STAFF MEMO

House price prediction using daily news data

NO. 5 | 2021

SARA J. KIRKEBY AND
VEGARD H. LARSEN

NORGES BANK

ISSN 1504-2596 (online)
ISBN 978-82-8379-196-9 (online)

House price prediction using daily news data*

Sara Jahr Kirkeby† and Vegard Høghaug Larsen‡

June 9, 2021

Abstract

We investigate whether information from news articles could improve predictions of house price inflation at a short forecast horizon. The Covid-19 pandemic led to a shutdown of the Norwegian economy on March 12th 2020. Large economic fluctuations posed challenges for models used to forecast economic developments. Our results indicate that news data contain valuable information about the direction of the housing market in periods of economic distress.

*The views and conclusions expressed in this publication are the authors' own and do not necessarily reflect those of Norges Bank. This paper should not be reported as representing the views of Norges Bank. Comments from Nina L. Midthjell, Sverre Mæhlum, Bjørn Helge Vatne improved the quality of this paper considerably. We also thank Ole Christian Bech-Moen, Henrik Borchgrevink, Kaja D. Haug and Torbjørn Hægeland for comments.

†Norges Bank, Financial Stability. E-mail: `Sara-Jahr.Kirkeby@norges-bank.no`

‡Norges Bank, Monetary Policy. E-mail: `Vegard-Hoghaug.Larsen@norges-bank.no; Corresponding author.`

1 Introduction

The Covid-19 pandemic led to a shutdown of the Norwegian economy on 12 March 2020. This crisis is unlike anything we have seen before and is generating large economic fluctuations that differ from traditional business cycles. The new environment is challenging for policy institutions trying to conduct mitigation policies because prediction models used for macroeconomic forecasting typically perform poorly when fluctuations are large. Knowing the state of the economy in real time and having good predictions for where it is heading are crucial to selecting the right mix of policy interventions. Throughout 2020, Norges Bank made use of new and nontraditional data sources to better understand the economic situation. Examples are transaction data from credit cards (Aastveit et al., 2020) and a financial news index for economic activity (Thorsrud, 2020) that provide information on consumption patterns and the growth rate of the economy.

The goal of our study is to analyse high-frequency data for predicting house price inflation, with particular focus on periods of distress. More precisely, we use data that is released daily or weekly, and made available in real time, to predict monthly house price inflation. The timing will be important, and in practice, house prices for a given month are not made publicly available until a few days into the following month. Our aim is to predict this monthly house price inflation using information that gradually becomes available within the same month, but before the actual house price statistics are made publicly available. Using daily and weekly data allows us to explore whether our predictions improve from having a few extra days of information.

Our main data source is text from printed newspaper articles. Our hypothesis is that news articles capture important developments in the economy and are a potential driver of economic fluctuations, see e.g. Beaudry and Portier (2014). We use Natural Language Processing (NLP) tools to calculate category-specific sentiment and uncertainty measures for the economy. We study four types of news labelled *Housing*, *Labour market*, *Monetary Policy*, and *Stock market*. In addition to the news data, we look at two high-frequency time series for the housing market. One series captures the stock of unsold houses, while the other captures the prices of sold houses relative to their asking price.

To test if our predictions improve with more up-to-date information, we estimate predictive models for monthly house price inflation where our independent variables are high-frequency news and housing data. Importantly, we include the variables as a 30-day backward-looking rolling mean where we calculate the rolling mean at different points within the month. That is, we use data from the first day, day six, day twelve, day eighteen and the last day of the month.

The model is estimated for the period 2004–2019 (from 2010 for the predictors from the housing market data), and the out-of-sample evaluation period is the Covid-19 pandemic from December 2019 until mid-March 2021. We find that our models perform better when more observations become available throughout the month. For the news data there is a monotonic improvement in model fit throughout the month.

For the Covid-19 pandemic, we find that the news data often correctly signal the direction of the housing market. In March 2020, when Norway closed down, we see an improvement in the prediction for house price inflation throughout the month, with the prediction becoming gradually more negative. Then, the model gives an increasingly positive signal about house price inflation when prices pick up again towards summer 2020. Interestingly, in contrast to most years, the model that only includes news data is the best model throughout the Covid-19 pandemic. This is in line with studies showing that text data is especially valuable for economic forecasting in periods of economic turmoil, see e.g., Kalamara et al. (2020) and Thorsrud (2020).

Several studies show that economic fundamentals do not explain all the observed variation in house prices and that expectations about future prices are to a great extent generated by information that is related to other factors (Shiller, 2007; Akerlof and Shiller, 2009; Ling, Ooi and Lee, 2015). More specifically, Akerlof and Shiller (2009) notably argue that “animal spirits” or an irrational exuberance of investors was a significant factor in the dramatic booms and busts in house prices in some markets. Ling et al. (2015) further examined the role of sentiment in house price dynamics. They construct sentiment measures based on surveys of home buyers, home builders and mortgage lenders in order to predict movements in house prices and find significant evidence that the prediction holds in the subsequent quarters. Moreover, they find that housing market sentiment and its effect on real house prices is highly persistent.

Recently, some studies have attempted to more directly examine the impact of news media sentiment on the housing market. First, Walker (2014) found a significant positive relationship between newspaper articles in the *Financial Times* and stock returns for listed companies engaged in the UK housing market. Walker (2016) subsequently analysed the direct housing market in the UK and found that news media Granger-caused real house price changes from 1993 to 2008. Soo (2018) first developed measures of housing sentiment for 34 cities across the United States by quantifying the qualitative tone of local housing news. She found that news media sentiment on the housing market has significant predictive power for future house prices, leading prices by nearly two years. More recently, Beracha, Lang and Hausler (2019) examined the relationship between news-based sentiment and commercial real estate performance in the United States. They found that news abstracts in the *Wall Street Journal* predict returns on commercial real estate up to four quarters in advance.

The above studies suggest that news-based sentiment can serve as an early housing market indicator, leading house prices by several quarters. However, to our knowledge, no academic research has investigated this effect using a shorter prediction horizon. This paper thus provides a first opportunity to examine the role of daily news media in the housing market.

The rest of the article is organised as follows: In Section 2 we describe the data and how we use the news articles. Section 3 describes the models we use for predicting house prices, while Section 4 presents the results. Section 5 concludes.

2 Data and text classification

In this section, we describe the data and how we transform the textual data into time series we can include in forecasting models.

