

Forecasting mortgages:

Internet search data as a proxy for mortgage credit demand

Branislav Saxa

Czech National Bank

Research Open Day, Prague, May 2015

The views expressed are the views of the author and do not necessarily represent the views of the affiliated institution.

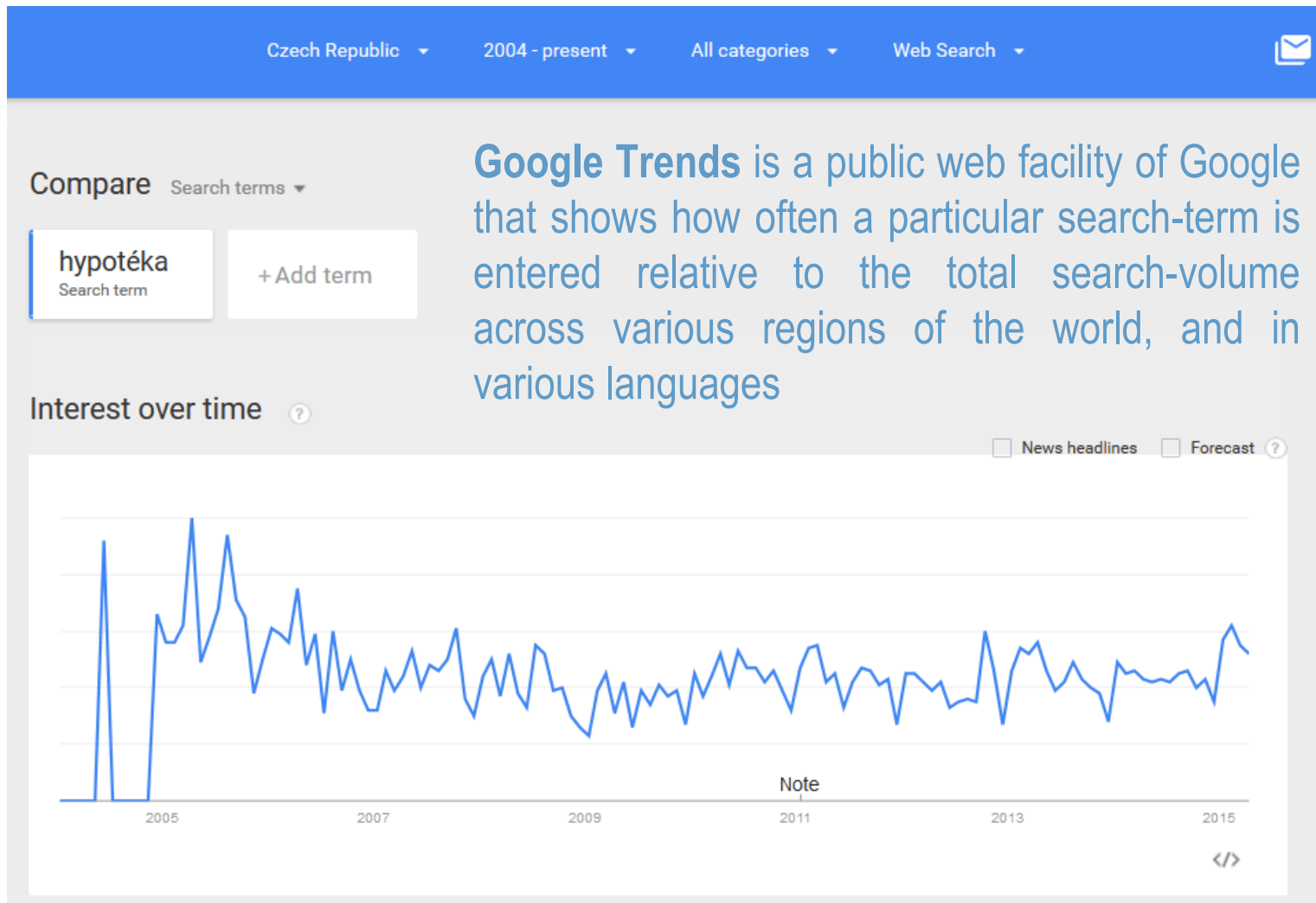
Motivation

- After the outbreak of crisis, **loan provision slowed** considerably in many countries
- Main reason unclear: **Lower demand** for loans or **lower willingness** of banks to provide loans?
- At about the same time, first applications employing **internet search data** appeared (e.g. Google Flu Trends)
- Challenge: Is it possible to use internet search data to **proxy the demand for mortgages?**

Overview

- Internet search data: What it is?
- Internet search data in the economic literature
- Empirical approach and results for the Czech Republic
 - Forecasting mortgages
 - Experimental indicator of restrictively tight bank lending standards and conditions
- Practical aspects of using Google Trends data

Internet search data: What it is?

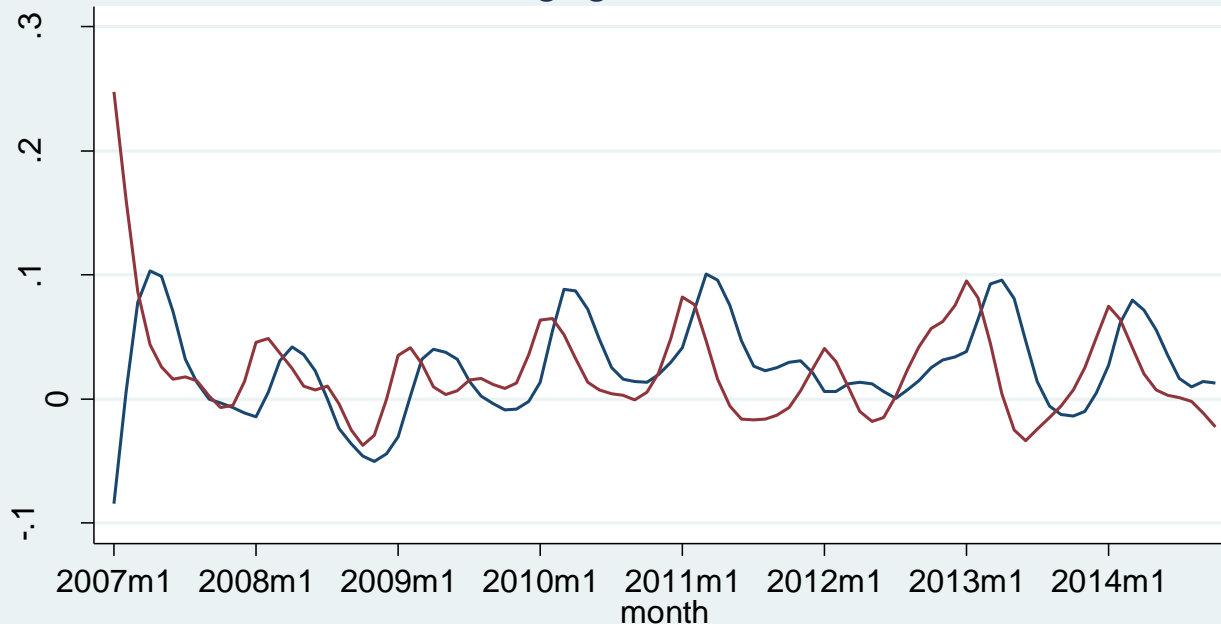


Internet search data in the economic literature

- Pioneers Choi and Varian (2009a, 2009b, 2012) use simple autoregressive models augmented with search engine data to produce near-term forecasts of **automobile sales**, **unemployment claims**, **travel destination** planning and **consumer confidence**
- Askitas and Zimmerman (2009), Pescyova (2011), McLaren and Shanbhogue (2011), D'Amuri and Marcucci (2012), Fondeur and Karame (2013): Nowcasts and near-term forecasts of **unemployment**
- Schmidt and Vosen (2009): Google Trends beat the forecasting performance of two most common **indicators of private consumption** in U.S. (the University of Michigan Consumer Sentiment Index and the Conference Board Consumer Confidence Index)

