Forecasting Building Permits with Google Trends

David Coble & Pablo Pincheira Brown

School of Business, Adolfo Ibáñez University, Santiago, Chile

BCRA, November 2017

This presentation is based on two papers

1. Nowcasting building permits with google trends (joint with David Coble).
2. The Concatenation of Forecasts (joint with Ignacio Finot). Work in progress.

What do we do in the first paper?

1. We provide strong evidence of the ability that some internet search queries have to generate accurate backcasts, nowcasts and forecasts of Building Permits (BP) in the US.
2. This is important because BP are useful to:
    1. Predict employment and construction activity.
    2. Perform assessments on housing programs.
    3. Forecast the demand of mortgage-related products.
    4. Investment analysis, risk evaluations, etc.

What do we do in the second paper?

1. We study the propagation of one-step-ahead forecast errors across multiple horizons.
2. This allows us to measure the impact of short run forecast accuracy in long run accuracy.
3. Put it differently, we analyze whether accurate nowcasts help to reduce forecast errors at longer horizons.
4. We also evaluate whether the concatenation of forecasts may increase long run forecast accuracy.

Outline

1. The propagation of forecast errors
2. Forecasting and nowcasting building permits.
3. Out-of-sample evaluation strategy.
4. Empirical results.
5. Conclusions.

The propagation of forecast errors

1. Sometimes it is believed that an accurate nowcast is relevant for the generation of accurate multistep ahead forecasts.
2. In a joint work with Ignacio Finot we show that sometimes this is the case, but sometimes it is not.
3. Let us suppose that we use a simple linear AR(p) model

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$

to generate, iteratively, multistep ahead forecasts according to the following rule:

$$\begin{aligned} Y_t^f(h) &= c + \phi_1 Y_t^f(h-1) + \phi_2 Y_t^f(h-2) + ... + \phi_p Y_t^f(h-p) \\ Y_t^f(h) &= y_{t+h} \; if \; h \leq 0 \end{aligned}$$

The propagation of forecast errors

1. Sometimes it is believed that an accurate nowcast is relevant for the generation of accurate multistep ahead forecasts.
2. In a joint work with Ignacio Finot we show that sometimes this is the case, but sometimes it is not.
3. Let us suppose that we use a simple linear AR(p) model

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$

to generate, iteratively, multistep ahead forecasts according to the following rule:

$$\begin{aligned} Y_t^f(h) &= c + \phi_1 Y_t^f(h-1) + \phi_2 Y_t^f(h-2) + ... + \phi_p Y_t^f(h-p) \\ Y_t^f(h) &= y_{t+h} \ if \ h \leq 0 \end{aligned}$$

The concatenation of forecasts

1. Let us suppose that at time t_0 we have available a very accurate exogenous one-step-ahead forecast for y_{t_0+1} that we call $Y_{t_0}^X(1)$. We also denote its forecast error as $e_{t_0}^X(1)$.
2. We can build a hybrid multi step ahead forecast as follows:

$$
\begin{aligned}
Y_{t_0}^C(1) &= Y_{t_0}^X(1) \\
Y_{t_0}^C(2) &= c + \phi_1 Y_{t_0}^C(1) + \phi_2 y_{t_0} + \ldots + \phi_p y_{t_0-p+2}
\end{aligned}
$$

and more generally

$$
\begin{aligned}
Y_{t_0}^C(h) &= c + \phi_1 Y_{t_0}^C(h-1) + \phi_2 Y_{t_0}^C(h-2) + \ldots + \phi_p Y_{t_0}^C(h-p) \\
Y_{t_0}^C(h) &= y_{t_0+h} \ \ if \ \ h \leq 0 \\
Y_{t_0}^C(1) &= Y_{t_0}^X(1)
\end{aligned}
$$

