Journal of Forecasting, *J. Forecast.* **32**, 289–298 (2013)
Published online 20 November 2011 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/for.1252

Nowcasting with Google Trends in an Emerging Market

YAN CARRIÈRE-SWALLOW* AND FELIPE LABBÉ
Central Bank of Chile, Santiago, Chile

ABSTRACT

Most economic variables are released with a lag, making it difficult for policy-makers to make an accurate assessment of current conditions. This paper explores whether observing Internet browsing habits can inform practitioners about aggregate consumer behavior in an emerging market. Using data on Google search queries, we introduce an index of online interest in automobile purchases in Chile and test whether it improves the fit and efficiency of nowcasting models for automobile sales. Despite relatively low rates of Internet usage among the population, we find that models incorporating our Google Trends Automotive Index outperform benchmark specifications in both in-sample and out-of-sample nowcasts, provide substantial gains in information delivery times, and are better at identifying turning points in the sales data. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS nowcasting; Google Trends; forecast accuracy; emerging markets

INTRODUCTION

The delay in data releases of key macroeconomic variables presents a limitation for decision-makers by restricting their ability to accurately assess current conditions. These lag times make nowcasting—or the prediction of the present—an important practice. The problem is of greater importance in emerging markets, where lags in the release of data are often longer than in developed countries. The availability of data that would allow decision-makers to observe trends as they unfold may improve the quality of economic assessments and, in turn, the decisions they inform.

In 2009, Google began the release of its users' search queries through a publicly accessible interface. The rapid expansion of the Internet into all aspects of modern life, together with Google's dominance in the search engine market, give the company a central role in the collection of market intelligence. The publication of user search queries offers researchers the tantalizing possibility to observe the interests of society in real time without carrying out costly surveys.

The company's Chief Economist, Hal Varian, published a research note signaling the data's potential uses shortly after the public release of the Google Trends portal (Choi and Varian (2009a)). The research question of immediate interest in the literature was whether the population's Internet tendencies contained a signal about their subsequent behavior. Choi and Varian (2009b) show that basic nowcasting models for unemployment claims in the USA can be substantially improved by incorporating search results for related keywords. Extending the scope of the concept, Della Penna and Huang (2009) use search results for retail goods to nowcast private consumption in the USA, and find that their index is a better predictor than the commonly used Michigan Consumer Sentiment Index and Conference Board Consumer Confidence Index. Vosen and Schmidt (2011) carry out a similar analysis using a principal component approach and also find that their index outperforms the survey-based indexes currently in use.

Suhoy (2009) conducts a growth cycle analysis for Israel, an emerging market country that has an unusually high level of Internet penetration (71% in 2006), and twice the income per capita of any South American country. She investigates whether Google query indexes could have predicted the 2008 downturn in real time, and finds that many web query categories do have predictive ability for real activity in their corresponding sector.

So far the literature establishing the link between Internet search patterns and observed consumer behavior has been limited to advanced economies, and we are unaware of attempts to test this link in an emerging market where penetration rates and wealth levels are much lower. It is not immediately clear that the Internet has become embedded into the consumer's purchasing decision in emerging economies. While Internet use is growing rapidly in many South American countries, penetration is only half of what is observed in the USA and Western Europe (World Bank, 2010). Furthermore, a larger share of private consumption is taken up by non-discretionary spending on basic items that do not require extensive research prior to purchasing. As such, there is reason to believe that the information contained in Internet search queries may represent too small a sample of the total population to provide a valuable signal about aggregate behavior.

* Correspondence to: Yan Carrière-Swallow, Central Bank of Chile, Agustinas 1180, Santiago, Chile. E-mail: ycarriere@bcentral.cl