Data on Czech mortgages and stylized facts

Smoothed month-on-month growth rates of mortgages and searches



— Mortgage growth (m-o-m), smoothed
— Search growth (m-o-m), smoothed

Original series:

- Nominal volume of mortgages newly provided to households by banks in the Czech Republic (monthly, publication lag 1 month)
- Google data on search volume of the mortgage related words in the Czech language searched from the computers in the Czech Republic (weekly, no publication lag)

Transformation:

- Month-on-month growth rates smoothed using the Hodrick-Prescott filter with $\lambda = 10$

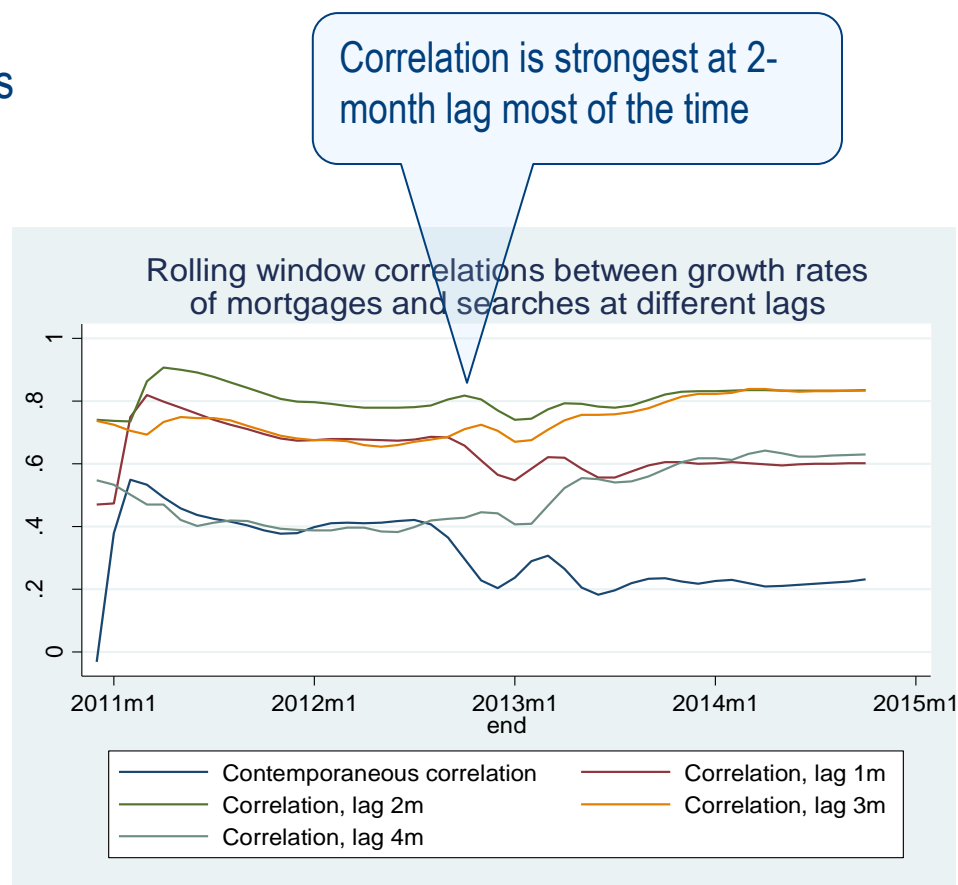
Data on Czech mortgages and stylized facts

What is the time lag between searching and providing mortgage? Is it changing over time?

Table A1: Crosscorrelations between mortgages and searches for different lags and subsamples

(significance levels in parentheses, lags with the highest correlation coefficient in bold)

Lag in months	Subsample			Whole sample
	2007m1-2009m12	2009m1-2011m12	2011m1-2014m10	2007m1-2014m10
0	-0.08 (0.66)	0.20 (0.24)	0.24 (0.11)	0.04 (0.67)
1	0.47 (0.00)	0.62 (0.00)	0.60 (0.00)	0.49 (0.00)
2	0.78 (0.00)	0.83 (0.00)	0.84 (0.00)	0.75 (0.00)
3	0.81 (0.00)	0.74 (0.00)	0.83 (0.00)	0.74 (0.00)
4	0.67 (0.00)	0.45 (0.01)	0.63 (0.00)	0.54 (0.00)



Forecasting mortgages

Out-of-sample forecasting exercise

- Estimation window extends from 2007m1–2008m8 to 2007m1–2014m9
- One and two month ahead forecasts are constructed
- MAE and RMSE of one-step ahead mortgage forecasts decreases by approximately 18% and 23%, respectively

	AR(1)	ARX	Change	Diebold-Mariano S(1)	p-value
One-step-ahead forecast					
MAE	0.1411	0.1162	-18%	4.25	0.00
RMSE	0.1919	0.1475	-23%		
Two-steps-ahead forecast					
MAE	0.1420	0.1150	-19%	4.27	0.00
RMSE	0.1924	0.1466	-24%		

$$AR(1): \Delta mortgage_t = \alpha + \beta \Delta mortgage_{t-1}$$

$$ARX: \Delta mortgage_t = \alpha + \beta \Delta mortgage_{t-1} + \gamma \Delta search_{t-2}$$

Forecasting mortgages

Big part of explained variation is seasonal. With seasonal term, searches still improve the forecast, but the improvement is not statistically significant

- MAE and RMSE of one-step ahead mortgage forecasts decreases by approximately 8% and 10%, respectively

	SAR(1)	SARX	Change	Diebold-Mariano S(1)	p-value
One-step-ahead forecast					
MAE	0.0985	0.0909	-8%	1.50	0.13
RMSE	0.1299	0.1168	-10%		
Two-steps-ahead forecast					
MAE	0.0992	0.0925	-7%	1.41	0.16
RMSE	0.1307	0.1182	-10%		

$$\text{SAR (1)} : \Delta \text{mortgage}_t = \alpha + \beta \Delta \text{mortgage}_{t-1} + \theta \Delta \text{mortgage}_{t-12}$$