2.1 Housing market data

We employ statistics for the housing market from *Eiendom Norge* and *Eiendomsverdi AS*. The statistics are based on sales brokered and advertised through *Finn.no*. Three time series variables are acquired: (1) monthly seasonally adjusted house price inflation, (2) daily number of unsold homes and (3) weekly achieved sales prices measured relative to asking prices. The first series is obtained from *Eiendom Norge*, while the last two are obtained from *Eiendomsverdi AS*. The data range for 'house price inflation' and 'unsold homes' variables are from January 2003 to March 2021, while the 'sales price/asking price' variable spans from January 2010 to March 2021. To ensure comparability of the variables and to mitigate noise in the daily and weekly data, the 'unsold homes' and 'sales price/asking price' variables are standardised and smoothed by calculating a 30-day backward-looking rolling mean. The housing market data is plotted in Figures 8 and 9 in Appendix B.

2.2 News data

The news data we use are articles from the printed version of *Dagens Næringsliv* available Monday to Saturday. The news articles were generously provided by the company Retriever through their "Atekst" database and were collected manually for the most recent years. The data we use are from the beginning of 2003 until mid-March 2021. Using printed news has the advantage that it has gone through an editorial process and should capture the most important business news.[1] Before

[1] There is also an editorial process for online news, but the much stricter page limit for printed news yields a more curated set of news stories.

Figure 1: Word clouds for selected news topics

Note: The news topics are represented as word clouds. The 100 words with the highest probabilities for a given news topic are shown, the size of the words corresponds to the probability of that word occurring in the topic distribution.

the classification, we remove numbers, punctuation, given names and last names. We also remove stopwords, which are words that do not convey any meaning about the underlying content of the text, such as *and*, *this*, and *could*. Lastly, we stem the words, i.e. we reduce inflected words to their word stem.

2.3 Classifying the textual data

The raw news data are unstructured and high-dimensional. To investigate if news is useful for predicting house prices, we classify the news articles into a set of underlying (latent) topics. For the classification, we rely on the topic model by Blei et al. (2003) called Latent Dirichlet Allocation

(LDA). The LDA is one of the most successful models developed for Natural Language Processing (NLP) and is widely used in various fields. One reason for the LDA's popularity is that it classifies text in almost the same way as humans would, and it generates topics that are easy to interpret, see Chang et al. (2009). The classification follows Larsen (2020), and further details can be found there. The classification is unsupervised, meaning that the news topics are identified without any external input or pre-labelled articles. The researcher only has to decide on the number of topics. In Larsen (2020), the data is classified into 80 categories.

We focus on a small subset of the identified news topics and select four types of news that can affect participants in the housing market. We think of these types of news as capturing information that matters for buyers and sellers in the housing market and that can also potentially drive agents' behaviour. The topics are presented as word clouds in Figure 1 together with their respective labels. The topic model does not assign a label to the word distributions, and the labels are subjectively chosen based on the word clouds. The first type of news is classified as *Housing* and should capture news about the housing market directly. The second type is labelled *Monetary policy*, and this is included to capture news about interest rates. Then we include news classified as *Stock market*, which covers developments in financial markets. Lastly, we include *Labour market* news to capture the state of the labour market.

2.4 Sentiment and uncertainty indices

In addition to the classification described above, we compute a sentiment and uncertainty score for each article. We do this to get a measure of whether the news is reported as positive, negative or reflecting uncertainty. The measures are calculated by counting words from the categories *positive*, *negative* and *uncertainty*: the word lists for these categories are given in Appendix A. For an article i, the sentiment and uncertainty scores are computed as follows:

$$\mathcal{S}_i = \frac{\#\text{positive words} - \#\text{negative words}}{\#\text{total words}},$$

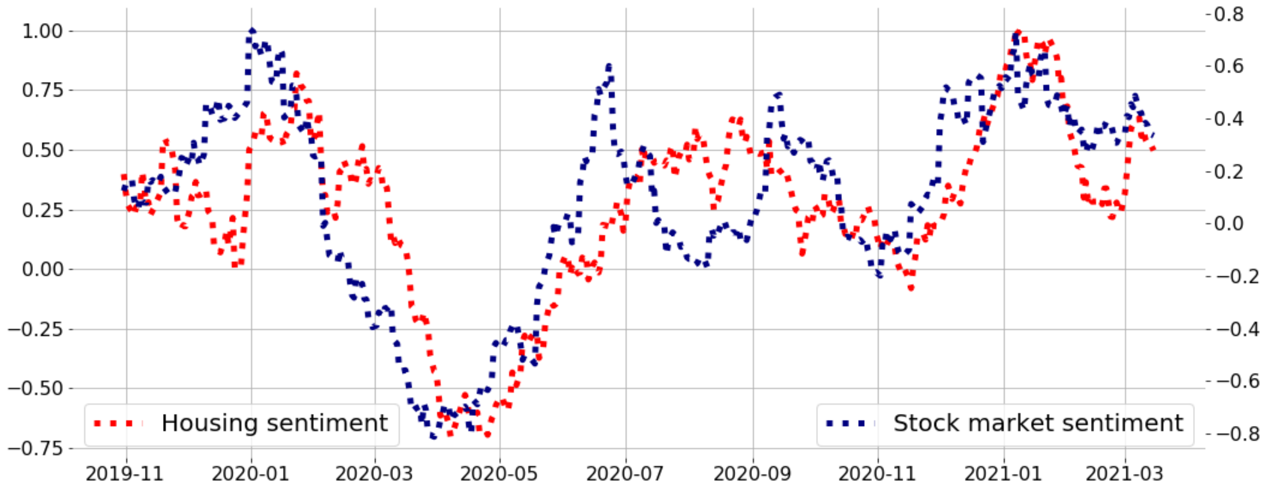
$$\mathcal{U}_i = \frac{\#\text{uncertainty words}}{\#\text{total words}}.$$

The goal is that a higher $\mathcal{S}_i$ corresponds to a more positive message in article i, and similarly a high $\mathcal{U}_i$ reflects high uncertainty. It is common to think of high uncertainty as negative for the economy (see Larsen (2020) for a discussion of positive and negative types of uncertainty). Combining the LDA classification of the news articles with the sentiment and uncertainty counts gives us topic specific time series for sentiment and uncertainty.

For some of the topics, the sentiment measure can be difficult to interpret because the positive and negative terms can have different meanings in different contexts. An example is the word *high*, which is part of the positive word list. However, this word is not positive if used in the context of *unemployment* or *inflation*. To avoid these issues with the interpretation of the sentiment measures, we instead choose the uncertainty measures for *Monetary policy* and *Labour market* news. For *Housing* and *Stock market* news, we choose the sentiment measures because they are less problematic to interpret. The sentiment data is shown in Figure 2, while the uncertainty data is shown in Figure 3.