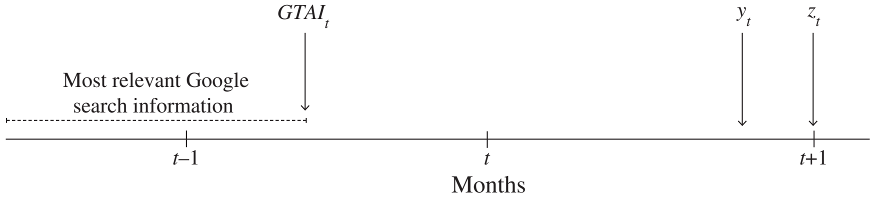

Figure 1. Timeline of data availability

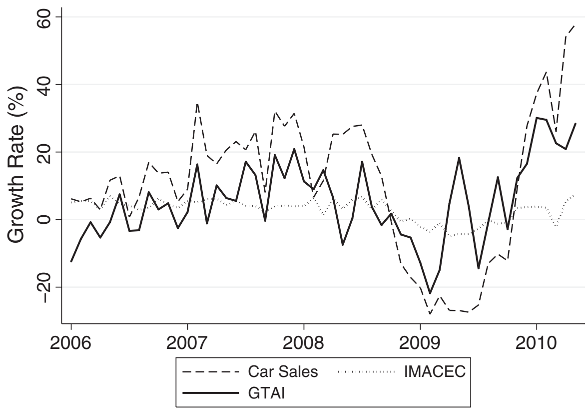

Figure 2. Data series employed

With these considerations in mind, we limit our analysis to the Chilean automotive sector for three reasons. First, the purchase of an automobile is a major event for most households, and one that requires a great deal of prior research into competing products. Second, the industry is dominated by a relatively small group of dedicated brands that can be adequately represented by a limited number of keywords. And finally, Chile has one of the highest rates of Internet usage in Latin America, at 35% of the population. The purpose of our empirical investigation is to evaluate the claim that Internet behavior correlates with consumer purchases in an emerging market. We do not aim to identify the most accurate model for nowcasting car sales, but rather to evaluate to what extent the popularity of Google search queries provides useful information about important macroeconomic variables in an emerging market.

DATA

Time series data on the volume of car sales are from the national statistics agency, the Instituto Nacional de Estadisticas de Chile, and include monthly sales of both new and used vehicles. The data are released during the last five days of the month following the close of the period in question, so sales are known with approximately a one-month lag. We use year-over-year changes in variables throughout the analysis to eliminate potential seasonal effects and non-stationarity issues. We denote the 12-month growth rate of the data y_t, which is available from January 2006 to May 2010.

As a measure of macroeconomic activity we employ the IMACEC series, released monthly with a five-week lag by the Department of National Accounts at the Central Bank of Chile.[1] We denote the year-over-year percentage change in this series as z_t. Figure 1 provides a timeline of data availability for the series employed in our exercise.

It should be noted that automobile sales are a volatile series. The standard deviation of the sales series y_t, 22.4%, is seven times greater than that of economic activity z_t. In Figure 2 we plot the data series employed in this paper against our Google Trends-based index (GTAI), which we present later.

The Internet search data is publicly available from Google through their Insights interface, and historical series beginning in January 2004 are available at a weekly frequency.[2] The interface returns one series per keyword for a given geographical area. For most countries, Google has also constructed categories that aggregate related keywords, but this feature has not yet been developed for Spanish-speaking countries.

The raw data undergo two transformations prior to public release. First, the data are normalized by the total number of search queries in the geographical region of interest. As such, any trends from growth in the total number of

[1] For simplicity, we will assume that these data are available with only a one-month lag. While this is not strictly the case, many of the components that make up the index are available one week prior to the aggregate.
[2] Available at http://www.google.com/insights/search/. Note that data are available at a daily frequency for short sample lengths, such that it should be possible to implement higher-frequency forecasts.

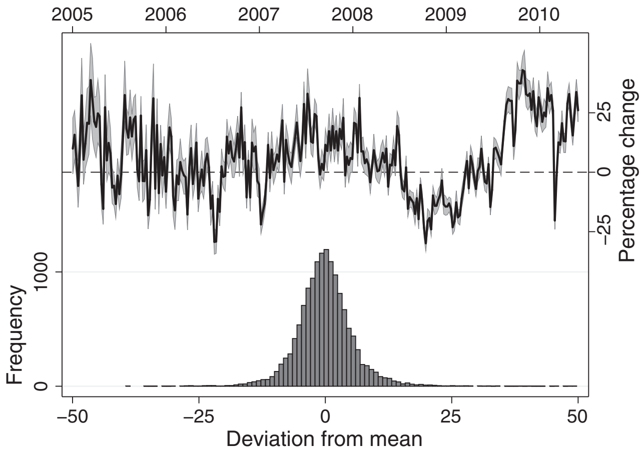

Figure 3. Sample noise: series returned for keyword 'Chevrolet'

Internet users or from a change in the relative popularity of Google as a search engine are removed from the data. Second, the normalized data are rescaled to an index with a maximum value of 100. This means that magnitudes are not directly comparable across series as a measure of relative popularity.

For our study, we use individual series on search queries for nine of the most popular automobile manufacturers in Chile by volume of sales, which together make up over 65% of sales by volume according to market share data from the Chilean National Automobile Association. The keywords used are: Chevrolet, Hyundai, Nissan, Kia, Toyota, Suzuki, Ford, Mitsubishi, and Mazda.

Suhoy (2009) suggests that Google popularity series for consumption-related keywords may exhibit a downward trend as the Internet becomes increasingly used for social-networking purposes, and uses first-differences to avoid stationarity problems. While we observe no such trend in the series used in our analysis, we acknowledge the possibility of such a problem in theory and encourage practitioners to check the data prior to modeling.