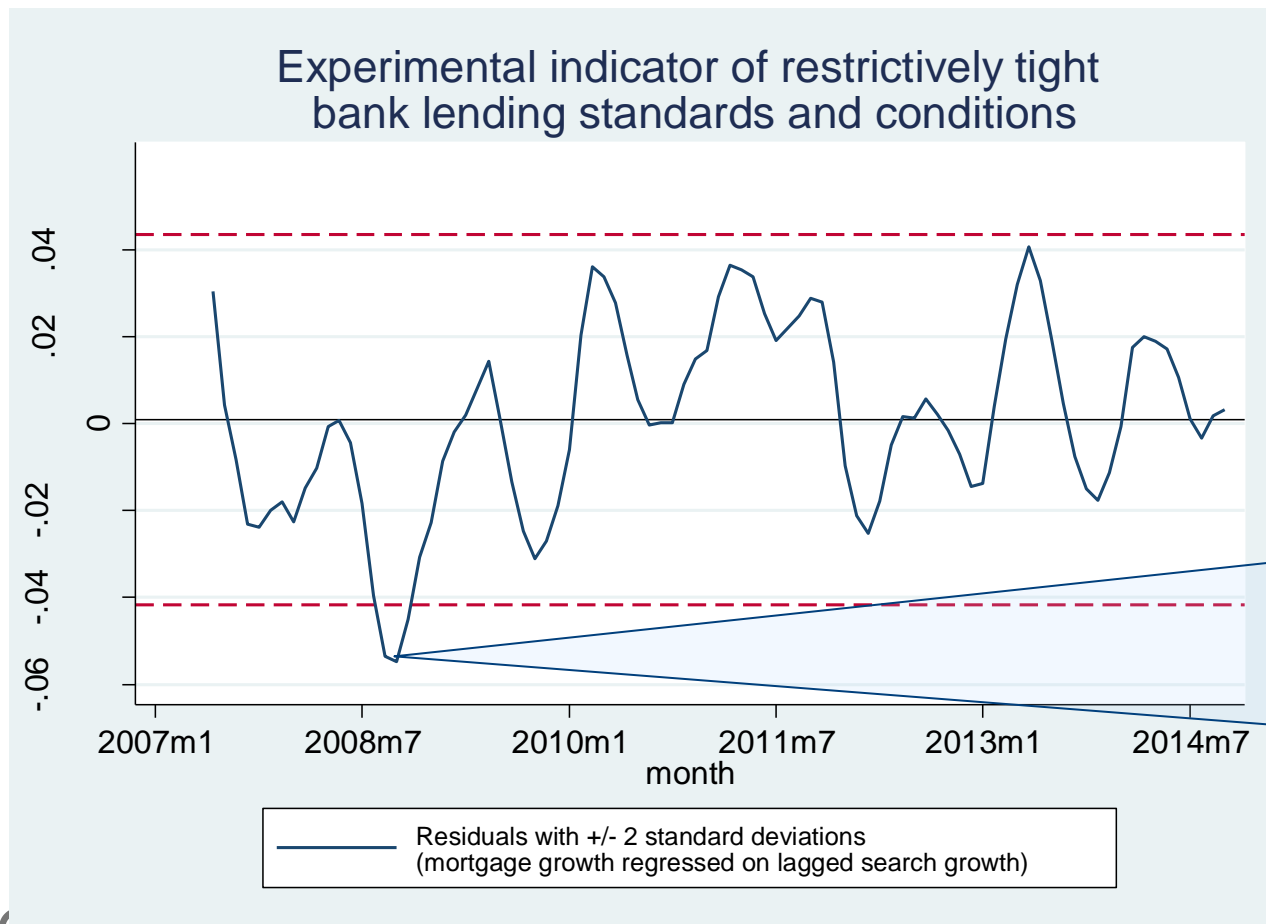
$$\text{SARX} : \Delta \text{mortgage}_t = \alpha + \beta \Delta \text{mortgage}_{t-1} + \theta \Delta \text{mortgage}_{t-12} + \gamma \Delta \text{search}_{t-2}$$

Experimental indicator of restrictively tight bank lending standards and conditions

- Experimental indicator of restrictively tight bank lending standards and conditions
 - Assumption so far: **Supply of mortgages is not limited**
 - Assumption from now on: Willingness of banks to provide mortgages changes over time and in some periods fewer mortgages are provided **not due to lower demand, but because of restricted supply**
 - Indicator: The smoothed growth rate of mortgages actually provided is regressed on the smoothed growth rate of searches lagged by two months. The **residuals** from this regression represent the part of the variation in mortgages that cannot be explained by the variation in demand for mortgages
 - Growth of demand substantially above the growth of mortgages actually provided can **signal a lower willingness of banks to provide mortgages.**

Experimental indicator of restrictively tight bank lending standards and conditions

Graph 3: Experimental indicator of restrictively tight bank lending standards and conditions



Eurozone BLS 3Q2008:
The net tightening of credit standards applied to loans to households for house purchase reached 36% (second-highest number in the history of the Eurozone bank lending survey; the only higher number was reported one quarter later)

Practical aspects of using Google Trends data

- Every data download provides indicator created using only random sample of all searches
 - Solution: Ten different data series obtained using the same query at different times were averaged for further use
- Use of 10 search terms¹ instead of one increased the usability of search data substantially

¹ “hypotéka” + “hypoteka” + “hypoteční” + “hypotecni” + “hypotéku” + “hypoteku” + “hypotéky” + “hypoteky” + “úvěr na bydlení” + “uver na bydlení”

