	8	1.44	✓
206	ARMA	Long	10	2			✓					✓	7.71E-06	9	1.92*	✓
14	ARMA	Long	2	0	✓								7.78E-06	10	1.77*	✓
350	ARMA	Long	10	2	✓		✓					✓	8.62E-06	11	2.78***	
211	ARMA	Long	10	7			✓					✓	8.88E-06	12	2.16**	✓

(Continued on next page)

Table 4 − continued

Num	Model	Sample	Lag k	Lag m	GDP	UR	GI1	GI2	GI3	GI4	GI5	Google	MSE	rank	DM	MCS
2509	ARMA	Long	10	1					✓			✓	9.04E-06	13	1.81*	✓
205	ARMA	Long	10	1			✓					✓	9.37E-06	14	2.99***	
214	ARMA	Long	10	10			✓					✓	9.51E-06	15	3.59***	
18 Step - No GI																
14	ARMA	Long	2	0	✓								7.78E-06	10	1.77*	✓
38	ARMA	Long	2	0	✓	✓							1.24E-05	17	3.49***	
26	ARMA	Long	2	0		✓							1.48E-05	27	3.48***	
2	ARMA	Long	2	0									2.03E-05	56	1.50	
61	AR	Long	1	0	✓								2.11E-05	59	3.90***	
62	AR	Long	2	0	✓								2.23E-05	65	4.01***	
13	ARMA	Long	1	0	✓								2.44E-05	81	2.07**	
49	AR	Long	1	0									2.50E-05	87	4.31***	
50	AR	Long	2	0									2.58E-05	94	4.24***	
15	ARMA	Long	3	0	✓								3.08E-05	147	4.61***	
37	ARMA	Long	1	0	✓	✓							3.12E-05	152	1.98**	
65	AR	Long	5	0	✓								3.38E-05	184	4.83***	
85	AR	Long	1	0	✓	✓							3.43E-05	190	5.41***	
1	ARMA	Long	1	0									3.46E-05	194	2.85***	
3	ARMA	Long	3	0									3.48E-05	196	4.69***	
24 Step - Overall																
2473	ARMA	Long	7	1					✓			✓	1.41E-06	1	0.00	✓
211	ARMA	Long	10	7			✓					✓	1.72E-06	2	0.29	✓
3663	ARMA	Long	10	3						✓		✓	2.40E-06	3	1.40	✓
3665	ARMA	Long	10	5						✓		✓	2.49E-06	4	1.13	✓
8	ARMA	Long	8	0									3.02E-06	5	0.96	✓
3664	ARMA	Long	10	4						✓		✓	3.09E-06	6	1.60	
3662	ARMA	Long	10	2						✓		✓	3.89E-06	7	2.24**	
207	ARMA	Long	10	3			✓					✓	5.31E-06	8	1.34	✓
3666	ARMA	Long	10	6						✓		✓	6.04E-06	9	2.62***	
2510	ARMA	Long	10	2					✓			✓	6.36E-06	10	2.30**	
3661	ARMA	Long	10	1						✓		✓	6.65E-06	11	2.21**	
14	ARMA	Long	2	0	✓								8.17E-06	12	4.02***	
206	ARMA	Long	10	2			✓					✓	8.63E-06	13	2.44**	✓
2	ARMA	Long	2	0									9.92E-06	14	5.22***	
2509	ARMA	Long	10	1					✓			✓	1.00E-05	15	2.72***	
24 Step - No GI																
8	ARMA	Long	8	0									3.02E-06	5	0.96	✓
14	ARMA	Long	2	0	✓								8.17E-06	12	4.02***	
2	ARMA	Long	2	0									9.92E-06	14	5.22***	
20	ARMA	Long	8	0	✓								1.33E-05	20	3.63***	
38	ARMA	Long	2	0	✓	✓							1.53E-05	23	6.46***	
10	ARMA	Long	10	0									1.61E-05	25	2.87***	
26	ARMA	Long	2	0		✓							1.64E-05	26	6.85***	
13	ARMA	Long	1	0	✓								2.85E-05	42	4.63***	
37	ARMA	Long	1	0	✓	✓							2.96E-05	44	4.75***	
61	AR	Long	1	0	✓								3.23E-05	50	4.49***	
62	AR	Long	2	0	✓								3.37E-05	56	4.56***	
22	ARMA	Long	10	0	✓								3.45E-05	57	5.06***	
49	AR	Long	1	0									3.52E-05	60	4.65***	
50	AR	Long	2	0									3.53E-05	61	4.53***	
25	ARMA	Long	1	0		✓							3.74E-05	70	5.79***	