Addressing sampling noise

An important characteristic of the data is that Google employs a sampling procedure that introduces measurement error into the series. Requests for an identical query on different days return slightly different series, while queries sent on the same day produce identical series. This suggests that the sampling takes place once per 24-hour period. We download the series for each keyword on 50 occasions in order to characterize the measurement error and in the hope of identifying the underlying historical series.

The upper section of Figure 3 plots the cross-sectional mean across the 50 samples returned for the keyword 'Chevrolet', with the 10th and 90th percentiles shaded in gray.[3] In the lower part of the figure, we plot the histogram of the 14,140 mean-centered weekly observations. The distribution has a standard deviation of 5.8% and a kurtosis of over 10, while the historical series—plotted in the upper portion of the figure—has a standard deviation of 15.25%. By using the cross-sectional mean at each time t as the historical time series, we hope to identify the underlying signal in the data.

This measurement error is a source of concern because it weakens the information content of the Google data, and thus makes it more difficult to reject the null hypotheses we will test in the following sections. It is worth bearing in mind that the strength of our results could be improved if Google were to make cleaner data available in the future.

METHODOLOGY

Index construction

In the absence of search categories for Chile, we are left with the task of aggregating individual keyword series into an index of automobile-related queries. Rather than using a method based on another source of data, such as the market share of each brand, we fit the series to a linear model and allow the weights to take on any value on the real line.[4] Let $\boldsymbol{W}$ be a matrix of the year-over-year percentage change in the popularity of related keywords, where a monthly observation has been generated as the arithmetic mean of daily observations expanded from the weekly data

[3] For making this point we show results with the keyword 'Chevrolet', but the noise behaves in a similar manner for the rest of the keywords.

[4] We also constructed an index by aggregating individual series using principal components analysis, similar to the strategy employed by Vosen and Schmidt (2011). All factors with a corresponding eigenvalue greater than unity were aggregated into an index using their eigenvalues as weights. The results presented below hold when using this alternative methodology, but we prefer to employ the least-squares methodology because the resulting index is somewhat easier to interpret.

that Google make available. We re-estimate the weights $\hat{\boldsymbol{\beta}}$ in each period t by fitting the following model using the observations up to $t-1$:[5]

$$y_t = \alpha + \boldsymbol{\beta} \boldsymbol{W}_t + \epsilon_t$$

where y_t is the year-on-year percentage change in car sales and ϵ_t is a white noise error term. The Google Trends Automobile Index (GTAI) for period t is then computed as the fitted values from the regression, excluding the constant:

$$\text{GTAI}_t = \hat{E}_t[\boldsymbol{\beta}|\boldsymbol{\Omega}_{t-1}] \cdot \boldsymbol{W}_t$$

where $\boldsymbol{\Omega}_{t-1}$ is the information set available at time $t-1$. Due to the transformations Google applies to the data before public release, our index is not a direct measure of search volume. Rather, it is the linear combination of the columns of $\boldsymbol{W}$ that best explains actual sales, and is thus more appropriately interpreted as the year-over-year percentage change in interest to purchase an automobile among Chilean Internet users.

Benchmark framework

We begin our empirical exercise by fitting linear autoregressive specifications of order $p = 1, \ldots, 6$ for car sales to determine the data's time series structure. The parsimonious AR(1) specification provides a strong fit, which is reduced by adding higher-order autoregressive terms. An examination of model residuals reveals that no statistically significant autocorrelations remain, so we consider the AR(1) model an appropriate benchmark specification.

To ensure that the Google data are not simply providing information already available from other sources, in our second benchmark we augment the AR(1) specification by including the lagged values of the IMACEC index of industrial activity, which is commonly used in macroeconomic forecasting models due to its monthly frequency.

In the interest of encountering the strongest possible time series characterization, while acknowledging the risk of over-fitting the data, we estimate a family of ARMA(p,q) models of order up to $p = 3$ and $q = 6$. The ARMA(2,2) is the preferred model on the basis of both the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), and is used as our final benchmark specification.