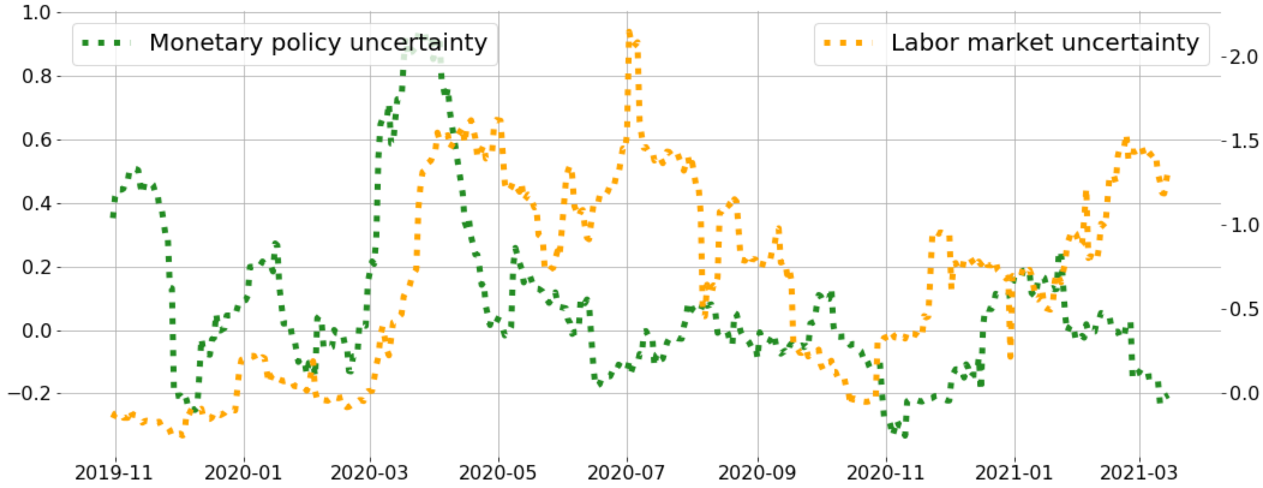
Looking at the measures we see that both *Housing* and *Stock market* sentiment falls sharply in the first part of 2020, and *Stock market* sentiment appears to lead *Housing* sentiment. Moving into the summer, we observe that the measures gradually improve, with some fluctuations in the rest of the period. The uncertainty measures both start to increase in late February. *Monetary policy* uncertainty appears to have been resolved relatively quickly, while *Labour market* uncertainty remained elevated until late 2020. To see the measures in a historical perspective, we plot the four news series from 1992 up until March 2021 in Figures 6 and 7 in Appendix B. Interestingly, we see that *Labour market* uncertainty reached unprecedented levels during the Covid-19 pandemic.

Figure 2: *Housing* and *Stock market* sentiment



Note: Housing sentiment on the left-hand scale and *Stock market* sentiment on the right-hand scale. The news measures are daily values, and we plot a 30-day backward-looking rolling mean.

Figure 3: *Monetary policy* and *Labour market* uncertainty



Note: Monetary policy uncertainty on the left-hand scale and *Labour market* uncertainty on the right-hand scale. The news measures are daily values, and we plot a 30-day backward-looking rolling mean.

3 Predictive models

To study the relationship between house price inflation and the high-frequency data, we estimate three predictive models for house price inflation with Ordinary Least Squares. The models are simple and focus on the predictive power of new information that becomes available through the month. The models are intended for prediction purposes only and should not be interpreted as capturing causal relationships.

First, Model 1 includes the four news measures from Figures 2 and 3 and lagged house price inflation as explanatory variables:

$$\Delta p_t = c + \beta_0 \Delta p_{t-1} + \beta_1 \mathcal{S}^{H}_{t,h} + \beta_2 \mathcal{S}^{SM}_{t,h} + \beta_3 \mathcal{U}^{LM}_{t,h} + \beta_4 \mathcal{U}^{MP}_{t,h} + \varepsilon_t, \tag{1}$$

where Δp is house price inflation, c is a constant, t indicates the month, h is days passed since the beginning of the month, $\mathcal{S}^{H}$ is housing sentiment, $\mathcal{S}^{SM}$ is stock market sentiment, $\mathcal{U}^{LM}$ is labour market uncertainty and $\mathcal{U}^{MP}$ is monetary policy uncertainty. The motivation for this specification is to start from an $AR(1)$-model, which often performs well in forecasting, and extend it with the news data to see whether the news can add some valuable information. Importantly, for all the models described here, we want to focus on the information set at different points in time, given by h. We estimate the models for five values of h: the first day of the month ($h = 1$), the sixth day of the month ($h = 6$), the twelfth day of the month ($h = 12$), the eighteenth day of the month ($h = 18$), and the last day of the month ($h \approx 25$). The main question we ask is what happens to our predictions when we have more information (i.e. when h increases)?

Next, Model 2 includes the two time series from *Eiendomsverdi AS* together with lagged house price inflation:

$$\Delta p_t = c + \beta_0 \Delta p_{t-1} + \beta_5 \mathcal{Q}_{t,h} + \beta_6 \mathcal{P}_{t,h} + \varepsilon_t, \tag{2}$$

where $\mathcal{Q}$ is the number of unsold properties and $\mathcal{P}$ is the price over the asking price. This specification is chosen with the same motivation as in Model 1, but here we want to investigate the value of the high frequency housing data as an alternative to the news data.

Lastly, Model 3 uses only news data:

$$\Delta p_t = c + \beta_1 \mathcal{S}^{H}_{t,h} + \beta_2 \mathcal{S}^{SM}_{t,h} + \beta_3 \mathcal{U}^{LM}_{t,h} + \beta_4 \mathcal{U}^{MP}_{t,h} + \varepsilon_t. \tag{3}$$

The motivation for this specification is to isolate the news effect and study what a pure news signal can tell us about house price growth. Our hypothesis is that for abrupt changes in the economy, the previous month's growth rate (Δp_{t-1}) might be of little importance. In some instances, it can potentially be better to ignore what happened last month and only focus on the current news.

In addition to the models described above, we estimate an $AR(1)$ model for use as a benchmark because we know it to be a good forecasting model historically, and it is often hard to beat.

4 Results

The results are presented in two parts. In Section 4.1 we present the in-sample results from the estimated models on the sample 2004–2019 (from 2010 for Model 2). Then, in Section 4.2 we describe how our models perform out-of-sample, when predicting house price inflation during the Covid-19 pandemic up to mid-March 2021.