Multistep ahead forecast accuracy

1. The relative accuracy of the hybrid forecast at multiple horizons is defined as
$$MSPE(h) = \mathbb{E}\left[e_{t_0}(h)\right]^2 - \mathbb{E}\left[e_{t_0}^{C}(h)\right]^2$$
2. It satisfies
$$MSPE(h) = A^2(h) + C(h)$$
where
$$\begin{aligned} A^2(h) &= \psi_{h-1}^2\left[\mathbb{E}\left[e_{t_0}(1)\right]^2 - \mathbb{E}\left[e_{t_0}^{X}(1)\right]^2\right] > 0 \\ C(h) &= 2\psi_{h-1}\mathbb{E}\left\{\left[e_{t_0+1}(h-1)\right]\left[e_{t_0}(1) - e_{t_0}^{X}(1)\right]\right\} \gtrless 0 \end{aligned}$$
3. The "accuracy" term $A^2(h)$ shows how the accuracy in the short run propagates (or vanishes) with the forecasting horizon.
4. The "correlation" term shows interesting interactions between one step forecast errors at time t_0 and multiple step ahead forecasts at time $t_0 + 1$.

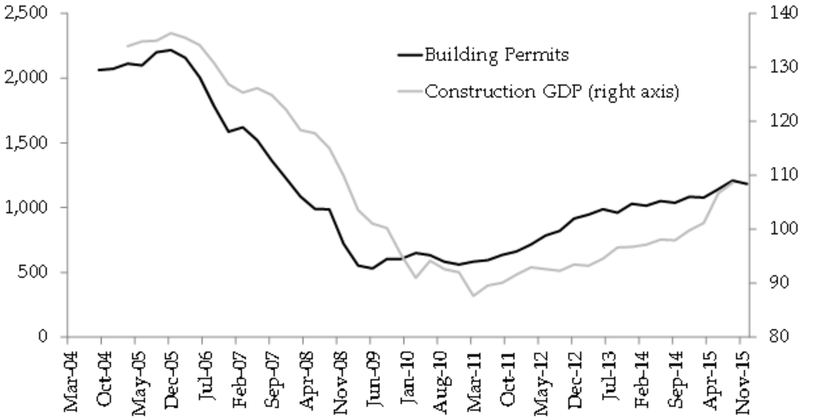
Implications for long run forecast accuracy

1. An accurate short run forecast implies accurate longer run forecasts depending on the time-series structure of the Forecast Generating Process (FGP), short run accuracy and intertemporal efficiency.
2. It might be extremely relevant for persistent processes, but not very relevant for short-lived processes.
3. We will see an example shortly.

Forecasting Building Permits

1. Building permits is the primary leading indicator of economic activity in the construction sector.
2. But it is released with a lag of almost two months.
3. Consequently, the current state of the business cycle in that sector cannot be known in real time.
4. Strategies to build reliable backcasts, nowcasts and forecasts of building permits are desirable.

Figure 1: Building Permits and Construction GDP



Source:US.Bureau of Economic Analysis and Census Bureau. BP are expressed in thousands of units. Construction GDP is expressed as a chain quantity index (2009=100). Both series are seasonally adjusted.

We propose to use Google Trends

1. We propose methods to predict building permits in the US, exploiting rich real-time data from web search queries.
2. Using Google Trends, we find some keywords with strong predictive information.
3. We show that they predict more accurately than our competing benchmarks both in-sample and out-of-sample.

Favorite keywords

1. We use the term "real state exam" which is connected with the supply side of the market.
2. We also work with search queries that are related to both sides of the market: supply and demand. These search queries are:
   1. "New construction"
   2. "New housing development"
   3. "New home construction"
3. The choice of these terms is based mainly on two criteria: meaningful economic connection and high correlation with BP.
4. The list of search terms is not intended to be exhaustive, but only a sample of the many ways researchers can use these freely available data to create their own indicators.