The three benchmark models are

$$y_t = \alpha_{1a} + \rho_{1a} y_{t-1} + \upsilon_t \tag{1a}$$

$$y_t = \alpha_{2a} + \rho_{2a} y_{t-1} + \delta_{2a} z_{t-1} + \psi_t \tag{2a}$$

$$y_t = \alpha_{3a} + \sum_{p=1}^{2} \rho_{3a,p} y_{t-p} + \epsilon_t + \sum_{q=1}^{2} \theta_{3a,q} \epsilon_{t-q} \tag{3a}$$

where υ_t, ψ_t, and ϵ_t are white noise error terms. We then introduce a family of augmented models that incorporate GTAI_t:

$$y_t = \alpha_{1b} + \rho_{1b} y_{t-1} + \gamma_{1b}\text{GTAI}_t + \nu_t \tag{1b}$$

$$y_t = \alpha_{2b} + \rho_{2b} y_{t-1} + \delta_{2b} z_{t-1} + \gamma_{2b}\text{GTAI}_t + \varphi_t \tag{2b}$$

$$y_t = \alpha_{3b} + \sum_{p=1}^{2} \rho_{3b,p} y_{t-p} + \gamma_{3b}\text{GTAI}_t + \varepsilon_t + \sum_{q=1}^{2} \theta_{3b,q} \varepsilon_{t-q} \tag{3b}$$

where ν_t, φ_t, and ε_t are white noise error terms.

It is not necessarily the case that Internet users search for automobile information and then proceed with their purchase in the very same week. To investigate the possibility of a lag between searches and purchases, we test the in-sample predictive power of the Google index when sliding the observation window backwards. In Figure 4, the solid line shows the in-sample fit of model (1b) for lag lengths of up to 30 days, with the dashed line reporting the fit of benchmark model (1a). The smallest in-sample root mean square error (RMSE) is achieved by using a one-month window of observations that ends 18 days prior to the close of the sales period being nowcast, so we employ this information to construct the GTAI_t variable.

We are interested in testing whether GTAI_t contains relevant information beyond that which is contained in previous values of the dependent variable and other macroeconomic variables. This conceptual framework was proposed by Granger for determining the forecasting relation between two variables.

[5] Note that GTAI_t for January 2006 to December 2007 is computed using all the data over that period.

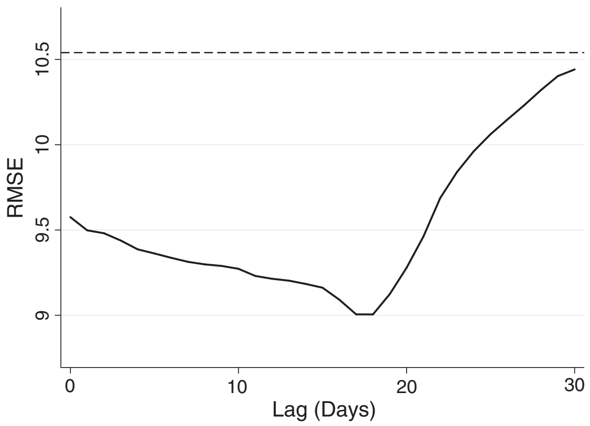

Figure 4. Identifying the optimal lag using an in-sample estimation of model (1b)

There is a lively ongoing discussion in the field of economic forecasting over how best to test such a hypothesis. Ashley *et al.* (1980) argue that an out-of-sample prediction approach is more in the spirit of Granger's conceptual framework, since it only uses information available at the time the forecast is generated. Chen (2005) uses a Monte Carlo experiment to test the relative power of the two methods, and finds that out-of-sample tests do indeed offer higher power when the dependent variable exhibits a structural break. When no structural break is present, however, the in-sample test provides a higher power for testing the null hypothesis.

The short length of our sample (53 observations) makes it difficult to perform rigorous tests of parameter stability. Since our aim is limited to establishing the relevance of the GTAI and not to identifying the best estimation procedure, we test the null hypothesis of no Granger causality using nowcasting results from an in-sample estimation and two out-of-sample estimation schemes to ensure that our findings are robust.

RESULTS

In-sample estimation

We begin our analysis by comparing the augmented models to their relevant benchmarks on the basis of in-sample predictive accuracy. To do so, we estimate model parameters using the complete sample and use the results to compute nowcast residuals. Estimation results are displayed in Table I. Model point accuracy is compared on the basis of the adjusted-R^2 and RMSE, and skill at identifying the direction-of-change is compared using a sign test.

Table I. In-sample estimation results. Dependent variable: car sales (y_t)

Ind. variable	Benchmarks			Including GTAI		
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
Car sales (y_t)						
$t-1$	*0.91	*0.99		*0.74	*0.72	
	(0.06)	(0.15)		(0.09)	(0.16)	
ARMA(2,2)			✓			✓
IMACEC (z_t)						
$t-1$		−0.60			0.12	
		(0.92)			(0.89)	
GTAI				*0.57	*0.58	*0.47
				(0.15)	(0.15)	(0.16)
N	52	52	51	52	52	51
R^2	0.75	0.76	0.81	0.82	0.82	0.84
Adj. R^2	0.75	0.75	0.79	0.82	0.71	0.83
RMSE	10.5	10.6	9.14	9.01	9.10	8.14

Notes:
(i) All variables are 12-month percentage changes.
(ii) White heteroskedasticity-robust standard errors are reported below coefficient values.
(iii) All regressions include a constant term.
(iv) *Significant at the 0.01 level.