4.1 In-sample results

Table 1 gives the parameter estimates from the three models, where we report results using data calculated on the first and the last day of the month. In Tables 3 – 5 in Appendix C, we report the estimated parameters for all models and all values of h. We include all the high frequency variables using a 30-day backward-looking rolling mean. That is, when $h = 1$, the high-frequency variables reflect approximately the average of the previous month, while on the last day, when $h \approx 25$, the variables reflect approximately the average of the current month.[2]

The estimated coefficients on the news variables are not significant in Model 1 when we use data from the first day. However, this changes when we move to the last day of the month, showing that more timely data improves the model. The coefficients on the sentiment measures are positive, which is intuitive given that a higher sentiment captures more positive news. For the uncertainty measures, we expect to see a negative effect, and this is the case for *Monetary policy* uncertainty. For *Labour market* uncertainty, we do not find any significant effects in any of the specifications. This is a bit surprising, but an explanation could be that the variation in *Labour market* uncertainty is captured by the other news-based indicators.[3]

For Model 2, we see that both $\mathcal{Q}$ and $\mathcal{P}$ are significant in all specifications. This result is as expected. More unsold houses predict lower house price inflation, while higher prices relative to asking prices predict higher house price inflation. The smaller β_0 in Model 2 suggests that there is some interaction between Δp_{t-1} and $\mathcal{P}_t$.

Model 3 is based on news data only and has a low explanatory power historically, with an explained variation (R^2) ranging from 9.3 % on the first day to 15.4 % on the last day of the month. For Model 3 we see significant coefficient estimates already from the first day, indicating that the news carries information from the start of the month.

The improvement in model fit moving into the month is reported in Table 6 in Appendix C. The table shows a gradual improvement in R^2 for all models, with the biggest improvement for Model 3 (66 %).

[2]We have experimented with using a backward-looking rolling window between 20 and 35 days, and our results are robust to changing this window.

[3]The highest monthly correlation between *Labour market* uncertainty and the other measures is 0.13 (*Housing* sentiment).

	h = First day of the month			h = Last day of the month		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
β_0 (Δp_{t-1})	0.5377***	0.2468**	–	0.5002***	0.2202**	–
c	0.0017***	0.0031***	0.0032***	0.0012**	0.0031***	0.0029***
β_1 ($\mathcal{S}^H$)	0.0014	–	0.0038**	0.0029*	–	0.0043**
β_2 ($\mathcal{S}^{SM}$)	0.0013	–	0.0040***	0.0027**	–	0.0053***
β_3 ($\mathcal{U}^{LM}$)	0.0016	–	0.0022	0.008	–	0.0009
β_4 ($\mathcal{U}^{MP}$)	−0.0014	–	−0.0049**	−0.0042**	–	−0.0062***
β_5 ($\mathcal{Q}$)	–	−0.0057**	–	–	−0.0076***	–
β_6 ($\mathcal{P}$)	–	0.0018***	–	–	0.0022***	–
R^2	0.339	0.233	0.093	0.378	0.280	0.154

Table 1: Model 1 and 3 are estimated for the period 2004 – 2019, while Model 2 is estimated for the period 2010–2019. The news variables are included at different h's as a 30-day backward-looking rolling mean. The independent variables (with the exception of Δp_{t-1}) are normalized. *, **, *** show significance at the 10%, 5%, and 1% level respectively.

To sum up, we find that including the last period's house price inflation is the most important factor for predicting the historical data. Including the high-frequency measures, both the news and the housing measures, gives a monotonic but modest improvement throughout the month. Next, we will see what our model can tell us in an episode of large and unexpected shocks.

4.2 Out-of-sample results

In this section, we use the models estimated up until 2019 to predict house price inflation through the Covid-19 pandemic (December 2019 until mid-March 2021).[4] We evaluate the short-run out-of-sample forecasting properties of the models. Table 2 reports the Mean Squared Error (MSE) relative to the first day for the models using data at the different points in time throughout the

[4]We stop our out-of-sample forecasting horizon at the cut-off date for the *Monetary Policy Report* by Norges Bank published in mid-March 2021.

	First day	Day 6	Day 12	Day 18	Last day	$AR(1)$
MSE for model 1 relative to day 1	1.00 (5.64)	0.99 (5.60)	0.92 (5.17)	0.88 (4.97)	0.79 (4.45)	1.09 (6.16)
MSE for model 2 relative to day 1	1.00 (7.35)	0.97 (7.10)	0.94 (6.94)	0.94 (6.92e-5)	0.88 (6.47)	0.80 (5.91)
MSE for model 3 relative to day 1	1.00 (5.25)	0.97 (5.08)	0.80 (4.21)	0.78 (4.07e-5)	0.65 (3.39)	1.17 (6.16)

Table 2: Mean Squared Error relative to day one throughout the Covid-19 pandemic. The last column gives the prediction error for the $AR(1)$ relative to day one for Models 1, 2 and 3. The MSEs in parentheses are multiplied by 10^5.

month.[5] We see that all models improve when more up-to-date information is included. Model 3 is the best model overall and we see large improvements moving into the month. Both news-based models (Models 1 and 3) are better than the $AR(1)$, while the $AR(1)$ beats Model 2 for this sample period.

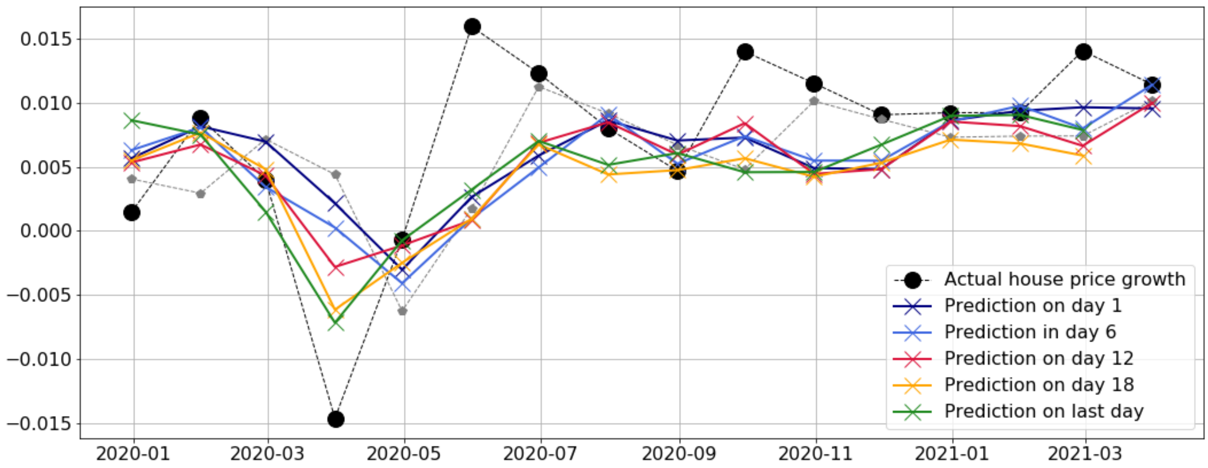
Figure 4 plots the predictions from the best performing model (Model 3) through the Covid-19 pandemic until mid-March 2021.[6] Zooming in on the March 2020 observation, we see a monotonic improvement of the predicted value throughout the month. We also observe that in the month after, the model captures the turning point, and the predicted value lies close to the actual value in April.