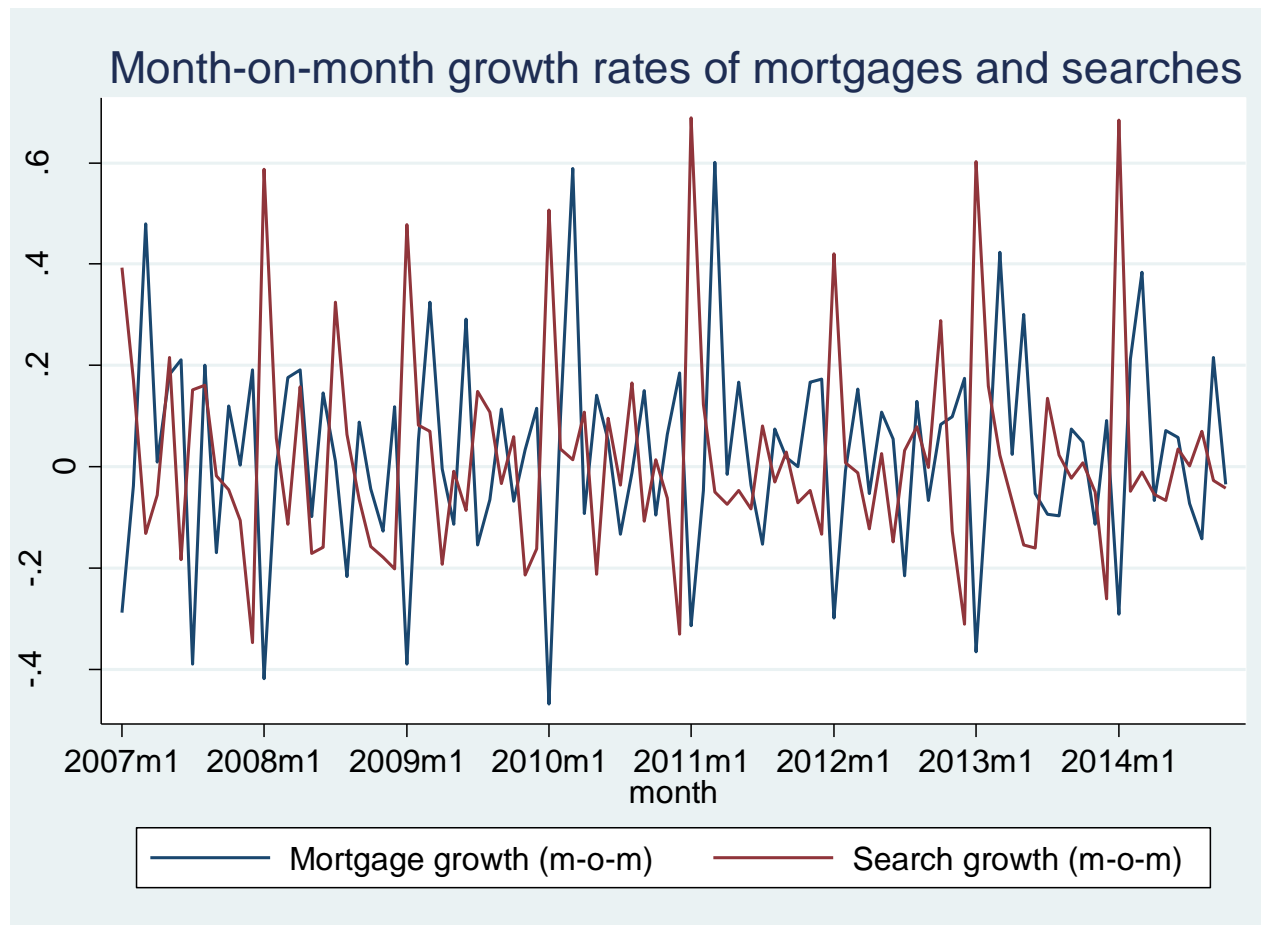
Conclusion

- The growth rates of searches and mortgages are **strongly correlated** and the volume of searches leads the volume of mortgages provided by **two months**
- Out-of-sample near-term forecast exercises show that the volume of searches **improves the short-term predictions** of mortgage lending
- Proposed **experimental indicator of restrictively tight mortgage credit standards and conditions** successfully identifies probably the most pronounced period of credit tightening in the history