Since the benchmark models are nested within the augmented models, testing whether GTAI_t Granger-causes y_t corresponds to a simple t-test of the null hypothesis that $\gamma = 0$ or, equivalently in bivariate models, whether $\text{MSE}_a \leqslant \text{MSE}_b$. As is reported in Table I, the null is rejected for each of the augmented models at the 99% confidence level. The introduction of Google information improves the fit over each benchmark specification, reducing the RMSE between 9% and 14%. Since it generates the best fit and its residuals appear homoskedastic, we focus the remaining discussion on models (3a) and (3b), while pointing out that the results are broadly consistent across the three benchmarks.

Besides testing Granger causality on a point-accuracy basis, as we have done so far, it is of interest to test the related but distinct hypothesis that the GTAI can identify turning points in the data. To test this claim, we conduct an event-forecast evaluation using the sign test methodology employed by Trefler (1995). We begin by computing the share of periods in which each model accurately predicts the sign of the first and second differences in sales. We find that both models (3a) and (3b) correctly predict a rise or fall in car sales in over 95% of periods, but that model (3a) correctly identifies the direction of the change in the growth rate in only 50% of months. The introduction of Google Trends data in model (3b) increases the success rate to 65%. We verify whether the model's success rate is statistically significant using the market timing test proposed by Henriksson and Merton (1981). While model (3a) generates a test statistic below 1, indicating poor timing skill, we are able to reject the null hypothesis of no timing skills for model (3b) at the 95% confidence level.

These results are consistent across each of the benchmark models we have mentioned above. The substantial improvements in in-sample nowcast accuracy over the benchmark models suggest that the GTAI indeed contains information beyond that which is contained in previous values of sales and macroeconomic variables.

Out-of-sample nowcast evaluation

The analysis presented thus far has demonstrated that the introduction of Google information outperforms univariate specifications in *ex post* in-sample estimations. We now investigate the claim that Google data can also improve nowcast accuracy in out-of-sample estimations. We undertake the exercise using two estimation schemes common in the literature, each of which has advantages and shortcomings. Under both schemes, each of the benchmark models exhibits lower mean squared prediction errors (MSPE) when the Google data are included. In the spirit of establishing Granger causality, we would like to test whether this improvement in forecast error is due to chance. Testing the relative performance of out-of-sample prediction models is a lively topic in the econometrics literature, leading us to use a number of tests to ensure the robustness of our findings. We will evaluate out-of-sample forecasting accuracy using conditional and unconditional tests of MSPE and forecast encompassing. We summarize the results in Table II, and plot the nowcast predictions from model (1b) against the actual sales data in Figure 5.

Estimation schemes

Under the recursive scheme, we begin by estimating the GTAI and model parameters over the first R periods corresponding to the interval of January 2006 to December 2007. These estimates are then used to formulate the first nowcast for period $R + 1$. We then re-estimate the models by extending the sample forward period by period for end dates $t \in \{R + 1, \ldots, T + 1\}$, where $T + 1$ is the number of periods in the full sample, and compute the nowcast error for contemporaneous car sales.[6] This method has the virtue of using all the information available at time t, such that parameter estimates are expected to converge to the in-sample estimates as $P \to T + 1$, where P is the number of periods used to generate the forecast.

While a visual inspection of estimated parameters from the recursive nowcast of model (1b) indicates that estimates may be unstable over time, the short length of our sample prevents us from making any defendable claims about parameter stability. Chen (2005) and Giacomini and White (2006) point out that forecast accuracy and test power can be improved by using a rolling window estimation scheme when parameters are unstable. In such a scheme, the model is re-estimated in each period $t \in (R + 1, \ldots, T + 1)$ using a sample of fixed size P. We estimate the models using a 24-month rolling window estimation to test the robustness of our findings.

We find that the AR(1) specifications produce more accurate out-of-sample nowcasts than the ARMA(2,2) specifications under both schemes, suggesting that the latter may be over-fitting the data. If the researcher were interested in identifying the best model to use, this inconsistency may be of concern. For our purposes, however, it suffices that the Google Trends-augmented models continue to outperform their respective reduced-form benchmarks.

Unconditional tests

We denote the one-step-ahead forecast error of model i by $\hat{e}_{i,t+1} \equiv y_{i,t+1} - E_t[\hat{y}_{i,t+1}]$. Since the benchmark models are nested within the augmented models, model residuals are identical (and thus are not independent) under the null

[6] We use only 18 samples to construct the final data point, in order to simulate the availability of data the nowcaster would have at their disposal. The historical Google search series are constructed using the full 50 samples.