Notes: Long sample: 1990:1-2009:12, Short sample: 2004:1-2009:12. In-sample ending with 2006:12; out of sample: 2007:1-2009:12. **Num** is model number, **Model** is either AR or ARMA, **Lag k** is the lag of the AR and MA part and of the exogenous leading indicators (GDP and UR) if present, **Lag m** is the lag for the Google index if present. **GDP**, **UR**, **GI1**, **GI2**, **GI3**, **GI4**, **GI5** are the leading indicators (GDP YoY change, unemployment rate YoY change, Google Index for keyword 'maternity', Google Index for keyword 'ovulation', Google Index for keyword 'pregnancy', the first principal component of the previous Google indexes, and Google Index for keyword 'KXMB', respectively.) A ✓ indicates that the row model adopts such leading indicator. **Google** indicates models using Google indeces. **MSE** is the mean squared prediction error of the row model, **rank** is the ranking with respect to the lowest MSE. **DM** is the Diebold-Mariano test for the null hypothesis of equal predictive accuracy (Diebold and Mariano, 1995) and **MCS** is the Model Confidence Set approach by Hansen, Lunde and Nason (2011). MCS has a ✓ when the row model is included in the final model confidence set at 5% confidence level. In all panels ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Table 5: Forecasting US birth rate in first differences ($d(br_t)$) - 2004-07 with AR(1) auxiliary model. Best 15 models at different forecast horizons (6, 12, 18 months ahead) with and without GI.