Backup slides

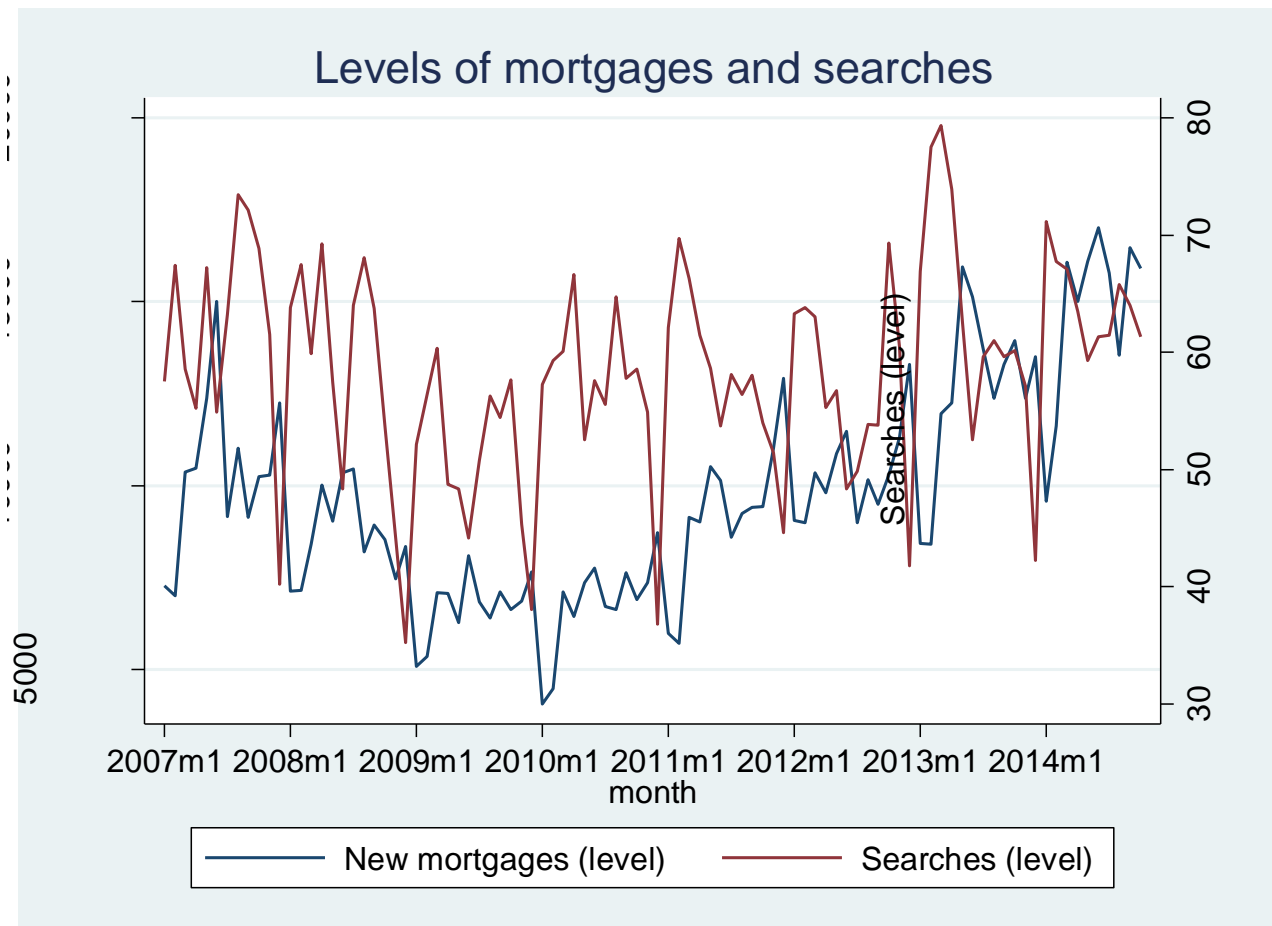
Raw m-o-m growth rates

Graph 1: Month-on-month growth rates of mortgages and searches



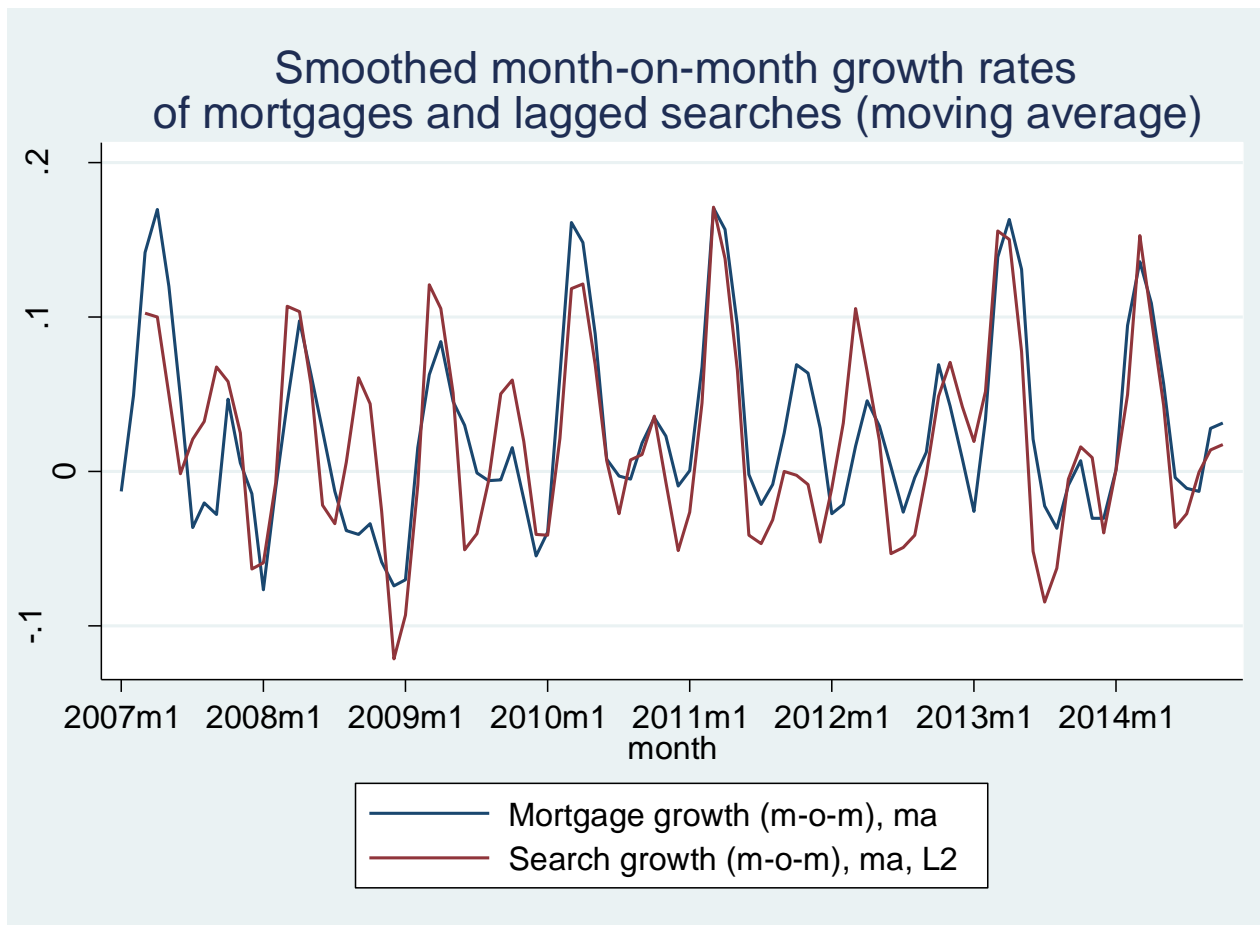
Levels of mortgages and searches

Graph A1: Levels of mortgages and searches



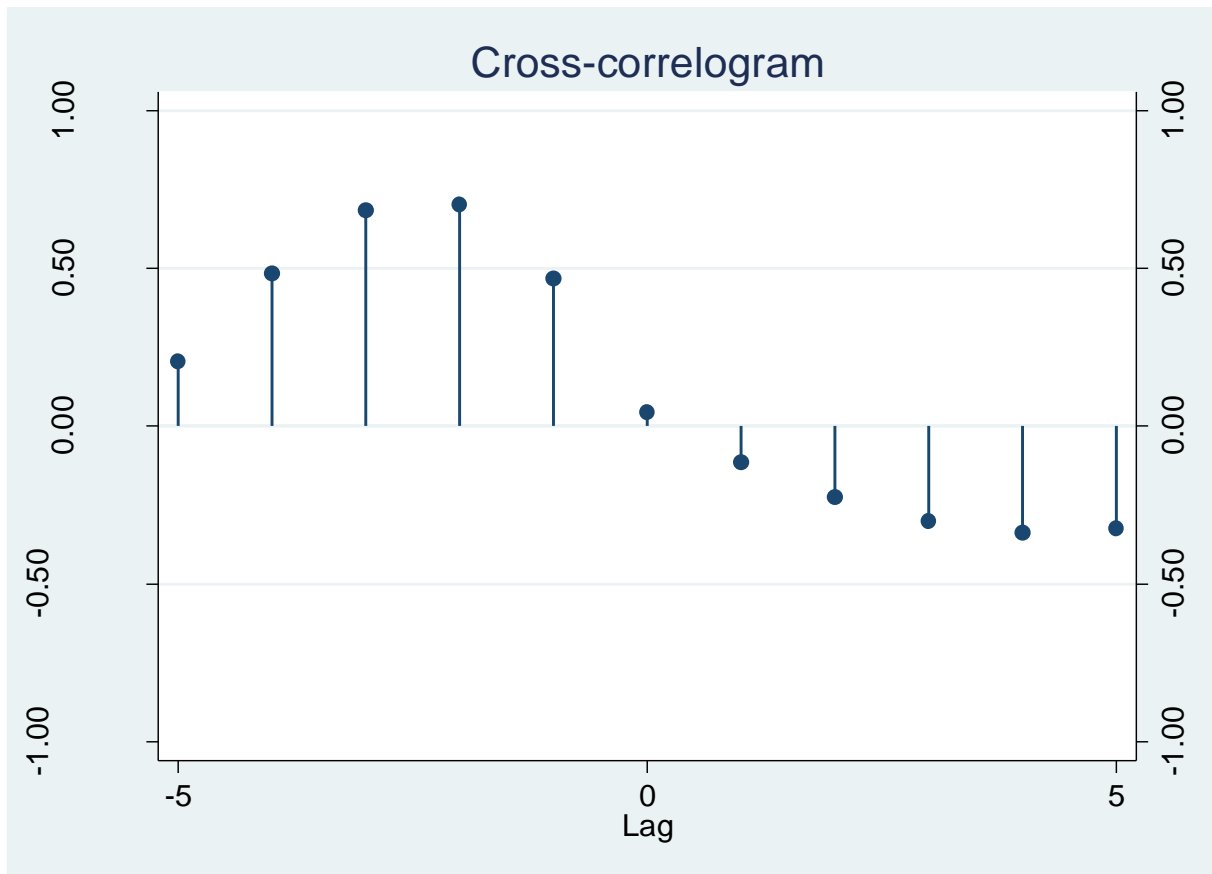
Moving average smoothing

Graph A2: Smoothed m-o-m growth rates of mortgages and searches (moving average)



Cross-correlogram

Graph A3: Cross-correlogram between smoothed m-o-m growth rates of mortgages and searches



Forecasting mortgages

Table 1: Variation in mortgage lending explained by amount of searching two months earlier
 (least squares estimation; the dependent variable is month-on-month growth of mortgage lending; standard errors in parentheses)

	AR(1)	ARX
L.Mortgage growth (m-o-m)	-0.24 ** (0.10)	-0.41 *** (0.09)
L2.Search growth (m-o-m)		0.58 *** (0.08)
Constant	0.03 * (0.02)	0.03 (0.02)
Adjusted R-squared	0.05	0.39
Number of observations	93	92

Note: * p<0.10, ** p<0.05, *** p<0.01

The amount of variation explained by the regression (proxied by adjusted R-squared) increases substantially once searches are included, from 0.05 to 0.39.