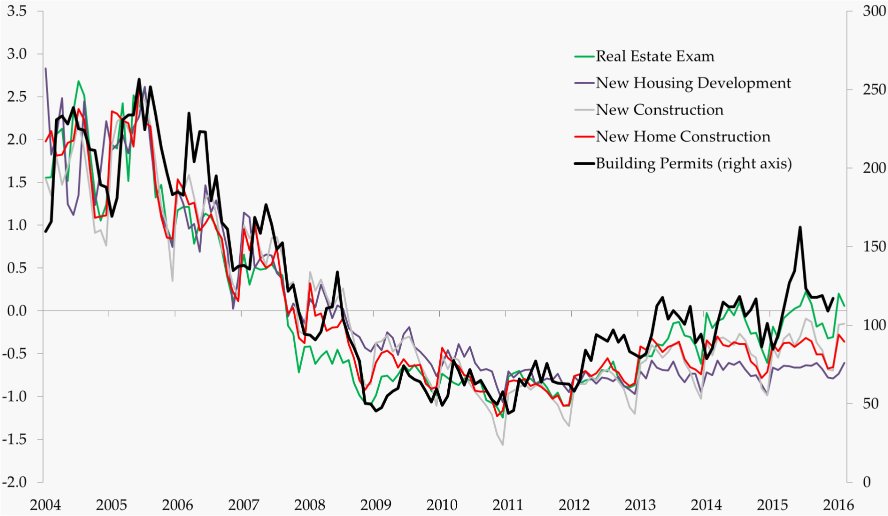
Data set: Building Permits

1. Monthly data on building permits (January 2004-December 2015, 144 observations)
2. Source: Census Bureau of Building Permits Survey.
3. Collected from a sample of local public permit-issuing agencies.
4. Representative at a national, state and city levels.
5. Released to the public on the 18th working day of the next month of reference, which is practically, a two month lag.
6. For instance, this would mean that in November 2017, the figure on BP is released today!!!

Data set: Google Trends

1. We have a total of 634 weekly observations for each search query.
2. Weekly data from January 2004 (first week) to February 2016 (fourth week).
3. We transform these weekly series into monthly series by taking averages.
4. We end up with 144 monthly observations of BP and 146 monthly observations for each search query.

Figure 2: Building Permits and Search Queries



Notes: For easier visualization, the intensity of search queries is standardized.

Key features in the data

1. Stochastic trends: traditional unit root tests cannot reject the null of a unit root in all our series. In first differences, this null is rejected.
2. Consequently, we build our models in first differences.
3. There is a strong seasonal pattern in the data.
4. The bulk of the analysis is carried out for not seasonally adjusted data, but we include the seasonal pattern in our models.
5. Seasonal adjustment is implemented for robustness with no critical differences.

Forecasting Models: In sample analysis

1. For our in-sample exercises we consider the following information set

$$I_t = \{g_{t+2}, g_{t+1}, g_t, g_{t-1}, ...; bp_t, bp_{t-1}, bp_{t-2} \dots\}$$

and the following direct regressions

$$b_{t+2+h} = \alpha^{(h)} + \beta_1^{(h)} b_t + \beta_2^{(h)} b_{t-1} + \beta_{12}^{(h)} b_{t+h-10} + \gamma^{(h)}(L) g_{t+2} + \varepsilon_{t+2+h}^{(h)}$$

2. Here

$$g_t = \Delta \ln(z_t)$$

where z_t denotes the search interest of the search query z at time t and

$$b_t = bp_t = \Delta \ln(BP_t)$$

3. Besides

$$\gamma^{(h)}(L) = \gamma_0^{(h)} + \gamma_1^{(h)} L + \gamma_2^{(h)} L^2 + ... + \gamma_{p(h)}^{(h)} L^{p(h)}$$

where L denotes the lag operator. We use BIC to find the lag length $p(h)$.

Backcasts

1. For instance, a backcast for bp_{t+1} built with information on I_t is constructed from the following regression

$$b_{t+1} = \alpha^{(-1)} + \beta_1^{(-1)} b_t + \beta_2^{(-1)} b_{t-1} + \beta_{12}^{(-1)} b_{t-11} + \gamma^{(-1)}(L) g_{t+2} + \varepsilon_{t+1}^{(-1)}$$

2. This expression can be estimated with observations 1 through T on bp_{t+1} and observations 2 through $T+1$ on g_t.
3. The backcast for b_{T+1} is constructed as

$$b_{T+1}^f = \alpha^{(-1)} + \beta_1^{(-1)} b_T + \beta_2^{(-1)} b_{T-1} + \beta_{12}^{(-1)} b_{T-11} + \gamma^{(-1)}(L) g_{T+2}$$

once population parameters are replaced with OLS estimates.