Num	Model	Sample	Lag k	Lag m	GDP	UR	GI1	GI2	GI3	GI4	Google	MSE	rank	DM	MCS
							6 Step - Overall								
4918	ARMA	Short	10	10			✓				✓	7.34E-07	1	0.00	✓
3874	ARMA	Long	3	10		✓				✓	✓	1.23E-06	2	1.40	✓
3865	ARMA	Long	3	1		✓				✓	✓	1.32E-06	3	1.27	✓
1715	ARMA	Long	3	11	✓	✓		✓			✓	1.42E-06	4	1.45	✓
4882	ARMA	Short	7	10			✓				✓	1.66E-06	5	2.04**	
3866	ARMA	Long	3	2		✓				✓	✓	1.75E-06	6	1.51	✓
121	ARMA	Long	3	1			✓				✓	1.75E-06	7	2.02**	
3721	ARMA	Long	3	1	✓					✓	✓	1.78E-06	8	1.80*	
4870	ARMA	Short	6	10			✓				✓	1.83E-06	9	2.28**	
1714	ARMA	Long	3	10	✓	✓		✓			✓	1.85E-06	10	1.60	
3577	ARMA	Long	3	1						✓	✓	1.88E-06	11	2.07**	
410	ARMA	Long	3	2		✓	✓				✓	1.90E-06	12	2.18**	
130	ARMA	Long	3	10			✓				✓	1.91E-06	13	3.16***	
4009	ARMA	Long	3	1	✓	✓				✓	✓	1.92E-06	14	2.08**	
3875	ARMA	Long	3	11		✓				✓	✓	2.02E-06	15	2.49**	
							6 Step - No Google								
14	ARMA	Long	2	0	✓							2.36E-06	28	2.54**	
37	ARMA	Long	1	0	✓	✓						2.61E-06	37	2.58***	
38	ARMA	Long	2	0	✓	✓						2.99E-06	56	3.09***	
20	ARMA	Long	8	0	✓							4.16E-06	138	3.35***	
2	ARMA	Long	2	0								4.17E-06	140	3.18***	
26	ARMA	Long	2	0		✓						4.20E-06	144	3.38***	
25	ARMA	Long	1	0		✓						4.31E-06	152	3.08***	
13	ARMA	Long	1	0	✓							4.50E-06	168	2.98***	
10	ARMA	Long	10	0								4.73E-06	182	4.12***	
61	AR	Long	1	0	✓							4.76E-06	188	2.81***	
62	AR	Long	2	0	✓							4.94E-06	203	2.83***	
4707	ARMA	Short	3	0								5.34E-06	247	3.57***	
4719	ARMA	Short	3	0	✓							5.86E-06	294	3.53***	
50	AR	Long	2	0								5.91E-06	302	2.83***	
49	AR	Long	1	0								6.00E-06	313	3.49***	
							12 Step - Overall								
3874	ARMA	Long	3	10		✓				✓	✓	1.21E-06	1	0.00	✓
4918	ARMA	Short	10	10			✓				✓	1.54E-06	2	0.49	✓
3851	ARMA	Long	1	11		✓				✓	✓	1.55E-06	3	0.52	✓
8579	ARMA	Short	3	11		✓				✓	✓	1.66E-06	4	0.88	✓
8567	ARMA	Short	2	11		✓				✓	✓	1.84E-06	5	0.82	✓
8566	ARMA	Short	2	10		✓				✓	✓	1.85E-06	6	0.87	✓
3875	ARMA	Long	3	11		✓				✓	✓	1.85E-06	7	1.90*	
8578	ARMA	Short	3	10		✓				✓	✓	1.90E-06	8	1.27	✓
121	ARMA	Long	3	1			✓				✓	2.18E-06	9	1.23	✓
8266	ARMA	Short	1	10						✓	✓	2.28E-06	10	2.09**	✓
8626	ARMA	Short	7	10		✓				✓	✓	2.41E-06	11	1.00	✓
1715	ARMA	Long	3	11	✓	✓		✓			✓	2.46E-06	12	1.12	✓
8555	ARMA	Short	1	11		✓				✓	✓	2.48E-06	13	0.92	✓
3721	ARMA	Long	3	1	✓					✓	✓	2.53E-06	14	1.58	
1691	ARMA	Long	1	11	✓	✓		✓			✓	2.55E-06	15	1.46	✓
							12 Step - No Google								
14	ARMA	Long	2	0	✓							3.36E-06	28	2.7978***	
37	ARMA	Long	1	0	✓	✓						4.50E-06	37	3.9663***	
38	ARMA	Long	2	0	✓	✓						4.96E-06	42	3.535***	
2	ARMA	Long	2	0								5.56E-06	56	3.4261***	
26	ARMA	Long	2	0		✓						5.79E-06	62	4.5083***	
10	ARMA	Long	10	0								7.24E-06	97	6.4331***	
25	ARMA	Long	1	0		✓						7.49E-06	104	5.7144***	
20	ARMA	Long	8	0	✓							8.01E-06	115	4.0027***	
13	ARMA	Long	1	0	✓							9.41E-06	147	2.8849***	
4719	ARMA	Short	3	0	✓							9.51E-06	149	2.8811***	
61	AR	Long	1	0	✓							1.10E-05	202	2.9079***	
62	AR	Long	2	0	✓							1.15E-05	224	3.9141***	
4707	ARMA	Short	3	0								1.30E-05	270	3.8219***	
50	AR	Long	2	0								1.32E-05	278	2.9597***	
49	AR	Long	1	0								1.36E-05	288	4.1893***	
							18 Step - Overall								
4918	ARMA	Short	10	10			✓				✓	1.29E-06	1	0.00	✓
8567	ARMA	Short	2	11		✓				✓	✓	2.11E-06	2	0.74	✓
8626	ARMA	Short	7	10		✓				✓	✓	2.13E-06	3	0.61	✓
3874	ARMA	Long	3	10		✓				✓	✓	2.14E-06	4	0.53	✓
3863	ARMA	Long	2	11		✓				✓	✓	2.43E-06	5	0.71	✓
8555	ARMA	Short	1	11		✓				✓	✓	2.92E-06	6	1.15	
3851	ARMA	Long	1	11		✓				✓	✓	3.44E-06	7	1.13	
130	ARMA	Long	3	10			✓				✓	3.91E-06	8	1.50	
553	ARMA	Long	3	1	✓	✓	✓				✓	4.03E-06	9	1.03	✓
396	ARMA	Long	1	12		✓	✓				✓	4.10E-06	10	1.60	
408	ARMA	Long	2	12		✓	✓				✓	4.11E-06	11	1.74*	
8566	ARMA	Short	2	10		✓				✓	✓	4.13E-06	12	1.65*	

(Continued on next page)

Table 5 – continued

Num	**Model**	**Sample**	**Lag k**	**Lag m**	**GDP**	**UR**	**GI1**	**GI2**	**GI3**	**GI4**	**Google**	**MSE**	**rank**	**DM**	**MCS**
3875	ARMA	Long	3	11		✓				✓	✓	4.41E-06	13	1.63	
409	ARMA	Long	3	1		✓	✓				✓	4.48E-06	14	1.14	
8579	ARMA	Short	3	11		✓				✓	✓	4.66E-06	15	1.27	
18 Step - No Google															
14	ARMA	Long	2	0	✓							9.61E-06	49	3.10***	
37	ARMA	Long	1	0	✓	✓						1.34E-05	96	3.42***	
38	ARMA	Long	2	0	✓	✓						1.40E-05	110	3.78***	
26	ARMA	Long	2	0		✓						1.52E-05	124	4.18***	
4719	ARMA	Short	3	0	✓							1.56E-05	139	4.92***	
2	ARMA	Long	2	0								1.65E-05	152	7.59***	
10	ARMA	Long	10	0								1.74E-05	171	3.46***	
25	ARMA	Long	1	0		✓						1.77E-05	179	8.11***	
20	ARMA	Long	8	0	✓							2.26E-05	257	6.12***	
4766	AR	Short	2	0	✓							2.48E-05	295	3.94***	
4743	ARMA	Short	3	0	✓	✓						2.79E-05	347	4.04***	
13	ARMA	Long	1	0	✓							2.81E-05	351	4.11***	
61	AR	Long	1	0	✓							3.26E-05	440	4.67***	
4765	AR	Short	1	0	✓							3.32E-05	455	2.82***	
62	AR	Long	2	0	✓							3.41E-05	470	4.58***	