Variation in mortgage lending

Table 3: Variation in mortgage lending explained by amount of searching two months earlier and seasonal term

(least squares estimation; the dependent variable is month-on-month growth of mortgage lending; standard errors in parentheses)

	SAR(1)	SARX
L.Mortgage growth (m-o-m)	-0.17 ** (0.08)	-0.28 *** (0.07)
L12.Mortgage growth (m-o-m)	0.67 *** (0.07)	0.47 *** (0.08)
L2.Search growth (m-o-m)		0.35 *** (0.08)
Constant	0.01 (0.02)	0.01 (0.01)
Adjusted R-squared	0.53	0.61
Number of observations	82	82

Note: * p<0.10, ** p<0.05, *** p<0.01

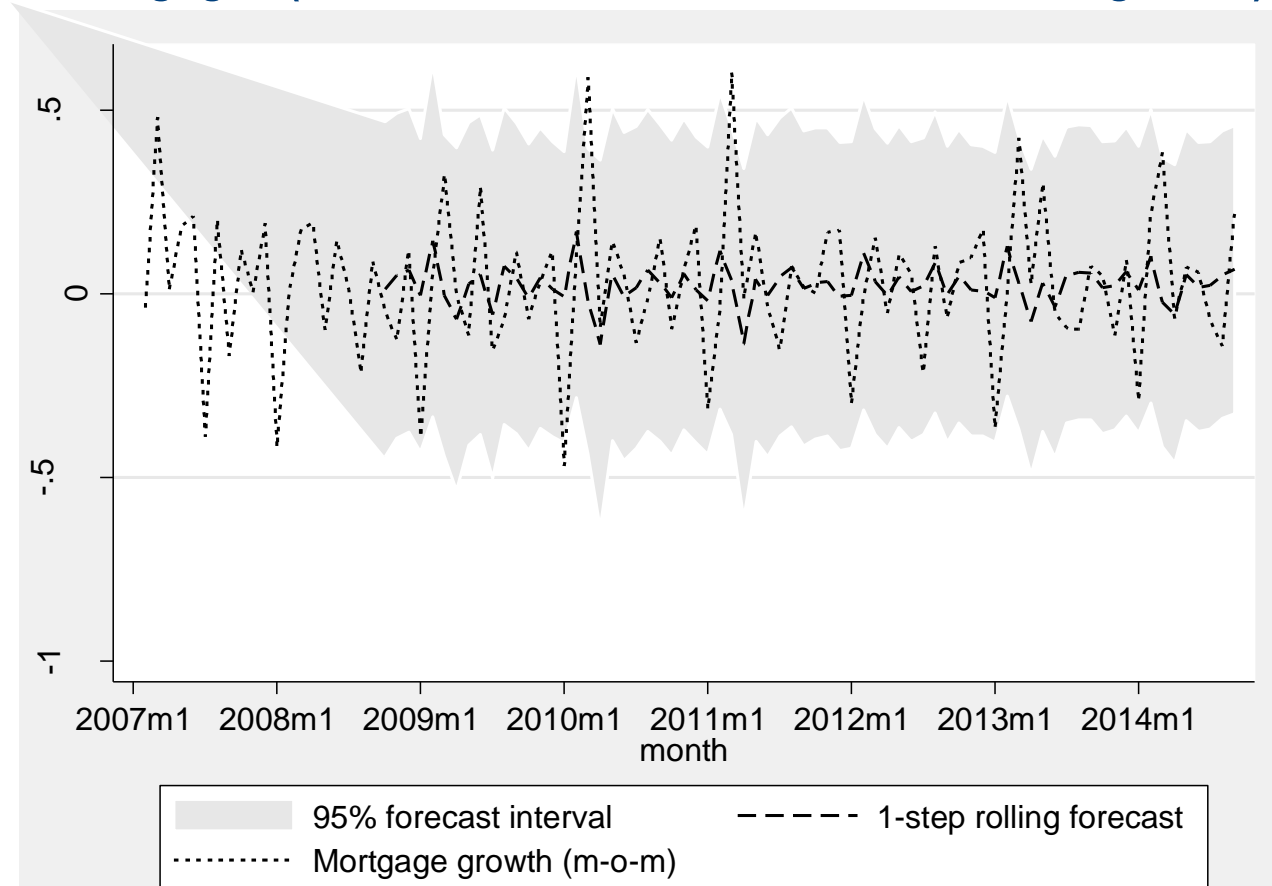
Summary statistics

Table A2: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
mortgages	94	9831.5	3027.5	4074.5	17021.4
searches	94	58.5	8.8	35.2	79.4
m-o-m mortgage growth	94	0.0250	0.2000	-0.4675	0.6015
m-o-m search growth	94	0.0224	0.2053	-0.3468	0.6889
smoothed m-o-m mortgage growth (HP filter, $\lambda=10$)	94	0.0250	0.0367	-0.0842	0.1031
smoothed m-o-m search growth (HP filter, $\lambda=10$)	94	0.0224	0.0406	-0.0376	0.2477
experimental index	92	0.0000	0.0234	-0.0943	0.0407