Statistical Inference: In-sample analysis

1. Null and alternative hypothesis:

$$\begin{aligned} H_0 &: \gamma^{(h)}(L) = 0 \\ H_A &: \gamma^{(h)}(L) \neq 0 \end{aligned}$$

2. We carry out inference using a traditional Wald Test.

Table 1: In sample results

Wald test and p-values							
		backcast	*nowcast*	*forecasts*			
Google Trend		h =-1	h =0	h =1	h =2	h =3	h =6
New home construction	Wald test	136.241	0.086	32.659	13.837	4.038	2.283
	df	7	1	3	2	1	1
	P-value	0.000	0.769	0.000	0.001	0.045	0.131
New construction	Wald test	142.763	33.653	16.972	10.644	0.157	3.625
	df	6	3	2	1	1	1
	P-value	0.000	0.000	0.000	0.001	0.692	0.057
New housing development	Wald test	0.363	0.002	1.287	1.620	0.055	0.216
	df	1	1	1	1	1	1
	P-value	0.547	0.962	0.257	0.203	0.814	0.642
Real estate exam	Wald test	89.943	0.149	16.801	4.569	1.799	0.510
	df	5	1	2	1	1	1
	P-value	0.000	0.700	0.000	0.033	0.180	0.475

Notes: Wald test built with HAC according to Newey & West (1987, 1994)

Table 2: Backcasts results

Dep. Var.: dlog(bp)	nhc	nc	nhd	rex
dlog(z(1))	*0.1693******	*0.3171******	*0.0292*	*0.1412******
	(0.0677)	*(0.0848)*	*(0.0487)*	*(0.0467)*
dlog(z)	*0.224******	*0.3128******		*0.0860*
	(0.0593)	*(0.0916)*		*(0.0681)*
dlog(z(-1))	*0.2600******	*0.2967******		*0.2072******
	(0.0839)	*(0.0928)*		*(0.0732)*
dlog(z(-2))	*0.4893******	*0.2967******		*0.3424******
	(0.0797)	*(0.0733)*		*(0.0740)*
dlog(z(-3))	*0.5969******	*0.6135******		*0.3167******
	(0.0637)	*(0.1161)*		*(0.0457)*
dlog(z(-4))	*0.3124******	*0.2387*****		
	(0.0898)			
dlog(z(-5))	*0.1673******			
	(0.0693)			
dlog(bp(-1))	*-0.5756******	*-0.5088******	*-0.1840******	*-0.4031******
	(0.0757)	*(0.0773)*	*(0.0481)*	*(0.0453)*
dlog(bp(-2))	*0.1886******	*-0.1278*	*0.1006*	*0.035*
	(0.0732)	*(0.0809)*	*(0.0661)*	*(0.0561)*
dlog(bp(-12))	*0.3843******	*0.3705******	*0.6735******	*0.5045******
	(0.0524)	*(0.0582)*	*(0.0617)*	*(0.0549)*
R-squared	0.663	0.644	0.483	0.586

In sample comments

1. Improvements in R^2 are relatively high.
    1. $R^2 = 0.48$ without google trends.
    2. $R^2 = 0.66$ with "new home construction".
    3. $R^2 = 0.64$ with "new construction".
    4. $R^2 = 0.48$ with "new housing development".
    5. $R^2 = 0.59$ with "real estate exam".
2. Even important gains in *in-sample* fit may be simply due to overfitting.
3. There is a real need to carry out inference with a method able to deal with this issue.

Out-of-sample analysis

1. We simulate a forecasting exercise in real time.
2. Suppose we have $T+1$ observations on BP and $T+3$ observations on z_t.
3. We split the total sample in an estimation window of size R, and a prediction window of size P, such that $R+P=T+1$.
4. We estimate our forecasting models in rolling windows of size $R=50$, and generate forecasts for future values of BP: $BP_{R+1}, BP_{R+2}, \ldots$
5. We compute forecast errors for the models with and without google trends.
6. If the information from google trends is useful to predict building permits, we expect the Root Mean Squared Prediction Errors to be smaller for the model with google trends:

$$\mathbb{E}\left[e_1^2\right] > \mathbb{E}\left[e_2^2\right]$$

Here model 2 is the model with google trends.