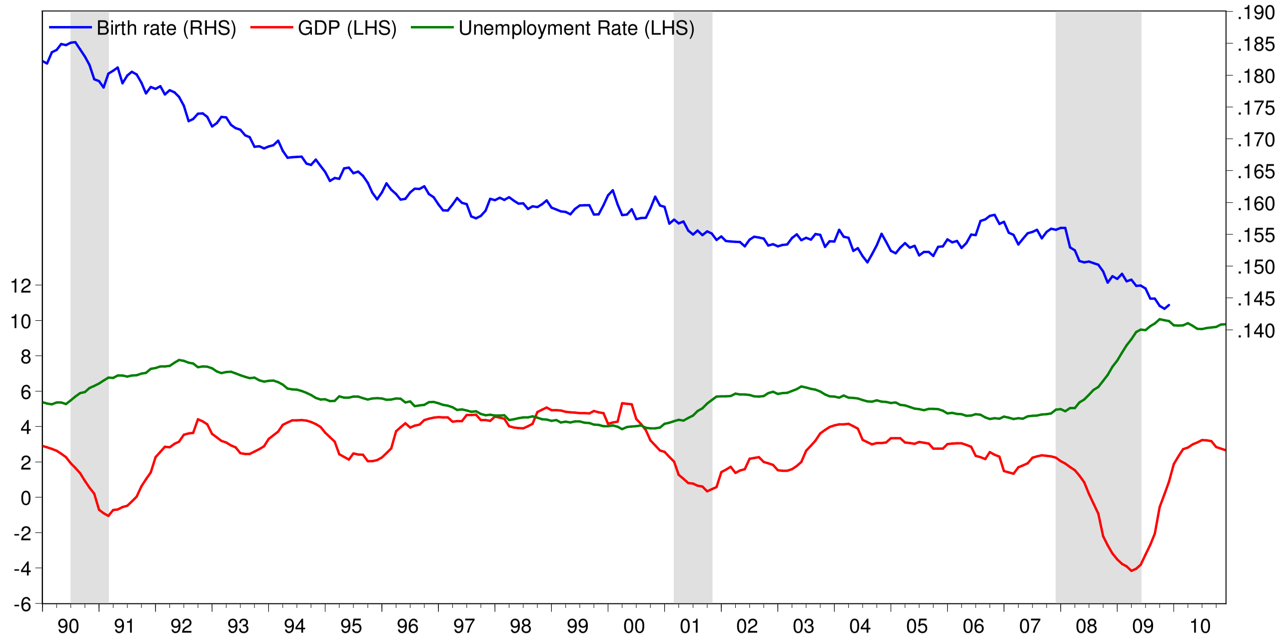
Notes: Long sample: 1990:1-2009:12, Short sample: 2004:1-2009:12. In-sample ending with 2006:12; out of sample: 2007:1-2009:12. **Num** is model number, **Model** is either AR or ARMA, **Lag k** is the lag of the AR and MA part and of the exogenous leading indicators (GDP and UR) if present, **Lag m** is the lag for the Google index if present. **GDP**, **UR**, **GI1**, **GI2**, **GI3**, **GI4** are the leading indicators (GDP YoY change, unemployment rate YoY change, Google Index for keyword 'maternity', Google Index for keyword 'ovulation', Google Index for keyword 'pregnancy', the first principal component of the previous Google indexes, respectively.) A ✓ indicates that the row model adopts such leading indicator. **Google** indicates models using Google indeces. **MSE** is the mean squared prediction error of the row model, **rank** is the ranking with respect to the lowest MSE. **DM** is the Diebold-Mariano test for the null hypothesis of equal predictive accuracy (Diebold and Mariano, 1995) and **MCS** is the Model Confidence Set approach by Hansen, Lunde and Nason (2011). MCS has a ✓ when the row model is included in the final model confidence set at 5% confidence level. In all panels ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Table 6: Forecasting birth rate in first differences ($d(br_t)$): one step ahead state level forecasts with AR(1) auxiliary model. Out of sample 2007:1-2009:12.