Table II. Out-of-sample test statistics

Source	Test	Models	Null hypothesis	Test stat.	p-value
Recursive scheme					
Clark and McCracken (2001)	ENC-T	1a vs. 1b	$E[e_a(e_a - e_b)] = 0$	3.36	0.01
Clark and McCracken (2001)	ENC-T	2a vs. 2b	$E[e_a(e_a - e_b)] = 0$	3.76	0.01
Clark and McCracken (2001)	ENC-T	3a vs. 3b	$E[e_a(e_a - e_b)] = 0$	n.a.	n.a.
Clark and McCracken (2001)	ENC-NEW	1a vs. 1b	$\text{MSPE}_a = \text{MSPE}_b$	14.1	0.01
Clark and McCracken (2001)	ENC-NEW	2a vs. 2b	$\text{MSPE}_a = \text{MSPE}_b$	12.7	0.01
Clark and McCracken (2001)	ENC-NEW	3a vs. 3b	$\text{MSPE}_a = \text{MSPE}_b$	n.a.	n.a.
Clark and West (2007)	MSPE-adjusted	1a vs. 1b	$\text{MSPE}_a = \text{MSPE}_b$	1.74	(0.01, 0.05]
Clark and West (2007)	MSPE-adjusted	2a vs. 2b	$\text{MSPE}_a = \text{MSPE}_b$	1.86	(0.01, 0.05]
Clark and West (2007)	MSPE-adjusted	3a vs. 3b	$\text{MSPE}_a = \text{MSPE}_b$	1.15	>0.10
Rolling window scheme					
Clark and McCracken (2001)	ENC-T	1a vs. 1b	$E[e_a(e_a - e_b)] = 0$	4.07	0.01
Clark and McCracken (2001)	ENC-T	2a vs. 2b	$E[e_a(e_a - e_b)] = 0$	3.38	0.01
Clark and McCracken (2001)	ENC-T	3a vs. 3b	$E[e_a(e_a - e_b)] = 0$	n.a.	n.a.
Clark and McCracken (2001)	ENC-NEW	1a vs. 1b	$\text{MSPE}_a = \text{MSPE}_b$	17.4	0.01
Clark and McCracken (2001)	ENC-NEW	2a vs. 2b	$\text{MSPE}_a = \text{MSPE}_b$	12.4	0.01
Clark and McCracken (2001)	ENC-NEW	3a vs. 3b	$\text{MSPE}_a = \text{MSPE}_b$	n.a.	n.a.
Clark and West (2007)	MSPE-adjusted	1a vs. 1b	$\text{MSPE}_a = \text{MSPE}_b$	1.77	(0.01, 0.05]
Clark and West (2007)	MSPE-adjusted	2a vs. 2b	$\text{MSPE}_a = \text{MSPE}_b$	1.45	(0.05, 0.10]
Clark and West (2007)	MSPE-adjusted	3a vs. 3b	$\text{MSPE}_a = \text{MSPE}_b$	1.09	>0.10
Giacomini and White (2006)	CPA	1a vs. 1b	$E[L_{a,t+1} - L_{b,t+1} \mid \mathbb{F}_t] = 0$	5.21	0.07
Giacomini and White (2006)	CPA	2a vs. 2b	$E[L_{a,t+1} - L_{b,t+1} \mid \mathbb{F}_t] = 0$	5.53	0.06
Giacomini and White (2006)	CPA	3a vs. 3b	$E[L_{a,t+1} - L_{b,t+1} \mid \mathbb{F}_t] = 0$	2.00	0.36

Notes:
(i) ENC-T and ENC-NEW tests were performed with $\pi = 1.2$ and critical values were obtained from Clark and McCracken (2001).
(ii) Under the rolling scheme we estimate the models using a 24-month rolling window.
(iii) Matlab codes for performing Giacomini and White (2006) were obtained from http://www.homepages.ucl.ac.uk/~uctprgi/http://www.homepages.ucl.ac.uk/ uctprgi/.

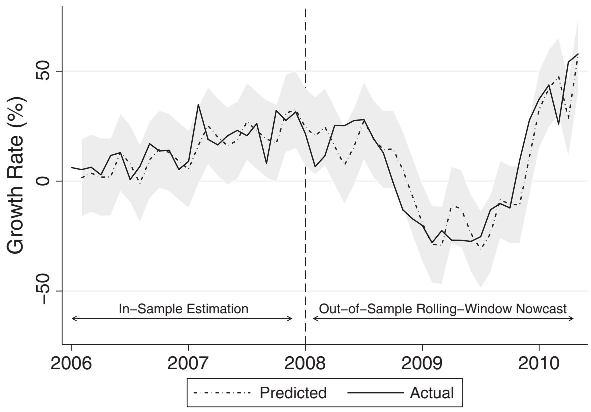

Figure 5. Nowcast for automobile sales using model (1b)

hypothesis of equal forecast performance. As a result, the test statistics proposed by Diebold and Mariano (1995) and West (1996) will not follow standard distributions.