Out-of-sample forecasting models

1. Model 1:

$$\begin{aligned} bp_t &= \alpha + \beta_1 bp_{t-1} + \beta_2 bp_{t-2} + \beta_{12} bp_{t-12} + \gamma(L) g_t + \varepsilon_t \\ g_t &= c + \phi_1 g_{t-1} + \phi_{12} g_{t-12} + u_t \end{aligned}$$

2. Model 2:

$$\begin{aligned} bp_t &= a + \gamma(L) g_t + v_t \\ (I - \theta L^{12})(I - \lambda_1 L - \lambda_2 L^2) v_t &= \omega_t \\ (I - \Theta L^{12})(I - \delta L) x_t &= u_t \end{aligned}$$

3. These models are used to construct multistep ahead forecasts using an iterated strategy.

Out-of-sample inference

1. We consider nested specifications because they allow us to evaluate whether the inclusion of Google search information is useful to reduce forecast errors.
2. Our out-of-sample evaluations consider the null hypothesis that all the coefficients in the lag polynomial $\gamma(L)$ are zero.
3. To test this null we consider the ENC-t test of Clark and McCracken (2001) using standard normal critical values as suggested by Clark and West (2007).
4. Pincheira and West (2016) show that this test performs well when evaluating multistep ahead forecasts using the iterated method (when the persistence of the processes is moderate).

ENC-t test or Clark and West test

1. We restrict ourselves to the following simple setup

$$\begin{aligned} Model\,1 \quad (null) \quad &: \quad y_{t+1} = X_t^T \beta + e_{t+1} \\ Model\,2 \quad (alternative) \quad &: \quad y_{t+1} = X_t^T \beta + Z_t^T \gamma + e_{t+1} \end{aligned}$$

2. Under the null hypothesis that $\gamma = 0$

$$\frac{2\sqrt{P}}{P} \sum_{t=R}^{R-P-1} \widehat{e}_{1t+1}(\widehat{e}_{1t+1} - \widehat{e}_{2t+1}) \rightsquigarrow N(0, V^*)$$

   where V^* is the long-run variance of $e_{1t+1}(e_{1t+1} - e_{2t+1})$.

3. Given a consistent estimator V of V^*, Clark and McCracken (2001) and Clark and West (2007) propose to test the null hypothesis with a one sided test based on the following t-statistic

$$ENC - t = \frac{\frac{2\sqrt{P}}{P} \sum_{t=R}^{R-P-1} \widehat{e}_{1t+1}(\widehat{e}_{1t+1} - \widehat{e}_{2t+1})}{\sqrt{V}}$$

Table 2: ENC-t tests with linear models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	MSPE-adjusted / CW							
	backcast	*nowcast*	*forecasts*					
Google Query	h=-1	h=0	h=1	h=2	h=3	h=6	h=9	h=12
New home construction	4.35 (2.37) 1.83**	12.02 (4.18) 2.87***	23.02 (7.70) 2.99***	36.69 (8.79) 4.17***	41.02 (9.17) 4.47***	29.15 (12.40) 2.35***	7.14 (14.51) 0.49	32.10 (34.83) 0.92
New construction	10.86 (3.66) 2.97***	21.67 (7.33) 2.96***	30.86 (10.66) 2.90***	53.04 (10.82) 4.90***	50.88 (9.09) 5.60***	41.30 (15.29) 2.70***	13.11 (21.73) 0.60	77.14 (48.19) 1.60*
New housing development	2.22 (2.68) 0.83	8.21 (3.89) 2.11**	17.23 (5.42) 3.18***	30.69 (9.31) 3.30***	24.07 (9.82) 2.45***	-0.51 (9.78) -0.05	-4.26 (15.72) -0.27	38.48 (34.84) 1.10
Real Estate Exam	8.61 (3.51) 2.45***	19.02 (6.10) 3.12***	25.31 (8.09) 3.13***	44.30 (9.63) 4.60***	45.78 (13.87) 3.30***	34.42 (18.29) 1.88**	29.06 (22.09) 1.32*	99.73 (52.11) 1.91**

Notes: ENC-t test built with HAC according to Newey & West (1987, 1994)