	Panel A				Panel B				Panel C			
State	**No Google**		**Google: GI1 & GI3**		**No Google**		**Google: GI1**		**No Google**		**Google: GI3**	
	mod #	**MSE**	**mod #**	**MSE**	**mod #**	**MSE**	**mod #**	**MSE**	**mod #**	**MSE**	**mod #**	**MSE**
0	56	6.41E-10	832	**9.36E-13**	56	6.41E-10	832	**9.36E-13**	56	6.41E-10	1983	**6.95E-11**
1	89	**1.91E-10**	2118	2.50E-09	89	**1.91E-10**	400	1.18E-07	89	**1.91E-10**	2118	2.50E-09
2	15	3.05E-09	1294	**4.14E-13**	15	3.05E-09	1223	**3.45E-10**	15	3.05E-09	1294	**4.14E-13**
3	7	1.93E-05	1251	**2.81E-11**	7	1.93E-05	2632	**6.61E-08**	7	1.93E-05	1251	**2.81E-11**
4	2412	1.56E-07	378	**3.92E-10**	2412	1.56E-07	378	**3.92E-10**	2412	1.56E-07	1848	**3.75E-09**
5	63	**1.48E-11**	711	3.77E-08	63	**1.48E-11**	711	3.77E-08	63	**1.48E-11**	1390	5.68E-08
6	2446	1.68E-07	1003	**5.27E-11**	2446	1.68E-07	1003	**5.27E-11**	2446	1.68E-07	3786	**1.28E-10**
7	37	2.66E-06	1629	**4.72E-10**	37	2.66E-06	2688	**7.39E-10**	37	2.66E-06	1629	**4.72E-10**
8	2453	2.10E-09	2731	**4.03E-11**	2453	2.10E-09	2731	**4.03E-11**	2453	2.10E-09	1900	**1.35E-10**
9	88	1.42E-08	3718	**2.81E-11**	88	1.42E-08	3081	**2.16E-09**	88	1.42E-08	3718	**2.81E-11**
10	37	2.83E-06	1986	**1.59E-12**	37	2.83E-06	2751	**2.28E-10**	37	2.83E-06	1986	**1.59E-12**
11	22	3.11E-06	4210	**1.29E-10**	22	3.11E-06	3069	**5.13E-10**	22	3.11E-06	4210	**1.29E-10**
12	24	3.75E-09	3099	**1.50E-13**	24	3.75E-09	3099	**1.50E-13**	24	3.75E-09	3868	**9.33E-12**
13	2419	7.54E-07	99	**7.35E-10**	2419	7.54E-07	99	**7.35E-10**	2419	7.54E-07	1371	**2.66E-09**
14	38	5.04E-11	4363	**7.51E-14**	38	**5.04E-11**	226	1.67E-10	38	5.04E-11	4363	**7.51E-14**
15	2418	3.80E-08	846	**1.77E-14**	2418	3.80E-08	846	**1.77E-14**	2418	3.80E-08	1415	**2.87E-11**
16	54	3.39E-10	4532	**4.52E-12**	54	3.39E-10	2965	**5.92E-11**	54	3.39E-10	4532	**4.52E-12**
17	2454	6.68E-08	720	**3.62E-14**	2454	6.68E-08	720	**3.62E-14**	2454	6.68E-08	1899	**4.99E-12**
18	82	5.46E-09	2917	**2.78E-11**	82	5.46E-09	2917	**2.78E-11**	82	5.46E-09	3927	**3.18E-10**
19	7	1.94E-09	4532	**1.73E-11**	7	1.94E-09	3210	**9.22E-10**	7	1.94E-09	4532	**1.73E-11**
20	16	8.11E-08	2629	**7.97E-10**	16	8.11E-08	2629	**7.97E-10**	16	8.11E-08	1692	**1.84E-08**
21	59	5.27E-09	1875	**7.62E-12**	59	5.27E-09	3325	**4.66E-10**	59	5.27E-09	1875	**7.62E-12**
22	2403	1.01E-09	3812	**6.18E-12**	2403	1.01E-09	144	**1.67E-11**	2403	1.01E-09	3812	**6.18E-12**
23	39	3.39E-10	2649	**1.40E-14**	39	3.39E-10	2649	**1.40E-14**	39	3.39E-10	1378	**7.78E-12**
24	6	2.05E-09	2559	**1.48E-12**	6	2.05E-09	2559	**1.48E-12**	6	2.05E-09	4255	**3.03E-12**
25	49	8.19E-06	3347	**3.19E-10**	49	8.19E-06	3347	**3.19E-10**	49	8.19E-06	1703	**7.18E-08**