Next, we do a comparison of the best news-based forecasts (Model 3) with the forecasts from the Norges Bank *Monetary Policy Report* (MPR) through the Covid-19 pandemic until mid-March 2021. The report is released four times a year, in the middle of the months of March, June, September and December. In 2020 there was also an update in May when the decision was made to cut the policy rate to zero. The cut-off date for adding information into the analyses in the reports

[5]The MSE is a measure of fit calculated as the average squared difference between the estimated and the actual values.

[6]The same figures for Models 1 and 2 can be found in Appendix C.

Figure 4: Predictions through the Covid-19 pandemic from Model 3

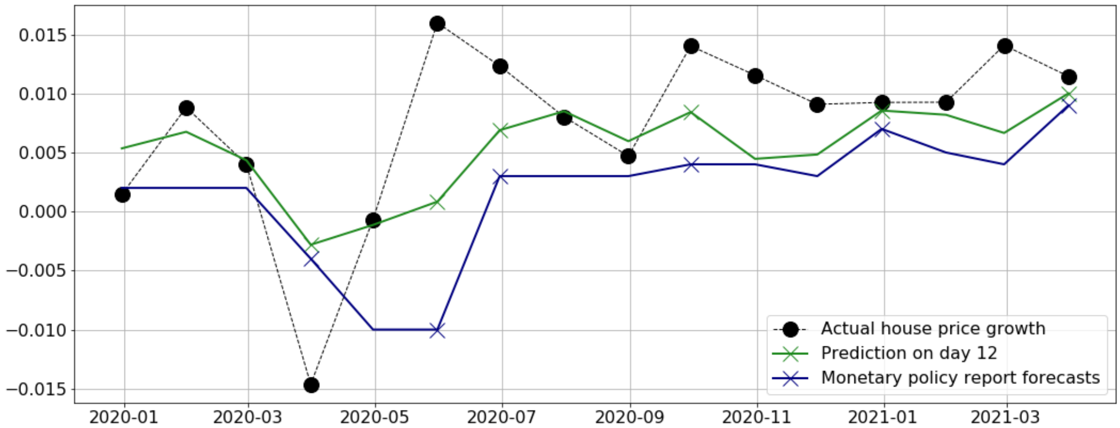


Note: Both predicted and actual house price inflation rates are plotted on the last day of the month, e.g., the growth rate from February to March is plotted on March 31.

usually approximately one week prior to publication. Hence, the most fair comparison is to compare the best-performing Model 3 news-based forecast from day 12 in each publication month with the corresponding MPR forecasts. Figure 5 shows the values through the Covid-19 pandemic predicted by the MPR and by Model 3, day 12. We find that the news-based model predicts slightly higher house price inflation than the MPR forecast throughout the period. Especially in April and May, when house price inflation unexpectedly bounced back from the initial fall in March, news data manage to capture the turning point. The results suggest that it would have been advantageous to employ news data as part of the forecasting analysis.

The models analysed in this paper are simple and do not rely to any great extent on history or dynamics. Our results suggest that information from news data becomes valuable in times of crisis. Given the simplicity of the models, it is not surprising that there is little to be gained from including the news variables in normal times when it is often hard to beat a simple $AR(1)$. We

Figure 5: Predictions through the Covid-19 pandemic from Model 3 versus Norges Bank Monetary Policy Report forecasts



Note: The *Monetary Policy Report* forecasts are published each quarter: in March, June, September and December. All forecasts are plotted on the last day of the month. The 'x' points in the figure show the MPR publication months through the Covid-19 pandemic. Comparing these with the news-based forecast on day 12 provides the fairest comparison.

focus on the Covid-19 pandemic, but we see a similar pattern for the Global Financial Crisis (GFC). Figures 12 and 13 in Appendix C show that this type of information could also help to indicate the direction of the housing market through the GFC in 2008.

For the Covid-19 pandemic, high-frequency news data are superior to high-frequency housing data in identifying the direction of house price growth. Our findings suggest that the news-based models could complement the existing forecasting models, and could provide useful information, especially when large shocks hit the economy.

5 Conclusion

We show that using more up-to-date information as it becomes available improves predictions for house price inflation. News data is especially valuable for predicting the direction of the housing market during episodes of economic distress and it can be particularly useful around turning points. In March 2020, when Norway closed down, news data gave a clear signal that prices were about to fall, and the prediction turned gradually more negative throughout the month. We also show that, for the Covid-19 pandemic, the best model is the one that uses only news data.

Comparing purely news-based forecasts with the forecasts in the Norges Bank *Monetary Policy Report*, we find that the news-based forecasts have smaller prediction errors throughout the Covid-19 pandemic. Going to March 2020, the news-based forecast captures the turning point earlier than any forecast we have access to. Our results suggest that the news-based forecast compliments the existing model framework, and policy makers should be especially interested in the news-based forecasts when large shocks hit the economy.

References

Aastveit, K. A., T. M. Fastbø, E. Granziera, K. S. Paulsen, and K. N. Torstensen (2020). Nowcasting Norwegian household consumption with debit card transaction data. Technical report, Norges Bank.

Akerlof, G. A. and R. J. Shiller (2009). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism.* Princeton university press.

Beaudry, P. and F. Portier (2014). News-driven business cycles: Insights and challenges. *Journal of Economic Literature 52*(4), 993–1074.

Beracha, E., M. Lang, and J. Hausler (2019). On the relationship between market sentiment and commercial real estate performance—a textual analysis examination. *Journal of Real Estate Research 41*(4), 605–638.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research 3*, 993–1022.

Chang, J., J. Boyd-Graber, C. Wang, S. Gerrish, and D. M. Blei (2009). Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems.*

Kalamara, E., A. Turrell, C. Redl, G. Kapetanios, and S. Kapadia (2020). Making text count: economic forecasting using newspaper text. Bank of England Working Paper.