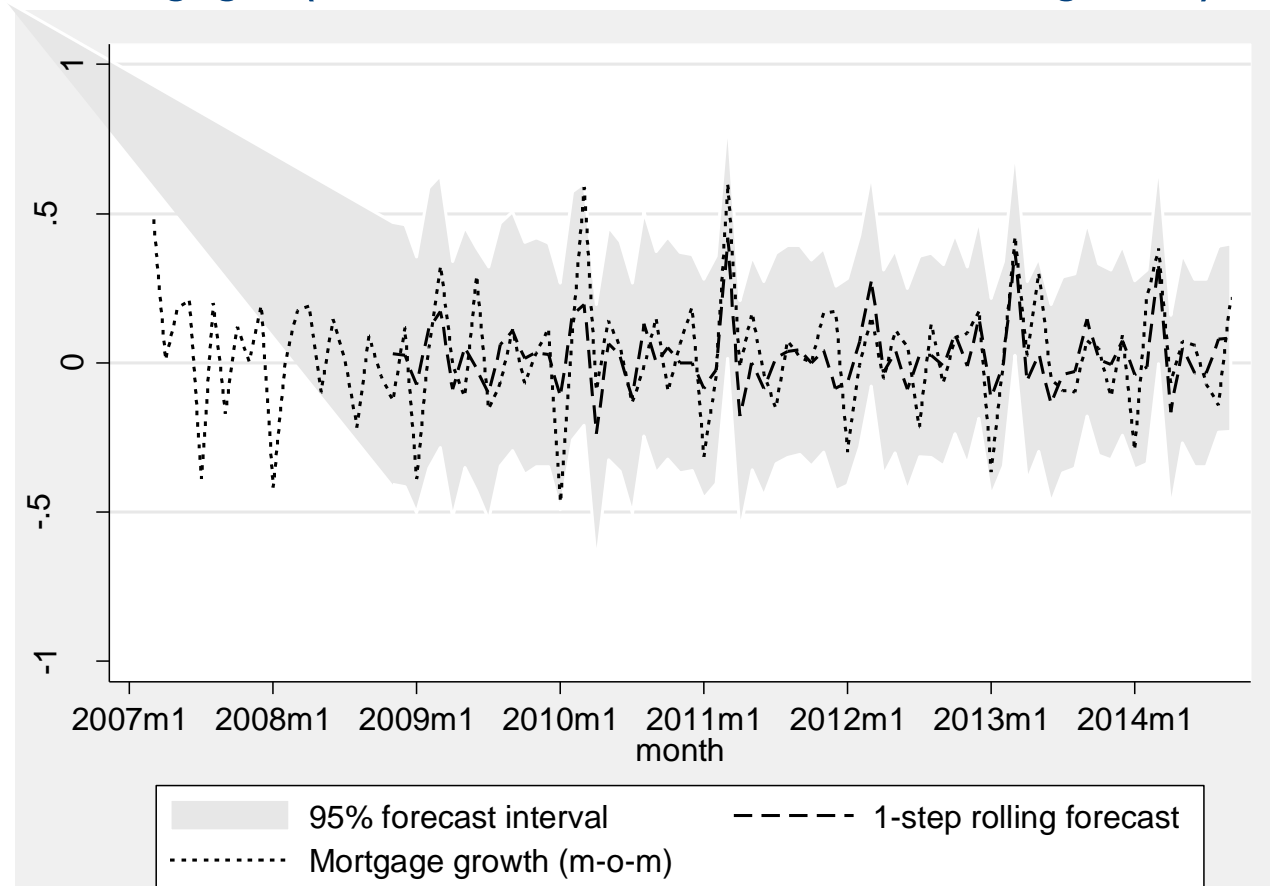
Out-of-sample forecasts

Graph A5: One-step-ahead out-of-sample forecasts of month-on-month growth rate of mortgages (without seasonal term, without search growth)



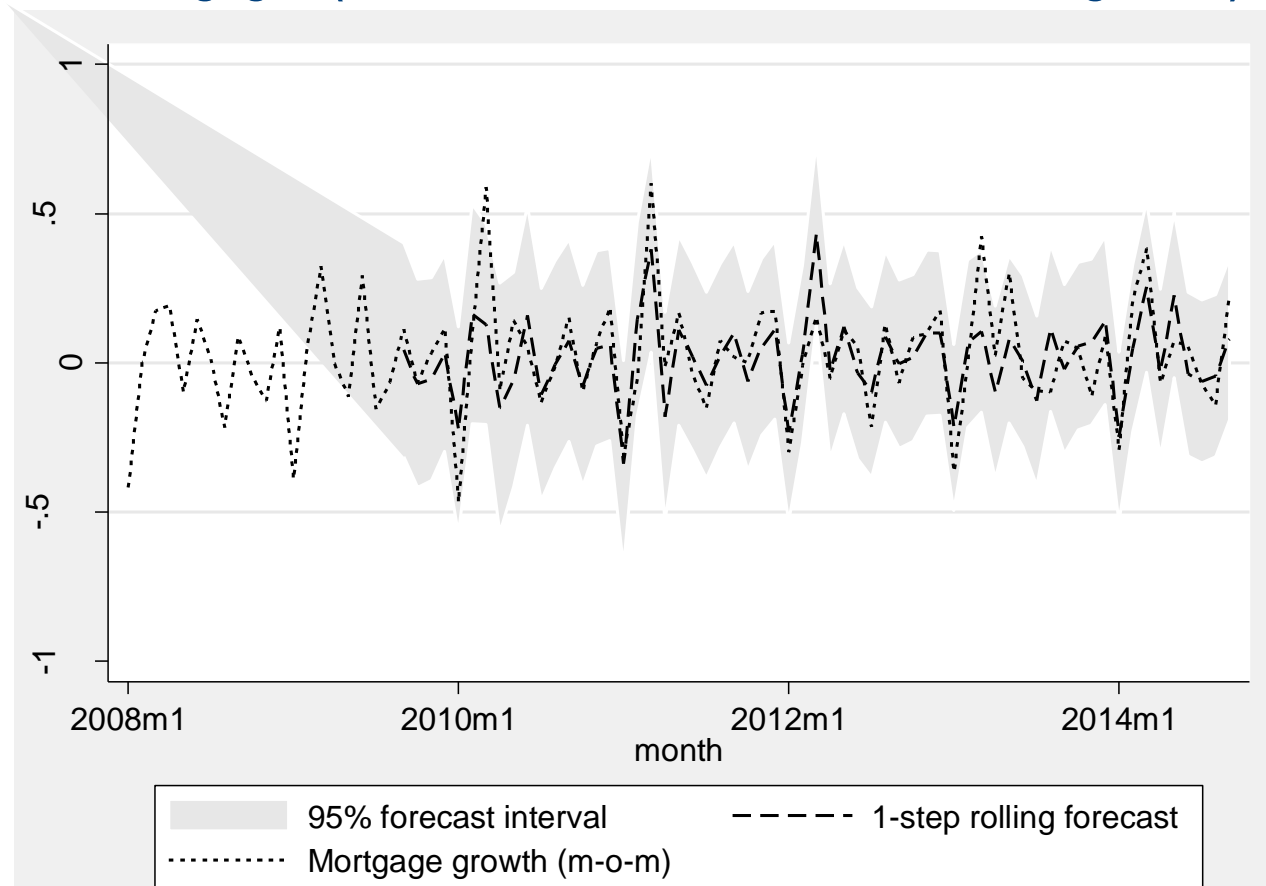
Out-of-sample forecasts

Graph A6: One-step-ahead out-of-sample forecasts of month-on-month growth rate of mortgages (without seasonal term, with search growth)