26	8	1.01E-08	1312	**6.77E-14**	8	**1.01E-08**	263	2.09E-08	8	1.01E-08	1312	**6.77E-14**
27	6	2.30E-09	828	**1.53E-09**	6	2.30E-09	828	**1.53E-09**	6	**2.30E-09**	1547	1.02E-07
28	2401	1.71E-06	3817	**1.46E-08**	2401	1.71E-06	2663	**2.55E-08**	2401	1.71E-06	3817	**1.46E-08**
29	2419	1.44E-08	3854	**1.33E-09**	2419	1.44E-08	2774	**1.72E-08**	2419	1.44E-08	3854	**1.33E-09**
30	29	**2.65E-11**	3735	2.76E-11	29	**2.65E-11**	849	1.08E-10	29	**2.65E-11**	3735	2.76E-11
31	2480	2.08E-10	2792	**2.14E-12**	2480	2.08E-10	2792	**2.14E-12**	2480	2.08E-10	3740	**7.87E-11**
32	49	2.79E-09	1777	**1.01E-09**	49	2.79E-09	345	**1.15E-09**	49	2.79E-09	1777	**1.01E-09**
33	2423	5.96E-10	3025	**4.63E-11**	2423	5.96E-10	3025	**4.63E-11**	2423	5.96E-10	2114	**2.90E-10**
34	2445	2.22E-07	981	**4.94E-12**	2445	2.22E-07	981	**4.94E-12**	2445	2.22E-07	4049	**7.50E-10**
35	2436	1.58E-07	347	**2.34E-10**	2436	1.58E-07	347	**2.34E-10**	2436	1.58E-07	2148	**1.11E-09**
36	29	1.42E-10	4378	**7.29E-13**	29	1.42E-10	3392	**1.02E-10**	29	1.42E-10	4378	**7.29E-13**
37	2411	**5.47E-13**	1358	6.72E-13	2411	**5.47E-13**	649	9.08E-09	2411	**5.47E-13**	1358	6.72E-13
38	44	5.60E-09	1117	**2.25E-11**	44	5.60E-09	1117	**2.25E-11**	44	5.60E-09	4675	**3.48E-10**
39	96	2.14E-09	763	**4.31E-14**	96	2.14E-09	763	**4.31E-14**	96	2.14E-09	1297	**1.26E-09**
40	21	1.61E-10	1335	**6.07E-11**	21	**1.61E-10**	356	3.81E-10	21	1.61E-10	1335	**6.07E-11**
41	18	2.52E-06	110	**3.03E-09**	18	2.52E-06	110	**3.03E-09**	18	2.52E-06	4479	**8.38E-09**
42	59	2.02E-09	3977	**1.16E-09**	59	**2.02E-09**	2640	4.85E-09	59	2.02E-09	3977	**1.16E-09**
43	2458	8.42E-10	4276	**1.18E-12**	2458	8.42E-10	596	**1.88E-11**	2458	8.42E-10	4276	**1.18E-12**
44	2470	1.65E-09	3717	**3.02E-12**	2470	1.65E-09	1017	**1.96E-11**	2470	1.65E-09	3717	**3.02E-12**
45	2458	2.74E-08	2571	**6.66E-10**	2458	2.74E-08	2571	**6.66E-10**	2458	2.74E-08	3675	**2.10E-09**
46	2457	7.91E-09	813	**6.04E-12**	2457	7.91E-09	813	**6.04E-12**	2457	7.91E-09	1256	**7.22E-12**
47	68	2.64E-08	3230	**1.17E-11**	68	2.64E-08	3230	**1.17E-11**	68	2.64E-08	4273	**6.45E-11**
48	36	7.31E-09	4517	**1.57E-10**	36	7.31E-09	3512	**7.10E-09**	36	7.31E-09	4517	**1.57E-10**
49	2439	8.15E-09	4552	**4.20E-12**	2439	8.15E-09	3398	**6.79E-11**	2439	8.15E-09	4552	**4.20E-12**
50	2412	1.47E-08	4385	**1.47E-12**	2412	1.47E-08	2506	**1.09E-10**	2412	1.47E-08	4385	**1.47E-12**
51	2449	6.55E-10	249	**4.23E-13**	2449	6.55E-10	249	**4.23E-13**	2449	6.55E-10	1562	**4.45E-12**
Percentage of best models with GI												
among first 5				89.6%				89.4%				81.5%
among first 10				89.2%				89.4%				77.9%
among first 15				88.8%				88.9%				78.3%