The unconditional tests of predictive ability and the distributions described by Clark and McCracken (2001) allow evaluating the relative average accuracy of competing nested models under certain conditions. These requirements, summarized in West (2006), are that: (i) the estimator must be nonlinear least squares (of which ordinary least squares is a special case); (ii) linear nested models are being compared; and (iii) the forecast horizon is one-step-ahead and forecast errors are homoskedastic. Our nowcasting framework meets the first two requirements, but none of the models considered in this paper are able to adequately capture the serial correlation in the error terms following the start of the 2008 financial crisis. However, Clark and McCracken (2005) demonstrate that requirement (iii) need not be met when the number of additional regressors in the augmented model is exactly one, which is the case in our tests.

We first carry out unconditional encompassing tests of predictability in which the null hypothesis is that $E[e_{a,t}(e_{a,t} - e_{b,t})] = 0$. The test statistics we employ are

$$\text{ENC-T} = P^{\frac{1}{2}} \frac{\bar{c}}{\sqrt{P^{-1} \sum_{t=R}^{T} (c_{t+1} - \bar{c})^2}}$$

and

$$\text{ENC-NEW} = P \cdot \frac{P^{-1} \sum_{t=R}^{T} \hat{e}_{a,t+1}(\hat{e}_{a,t+1} - \hat{e}_{b,t+1})}{P^{-1} \sum_{t=R}^{T} \hat{e}^2_{b,t+1}}$$

where $c_{t+1} = \hat{e}_{a,t+1}(\hat{e}_{a,t+1} - \hat{e}_{b,t+1})$ and $\bar{c} = P^{-1} \sum_{t=R}^{T} c_{t+1}$. We compare these test statistics to the critical values reported in Clark and McCracken (2001), and are able to reject the null hypothesis for models 1 and 2 at the 0.99 confidence level. The interpretation of this result is that the benchmark forecasts do not encompass the Google-augmented forecasts, such that incorporating the GTAI_t variable adds relevant information to the models.

Next, we compute Clark and West's (2007) unconditional test of mean squared prediction error using the MSPE-adjusted test statistic. This test statistic adjusts the upward bias in the MSPE of the larger model. The intuition is that under the null hypothesis that the parsimonious model is correct, the alternative model introduces noise into the forecast by estimating parameters with population values equal to zero. This procedure yields a test statistic better centered around zero. The requirements for the validity of the test are less restrictive than those for the tests mentioned above, and allow for multi-step forecasts with autocorrelated forecast errors.

The null hypothesis for the test is that $Ee^2_{a,t} - Ee^2_{b,t} = 0$. We regress $\hat{f}_{t+1} = (y_{t+1} - \hat{y}_{a,t+1})^2 - [(y_{t+1} - \hat{y}_{b,t+1})^2 - (\hat{y}_{a,t+1} - \hat{y}_{b,t+1})^2]$ on a constant, where the resulting t-statistic for a zero coefficient corresponds to the MSPE-adjusted statistic. By estimating the standard error of the test statistic using the Newey–West heteroskedasticity and autocorrelation consistent covariance matrix, the test is robust to serial correlation in the error terms. We are able to reject the null hypothesis at the 90% confidence level under both recursive and rolling estimation schemes, implying that, on average, the Google-augmented models produce lower out-of-sample forecast errors and are thus more efficient than the benchmark specifications.

Conditional test of predictive ability

The conditional framework allows one to test which model will perform better in the future rather than simply evaluating which model has performed better on average in the past, which we have established in the previous section. The test is conditional on the information available to the forecaster, and its implications deal with sample parameter estimates rather than population values. As this approach relies on a bounded estimation window, its results are more suitable for practitioners who dispose of a small number of observations, such that the estimated parameters have not yet converged to their population value.