Larsen, V. H. (2020). Components of uncertainty. *International Economic Review (forthcoming).*

Ling, D. C., J. T. Ooi, and T. T. Le (2015). Explaining house price dynamics: Isolating the role of nonfundamentals. *Journal of Money, Credit and Banking 47*(S1), 87–125.

Shiller, R. J. (2007). Understanding recent trends in house prices and home ownership. Technical report, National Bureau of Economic Research.

Soo, C. K. (2018). Quantifying sentiment with news media across local housing markets. *The Review of Financial Studies 31*(10), 3689–3719.

Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics 38*(2), 393–409.

Walker, C. B. (2014). Housing booms and media coverage. *Applied Economics 46*(32), 3954–3967.

Walker, C. B. (2016). The direction of media influence: Real-estate news and the stock market. *Journal of Behavioral and Experimental Finance 10*, 20–31.

A Word lists

Positive words:

akselerere, befeste, befestet, blomstre, bra, bedre, best, beste, begeistre, begeistret, opp, positiv, positive, positivitet, god, gode, godt, overveldene, ekspansjon, fantastisk, frisk, friskere, florerere, flott, frydefull, forlenge, fremgang, fremgangsrik, fremragende, rik, rikelig, rikere, robust, blomstre, overflod, tilfreds, tilfredsstille, utvide, utvidelse, generere, gullgruve, vant, vinne, vinner, vokse, vokser, vekst, vunnet, høy, høyne, forbedre, forbedring, forfremmelse, velstående, solid, solide, soliditet, stabil, stabile, steget, stige, stiger, steg, sterk, styrke, sunn, sunnhet, øke, øker, økning, super

Negative words:

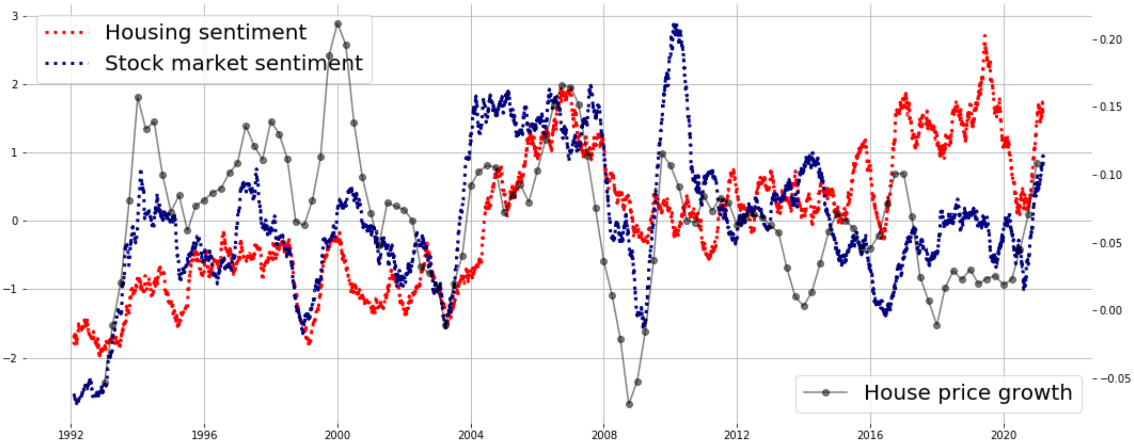
avta, blekne, brist, briste, dipp, død, dårlig, dårlige, dårligere, elendig, fall, falle, fallt, fallit, fiasko, havari, mangel, mangelfull, mislighold, ned, nedbemanning, mannefall, miserabel, nedbemanning, nederlag, nedgang, nedgangstider, nedskjæring, negativ, negative, negativitet, kollaps, kollapse, krakk, krasj, kritisk, kutt, kutte, resesjon, redusere, reduksjon, langsom, tilbake, tilbakegang, tilintetgjort, tragedie, treg, sank, senke, sunket, skrumpe, skuffe, skuffelse, sjokk, sparke, sparkes, sunket, svak, svakhet, svekke, svekket, svikte, sviktet, svinne, tap, tape, tapte, umulig, undergang, underskudd, ødeleggelse

Uncertainty words:

usikker, usikre, usikkert, usikkerhet, usikkerheter, usikkerheten, usikkerhetene

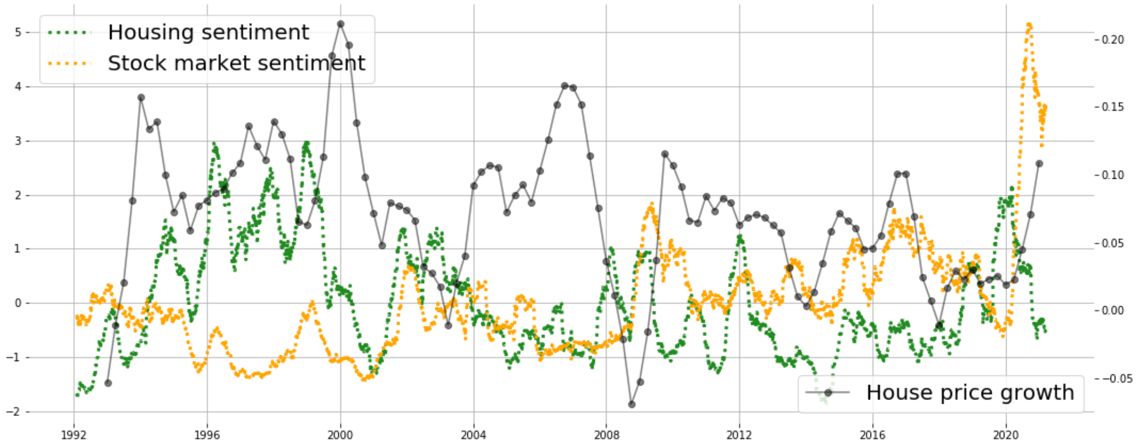
B Data

Figure 6: *Housing* and *Stock market* sentiment from 1988 to March 2021



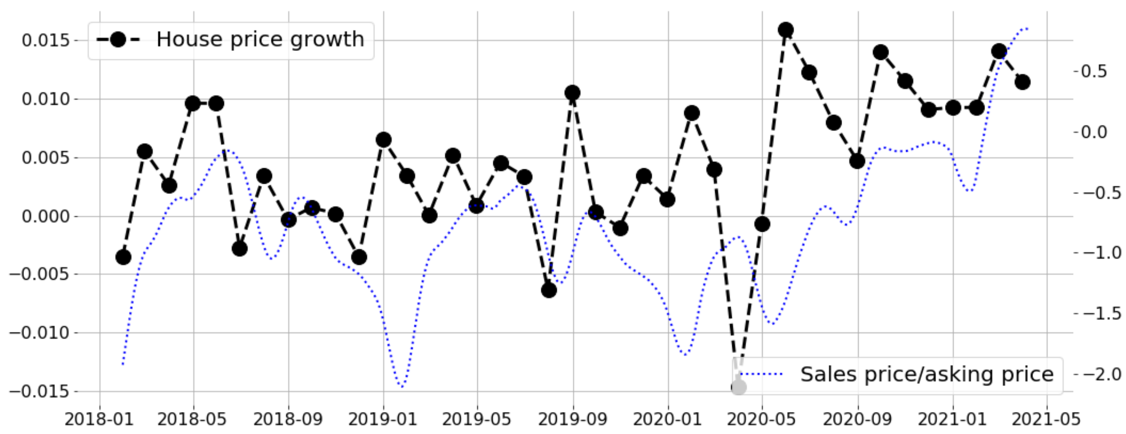
Note: The news measures on the left axis are normalized daily values and we plot a 180-day backward-looking rolling mean. The house price growth on the right axis are year-on-year growth rates of quarterly price index for existing dwellings from Statistics Norway.

Figure 7: *Monetary policy* and *Labour market* uncertainty from 1988 to March 2021



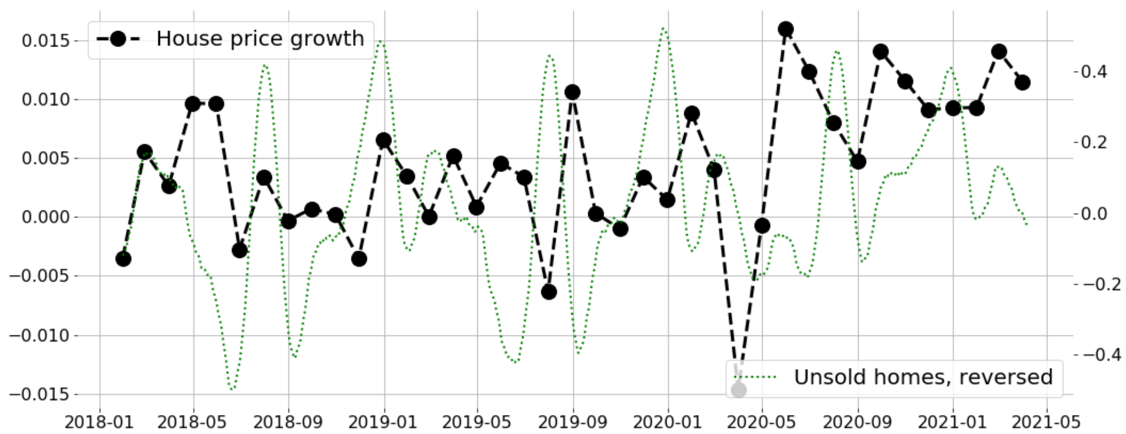
Note: See note to Figure 6.

Figure 8: *Sales price/asking price* and house price inflation from January 2018 to March 2021



Note: Sales price/asking price on the right-hand scale and *house price inflation* on the left-hand scale. The 'sales price/asking price' series is standardised, and we plot a one-month backward-looking rolling mean. Sources: Eiendom Norge, Eiendomsverdi AS and Finn.no.

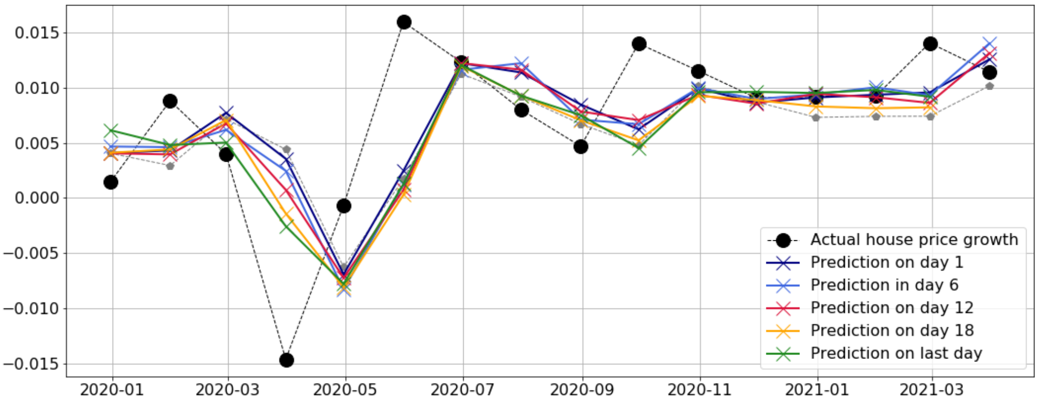
Figure 9: *Unsold homes* and house price inflation from January 2018 to March 2021



Note: Unsold homes, reversed on the right-hand scale and *house price inflation* on the left-hand scale. The 'unsold homes' series is standardised, and we plot a one-month backward-looking rolling mean. Sources: Eiendom Norge, Eiendomsverdi AS and Finn.no.

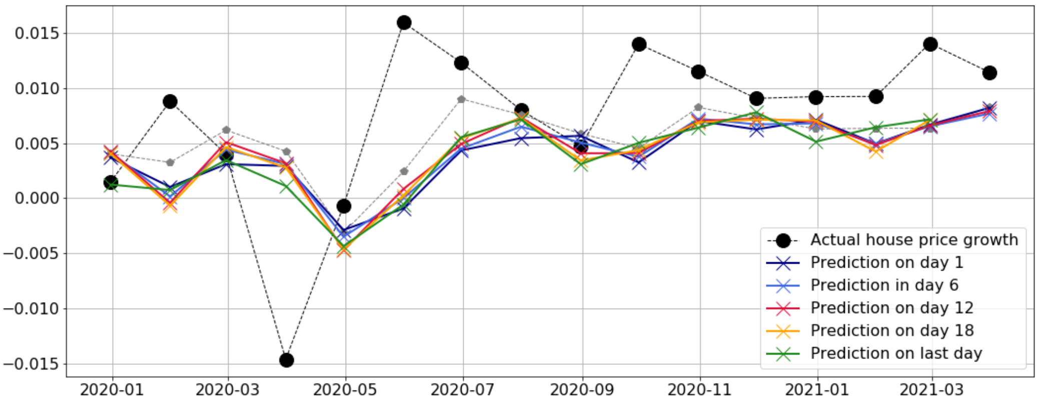
C Additional results

Figure 10: Predictions through the Covid-19 pandemic from Model 1



Note: See note to Figure 4.