Table 3: ENC-t tests with SARIMA models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	MSPE-adjusted / CW							
	backcast	*nowcast*	*forecasts*					
Google Query	h=-1	h=0	h=1	h=2	h=3	h=6	h=9	h=12
New home construction	13.54	23.18	30.31	51.18	48.98	47.75	53.18	137.98
	(4.19)	(7.16)	(9.96)	(12.25)	(12.52)	(20.88)	(23.68)	(65.67)
	3.23***	3.24***	3.04***	4.18***	3.91***	2.29**	2.25**	2.10**
New construction	11.32	23.60	30.17	53.31	51.15	66.05	63.01	162.53
	(6.26)	(10.98)	(15.23)	(15.58)	(19.30)	(23.37)	(30.00)	(68.78)
	1.81**	2.15**	1.98**	3.42***	2.65***	2.83***	2.10**	2.36***
New housing development	-2.24	0.99	2.75	6.77	4.28	-0.34	14.47	38.49
	(2.12)	(2.10)	(3.27)	(4.33)	(6.38)	(12.10)	(18.83)	(35.04)
	-1.05	0.47	0.84	1.56*	0.67	-0.03	0.77	1.10
Real estate exam	3.34	6.88	10.17	17.96	22.41	11.88	11.63	47.39
	(3.59)	(5.73)	(7.30)	(10.35)	(13.11)	(16.18)	(20.37)	(44.35)
	0.93	1.20	1.39*	1.73**	1.71**	0.73	0.57	1.07

Notes: ENC-t test built with HAC according to Newey & West (1987, 1994)

Table 4: DMW test with linear models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		RMSPE and DMW test							
	RMSPE/DMW	*backcast*	*nowcast*	*forecasts*					
Google Query		h=-1	h=0	h=1	h=2	h=3	h=6	h=9	h=12
New home construction	RMSPE	8.48	9.74	10.93	12.42	12.15	13.92	16.44	21.44
	DMW	-0.58	1.46*	1.71**	3.11***	3.76***	1.21	-0.39	-0.18
New construction	RMSPE	8.34	9.51	10.89	12.13	12.01	13.84	16.48	21.11
	DMW	0.22	1.55*	1.41*	3.40***	4.08***	1.08	-0.36	0.11
New housing development	RMSPE	8.51	9.88	11.08	12.53	12.76	14.95	16.79	21.15
	DMW	-0.74	0.74	2.03**	2.68***	1.71**	-1.33	-1.17	0.13
Real Estate Exam	RMSPE	8.34	9.70	11.09	12.38	12.24	14.12	15.96	20.35
	DMW	0.24	1.12	1.09	3.05***	2.46***	0.63	0.40	0.85
	RMSPE benchmark	8.38	10.02	11.47	13.39	13.34	14.50	16.23	21.26

Notes: DMW test built with HAC according to Newey & West (1987, 1994)

Table 5: DMW tests with SARIMA models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		RMSPE and DMW test							
	RMSPE/DMW	*backcast*	*nowcast*	*forecasts*					
Google Query		h=-1	h=0	h=1	h=2	h=3	h=6	h=9	h=12
New home construction	RMSPE	8.15	9.05	9.74	10.54	10.90	13.50	16.29	21.09
	DMW	0.67	1.33*	1.49*	2.97***	2.85***	1.18	1.18	0.89
New construction	RMSPE	8.68	9.41	10.38	11.17	11.49	13.30	16.34	21.05
	DMW	-0.66	0.13	-0.01	1.03	0.93	1.22	0.87	0.84
New housing development	RMSPE	8.59	9.60	10.40	11.77	12.25	14.68	17.02	21.97
	DMW	-1.99	-1.00	-0.18	0.61	-0.12	-0.78	0.26	0.63
Real estate exam	RMSPE	8.48	9.67	10.46	11.71	11.93	14.61	17.35	22.34
	DMW	-0.88	-0.74	-0.31	0.48	0.75	-0.49	-0.38	0.12
	RMSPE benchmark	8.31	9.49	10.37	11.87	12.22	14.37	17.15	22.43

Notes: DMW test built with HAC according to Newey & West (1987, 1994)

Table 6: Sample RMSPE ratios between models with and without internet search queries:

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
RMSPE ratios									
Expanding Windows									
Google Query	Linear/Non Linear	*backcast* h=-1	*nowcast* h=0	*forecasts* h=1	h=2	h=3	h=6	h=9	h=12
New home construction	Linear	0.99	0.97	0.93	0.91	0.89	0.95	0.99	0.99
	Non Linear	0.95	0.94	0.91	0.88	0.88	0.94	0.96	0.96
New construction	Linear	0.96	0.93	0.94	0.88	0.88	0.93	0.97	0.93
	Non Linear	0.98	0.95	0.95	0.89	0.90	0.91	0.93	0.90
New housing development	Linear	0.99	0.97	0.97	0.94	0.96	1.02	1.01	0.98
	Non Linear	1.01	1.00	1.00	0.99	1.00	1.03	1.01	1.00
Real Estate Exam	Linear	0.99	0.97	0.95	0.90	0.90	0.97	0.99	0.96
	Non Linear	1.01	1.01	1.00	0.99	1.00	1.05	1.05	1.03
Rolling Windows									
Google Query		h=-1	h=0	h=1	h=2	h=3	h=6	h=9	h=12
New home construction	Linear	1.01	0.97	0.95	0.93	0.91	0.96	1.01	1.01
	Non Linear	0.98	0.95	0.94	0.89	0.89	0.94	0.95	0.94
New construction	Linear	0.99	0.95	0.95	0.91	0.90	0.95	1.02	0.99
	Non Linear	1.04	0.99	1.00	0.94	0.94	0.93	0.95	0.94
New housing development	Linear	1.01	0.99	0.97	0.94	0.96	1.03	1.03	0.99
	Non Linear	1.03	1.01	1.00	0.99	1.00	1.02	0.99	0.98
Real Estate Exam	Linear	0.99	0.97	0.97	0.92	0.92	0.97	0.98	0.96
	Non Linear	1.02	1.02	1.01	0.99	0.98	1.02	1.01	1.00

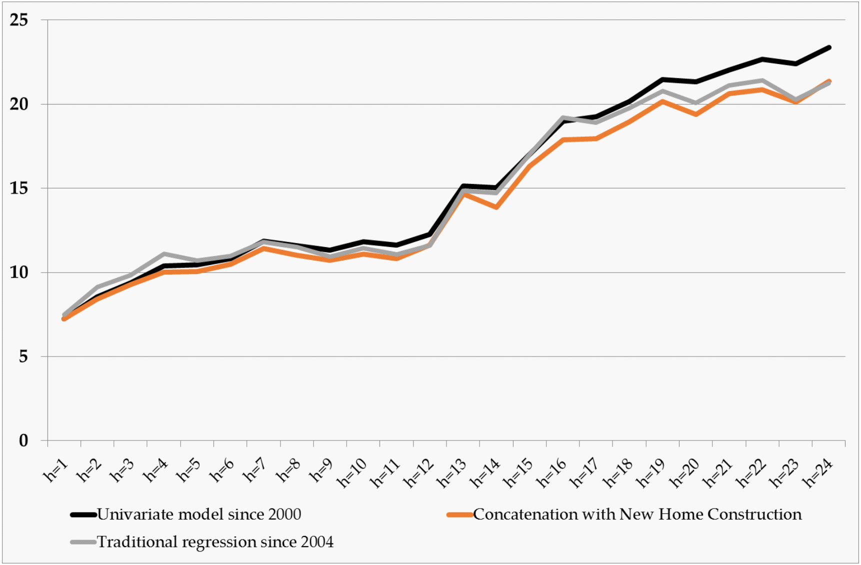
Using Google Trends Helps but

1. We need to restrict ourselves to use data only available since 2004. Nevertheless the time series on building permits is longer than that.
2. Key idea: to build an accurate backcast of building permits using google trends, and then to iterate forward using a good univariate model for building permits.
3. We consider the following univariate specification

$$bp_t = \alpha + \beta_1 bp_{t-1} + \beta_2 bp_{t-2} + \beta_{12} bp_{t-12} + \varepsilon_t$$

4. We construct the backcast using this specification augmented with new home construction as search query. For the univariate specification we consider data since January 2000.

Figure 3: RMSPE at different horizons



Concluding remarks

1. We provide strong evidence of the ability that some internet search queries have to generate backcasts, nowcasts and forecasts of building permits in the U.S.
2. In particular, search queries such as *new construction* and *new home construction* are shown to have relevant predictive information.
3. A natural avenue of future research considers the extension of this paper to evaluate predictability of measures of economic activity using techniques for data in mixed frequencies.