Using Giacomini and White's (2006) conditional predictive ability (CPA) procedure, we test the null hypothesis that the estimated MSPE from models a and b are equal, or that

$$E\left[L\left(y_{t+1}, y_{a,t+1}\left(\hat{\beta}_{a,t}\right)\right) - L\left(y_{t+1}, y_{b,t+1}\left(\hat{\beta}_{b,t}\right)\right)|\mathbb{F}_t\right] = 0$$

where L is a user-defined loss function. The conditions for the validity of this test differ from those presented in the previous section. The test can be carried out on both nested and non-nested models, and even on misspecified models. While heteroskedastic errors are admissible, the test requires an estimation window of a fixed size, precluding the use of the recursive estimation scheme. We employ the usual quadratic loss function for L, and select an $\mathbb{F}_t$-measurable vector h_t that contains lagged differences of the loss function, such that $h_t = \text{vec}(1, \Delta L_t)$. Our choice of h_t is an assumption that past forecast performance contains information about subsequent forecast performance. As such, the interpretation of our hypothesis test is that, given that the Google-augmented models have outperformed the baselines so far, they will continue to do so in the following period. We are able to reject the null hypothesis at the 90% confidence level for models 1 and 2.

The evidence presented confirms that the inclusion of information on Google search queries improves both the in- and out-of-sample performance of models for automobile sales in Chile.

CONCLUSION

We have examined whether Google search results provide relevant information about sales of automobiles in an emerging market. We constructed a Google Trends Automotive Index (GTAI) using search queries for Chile, and included this index as a regressor in a family of simple nowcasting models. Our results show that models incorporating Google search results outperform competing benchmark specifications in both in- and out-of-sample nowcasting exercises, improving in-sample efficiency by up to 14%. The Google data have a number of characteristics that should make them particularly attractive to decision-makers in emerging markets: (i) they are derived directly from micro user data; (ii) they contain information on a large proportion of Internet users, which is a far more extensive sample than is commonly employed by surveying agencies; and (iii) they are released at high frequency and at regular intervals. However, the data as currently available contain a substantial measurement error, which we have attempted to characterize. Our finding that the accuracy of nowcasting models for automobile sales in an emerging market can be improved using contemporaneous search patterns suggests that Google data are a promising source of information for nowcasting components of aggregate demand in short-run models in these economies—an exercise that we leave to future research.

ACKNOWLEDGEMENTS

We thank Luis Felipe Céspedes, Charles Engel, Òscar Jordà, Pablo Pincheira, Miguel Ricaurte, an anonymous referee, and seminar participants at the Central Bank of Chile for their valuable comments. Remaining errors are entirely our own.

REFERENCES

Ashley R, Granger C, Schmalensee R. 1980. Advertising and aggregate consumption: an analysis of causality. *Econometrica* **48**: 1149–1167.

Chen S. 2005. A note on in-sample and out-of-sample tests for Granger causality. *Journal of Forecasting* **24**: 453–464.

Choi H, Varian H. 2009a. *Predicting the present with Google Trends*. Economics Research Group, Google.

Choi H, Varian H. 2009b. *Predicting initial claims for unemployment benefits*. Economics Research Group, Google.

Clark TE, McCracken MW. 2001. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* **105**: 85–110.

Clark TE, McCracken MW. 2005. Evaluating direct multistep forecasts. *Econometric Reviews* **24**(4): 369–404.

Clark TE, West KD. 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* **138**: 291–311.

Della Penna N, Huang H. 2009. *Constructing consumer sentiment index for U.S. using Google searches*. Working Papers 2009-26, Department of Economics, University of Alberta.

Diebold FX, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**(3): 253–263.
Giacomini R, White H. 2006. Tests of conditional predictive ability. *Econometrica* **74**(6): 1545–1578.
Henriksson R, Merton R. 1981. On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business* **54**(4): 513–533.
Suhoy T. 2009. *Query indices and a 2008 downturn: Israeli data.* Discussion Paper No. 2009-06, Bank of Israel.
Trefler D. 1995. The case of the missing trade and other mysteries. *American Economic Review* **85**(5): 1029–1046.
Vosen S, Schmidt T. 2011. Forecasting private consumption: survey-based indicators vs. Google Trends. *Journal of Forecasting* **30**(6): 565–578.
West KD. 1996. Asymptotic inference about predictive ability. *Econometrica* **64**: 1067–1084.
West KD. 2006. Forecast evaluation. In *Handbook of Economic Forecasting*, Vol. 1, Elliott G, Granger C, Timmermann A (eds). Elsevier: Amsterdam: 94–134.
World Bank. 2010. World Development Indicators 2010. The World Bank Group (CD-ROM).

Authors' biographies:
Yan Carrière-Swallow is an economist at the Central Bank of Chile. He studied as an undergraduate at the University of British Columbia, and holds an M.A. from Queen's University.

Felipe Labbé is an economist at the Central Bank of Chile. He received his B.A. and M.A. degrees from the Pontificia Universidad Católica de Chile.

Authors' addresses:
Yan Carrière-Swallow and **Felipe Labbé,** Central Bank of Chile, Agustinas 1180, Santiago, Chile.