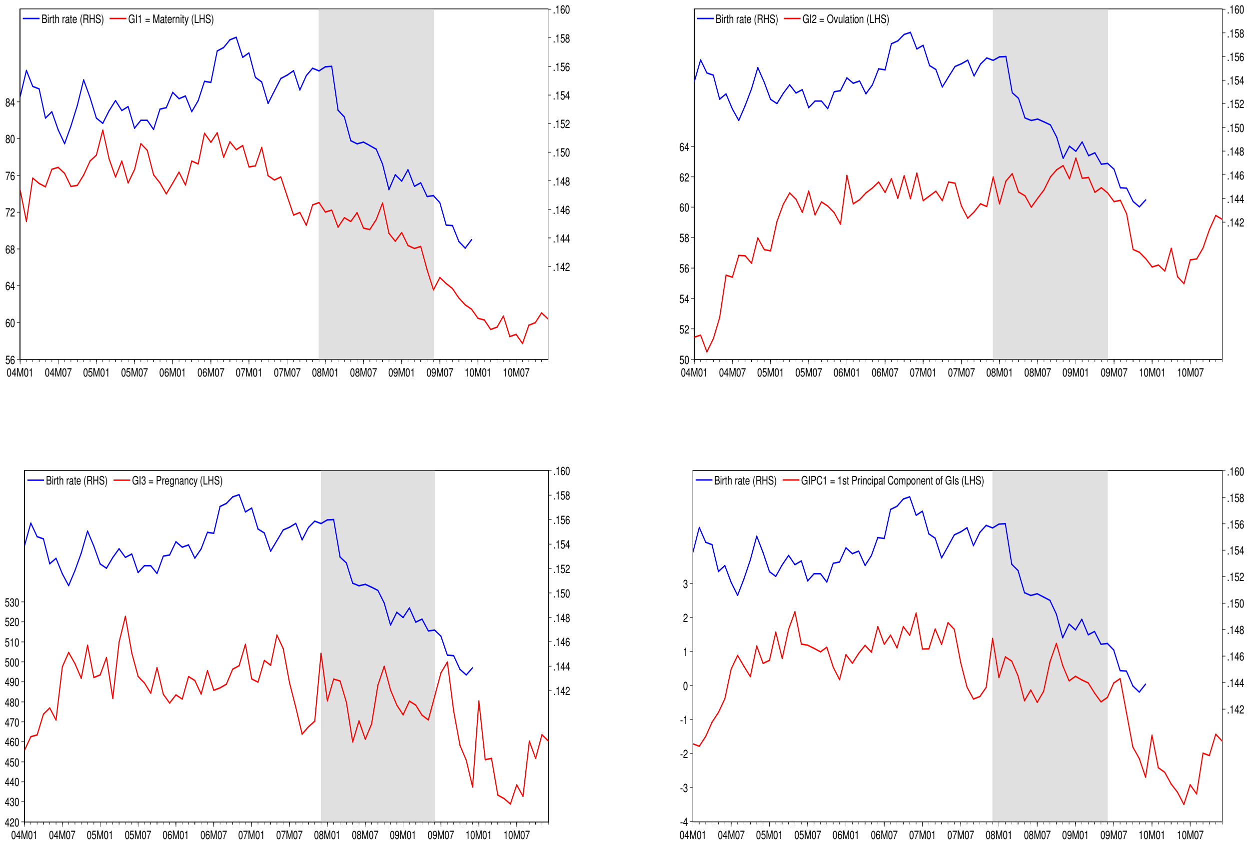
Notes: GI1 is the GI for 'maternity', GI3 is the GI for 'pregnancy', the only two GIs available at the state level. In-sample ending with 2006:12; out of sample: 2007:1-2009:12. **State** reports the State code (the code is set equal to zero for the federal level estimate and we consider also District Columbia) **mod #** is model number, **MSE** reports the lowest mean squared error. In each row, the MSE in bold indicates the best model. In panel A, the forecast comparison takes place between all non-google models and otherwise identical models augmented alternatively with GI1 or GI3; in panel B (C) the comparison *does not* feature GI3 (GI1) models.