Figure 11: Predictions through the Covid-19 pandemic from Model 2



Note: See note to Figure 4.

	Day 1	Day 6	Day 12	Day 18	Last day
β_0 (Δp_{t-1})	0.5377***	0.5210***	0.5131***	0.5106***	0.5002***
c	0.0017***	0.0014**	0.0014**	0.0016**	0.0012**
β_1 ($\mathcal{S}^H$)	0.0014	0.0021	0.0025	0.0021	0.0029*
β_2 ($\mathcal{S}^{SM}$)	0.0013	0.0018	0.0015	0.0015	0.0027**
β_3 ($\mathcal{U}^{LM}$)	0.0016	0.0020	0.0016	0.0004	0.008
β_4 ($\mathcal{U}^{MP}$)	−0.0014	−0.0028	−0.0039**	−0.0043**	−0.0042**
R^2	0.339	0.354	0.356	0.361	0.378
Adj. R^2	0.320	0.336	0.338	0.342	0.360

Table 3: Parameter estimates for Model 1. Standard errors in parentheses. *, **, *** shows significance at the 10%, 5%, and 1% level respectively.

	Day 1	Day 6	Day 12	Day 18	Last day
β_0 (Δp_{t-1})	0.2468**	0.2610***	0.2832***	0.2764***	0.2202**
c	0.0031***	0.0030***	0.0029***	0.0029***	0.0031***
β_5 ($\mathcal{S}$)	−0.0057**	−0.0052**	−0.0045**	−0.0039**	−0.0076***
β_6 ($\mathcal{P}$)	0.0018***	0.0016***	0.0015**	0.0017**	0.0022***
R^2	0.233	0.232	0.241	0.280	0.280
Adj. R^2	0.210	0.209	0.219	0.259	0.259

Table 4: Parameter estimates for Model 2. Standard errors in parentheses. *, **, *** shows significance at the 10%, 5%, and 1% level respectively.

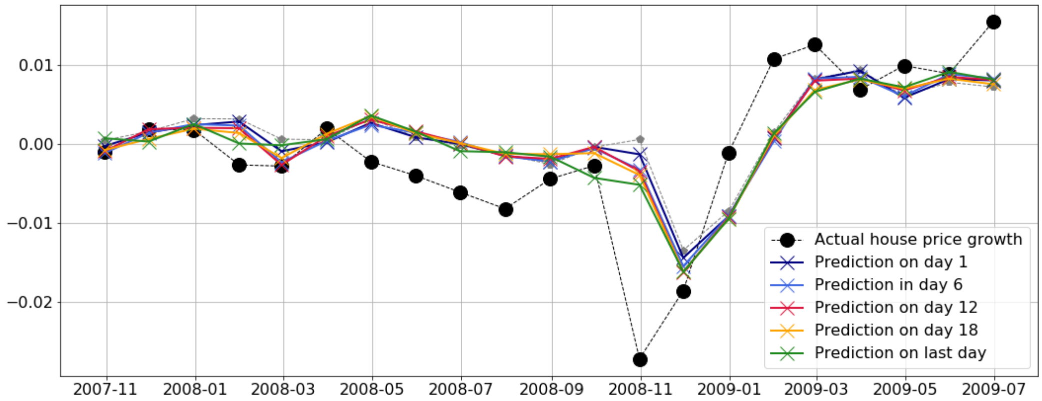
	Day 1	Day 6	Day 12	Day 18	Last day
c	0.0032***	0.0029***	0.0032***	0.0034***	0.0029***
β_1 ($\mathcal{S}^H$)	0.0038**	0.0043**	0.0039**	0.0037**	0.0043**
β_2 ($\mathcal{S}^{SM}$)	0.0040***	0.0042***	0.0038**	0.0038***	0.0053***
β_3 ($\mathcal{U}^{LM}$)	0.0022	0.0022	0.0017	-0.0003	0.0009
β_4 ($\mathcal{U}^{MP}$)	-0.0049**	-0.0054**	-0.0069***	-0.0067***	-0.0062***
R^2	0.093	0.114	0.124	0.132	0.154
Adj. R^2	0.072	0.094	0.104	0.112	0.134

Table 5: Parameter estimates for Model 3. Standard errors in parentheses. *, **, *** shows significance at the 10%, 5%, and 1% level respectively.

	First day	Day 6	Day 12	Day 18	Last day
Change in explained variance for model 1	- (0.339)	4% (0.354)	5% (0.356)	6% (0.361)	11% (0.378)
Change in explained variance for model 2	- (0.233)	0% (0.232)	3% (0.241)	3% (0.239)	20% (0.280)
Change in explained variance for model 3	- (0.093)	23% (0.114)	33% (0.124)	42% (0.132)	66% (0.154)

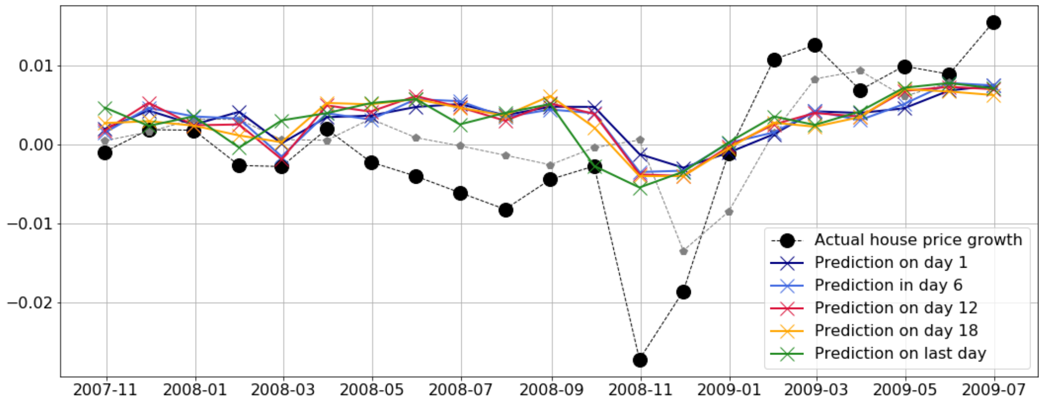
Table 6: Improvement in explained variation for the models relative to the first day of the month. For Model 1 and Model 3 the sample is 2004–2019, while for Model 2 it is 2010–2019. R^2 in parentheses.

Figure 12: Predictions through the financial crisis from Model 1



Note: See note to Figure 4.

Figure 13: Predictions through the financial crisis from Model 3



Note: See note to Figure 4.