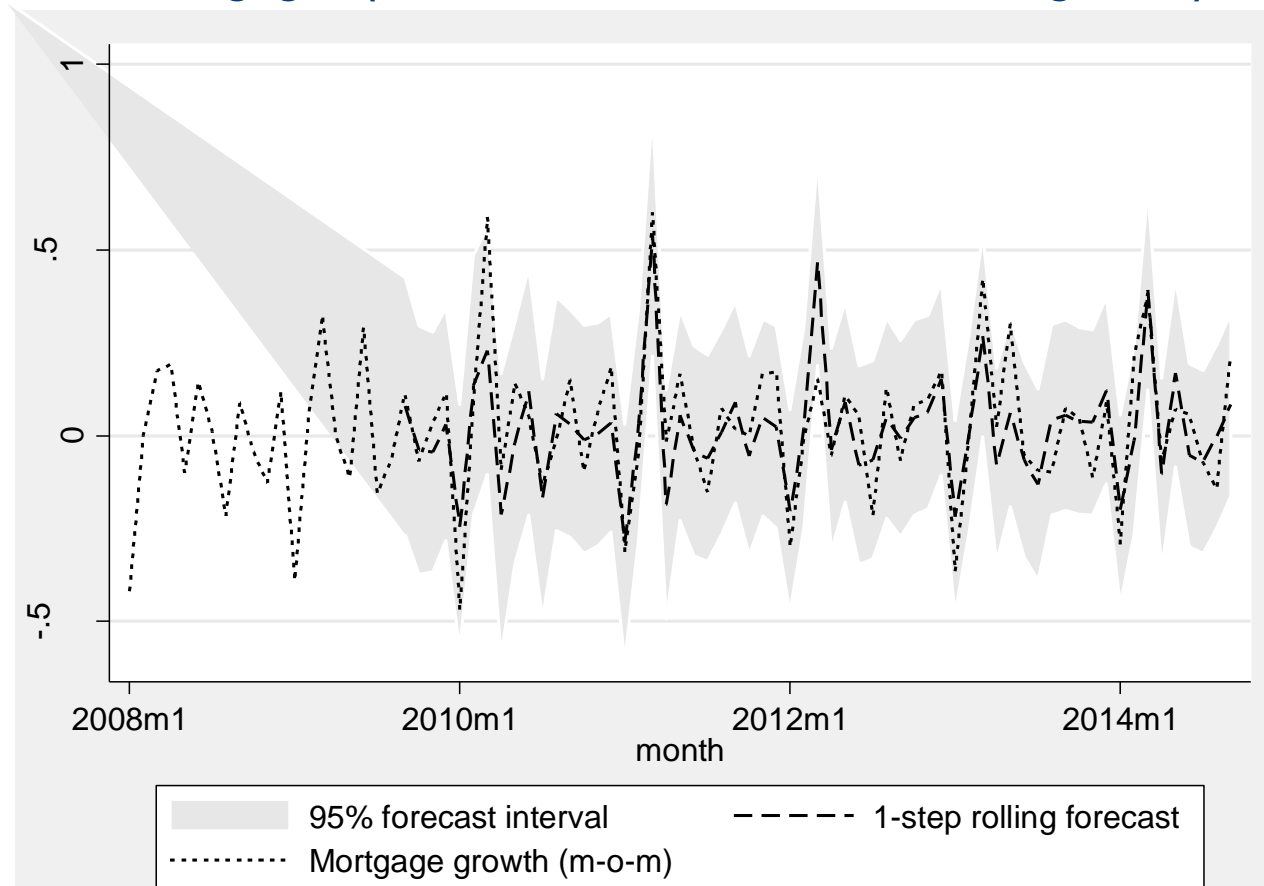
Out-of-sample forecasts

Graph A7: One-step-ahead out-of-sample forecasts of month-on-month growth rate of mortgages (with seasonal term, without search growth)



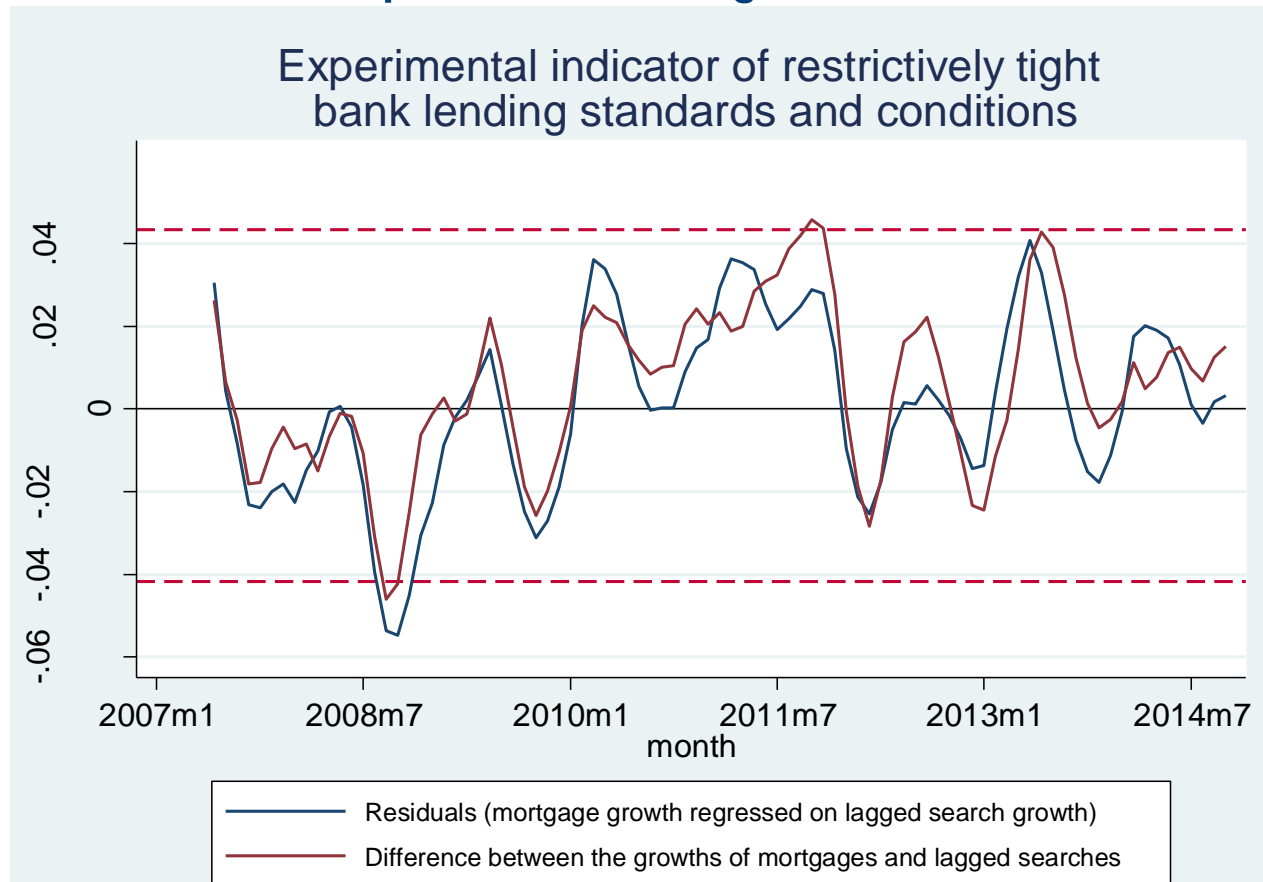
Out-of-sample forecasts

Graph A8: One-step-ahead out-of-sample forecasts of month-on-month growth rate of mortgages (with seasonal term, with search growth)



Experimental indicator

Graph A9: Comparison of baseline experimental indicator with version constructed as simple difference of growth rates



Experimental indicator

Graph A10: Comparison of baseline experimental indicator with version assuming lag of three months instead of two months