Figure 1: Birth rate, GDP and Unemployment Rate - Long sample: 1990:1-2009:12



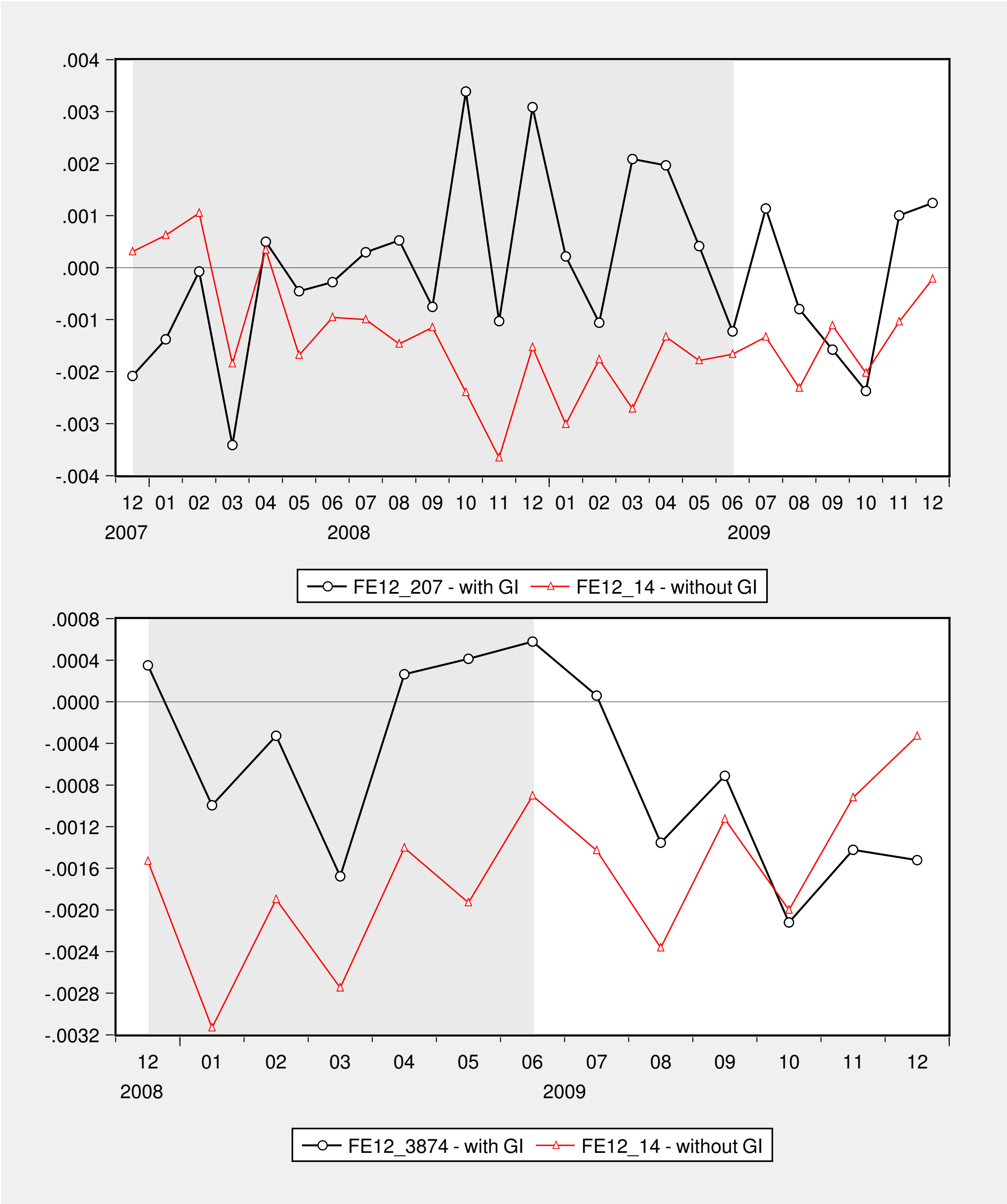
Notes: All vars are in percentage points. Shaded areas depict the NBER recessions.

Figure 2: Birth rate and Google indices - Short sample: 2004:1-2009:12



Notes: GI1 is the monthly average of the google index for 'maternity', GI2 is the monthly average of the GI for 'ovulation', GI3 is the monthly average of the GI for 'pregnancy', and GIPC1 is the first principal component of the previous three Google indices. Their sample is 2004:1-2009:12. The sample for the birth rate (br_t) is 1990:1-2009:12. The GI index takes a value of a 100 in the week in which the ration between the number of searches for the keyword 'maternity' was the highest. We normalize each index in order for it to be comparable to the one for *'maternity'*; in other words, a value of 120 for *'pregnancy'* in a given months means that the volume of searches for this keyword was 20% higher than the all-time peak reached by the volume of searches for *'maternity'*. Shaded areas identify NBER recessions.

Figure 3: Forecast errors of best models with and without GI - 12-month-ahead forecast errors. In-Sample 2004-06 (upper panel), 2004-07 (lower panel)



Notes: 12-month-ahead forecast errors from best model with GI and best model without GI. The upper graph shows the forecast errors for the in-sample 2004-06 of Table 4, while the lower graph depicts the forecast errors for the in-sample 2004-07 of Table 5. Shaded areas depict the NBER recessions.