Prepayments in depth - part 3: Drivers of prepayments and systematic usage of preliminary prepayments

Anders S. Aalund & Peder C. F. Møller

December 10, 2018

Contents

1 Introduction

In this paper we search for variables that can improve the estimation of prepayments. We test many types of variables in our random forest model and find that a combination of many variables can reduce the model bias we have observed over the last 9 years. We conclude that in order to significantly reduce the estimation errors from quarter to quarter it requires information that is not publicly available, such as the mortgage institutes campaign's for refinancing, development in margin fees on different loan products etc.
We describe how a classic prepayment model can benefit from a more systematic usage of preliminary prepayments. The findings are so promising that we expect that the set-up will be included in Nordeas official callable bond model when a model update is released during 2019.
Finally we show how a random forest model using many variables as well, as preliminary prepayments, long time ahead of the notification date can remove the observed bias. It can also notably reduce the noise in the individual observations.

2 Drives of estimation errors

We observe that on the overall level our prepayment model sometimes underestimate and sometimes overestimate prepayments. If we are able to forecast prepayments without this noise we can achieve more predictable returns for callable bonds which again makes the asset class more attractive for investors. In this section we use our machine learning set-up to search for variables that contains information that can be used to reduce the forecasting noise.

2.1 One swallow doesn't make a summer - but many indicates one!

In the previous articles in this series we showed that the random forest was more accurate than a classic prepayment model when estimating prepayments. One advantage of the random forest technique is, that it is more straight forward to test for more subtle relationship between a set of many variables.

We test a long list of variables such as *IO/non-IO, previous prepayments, change in yield curve slope, car sales, change in unemployment rate* etc. We find that each variable in itself doesn't improve the estimate much, but that the combination of many does.

On the out-of-sample analysis the average error becomes close to zero as opposed to now where the model on average estimate to high prepayments[1]. The magnitude of the *absolute* error is however only reduced slightly as can be seen in figure 1. In other words, we are able to locate a combination of variables that on average gives a much less biased fit, but still has significant variance.

This will make returns on high coupon portfolios more predictable and it will also make issuance estimation more accurate.

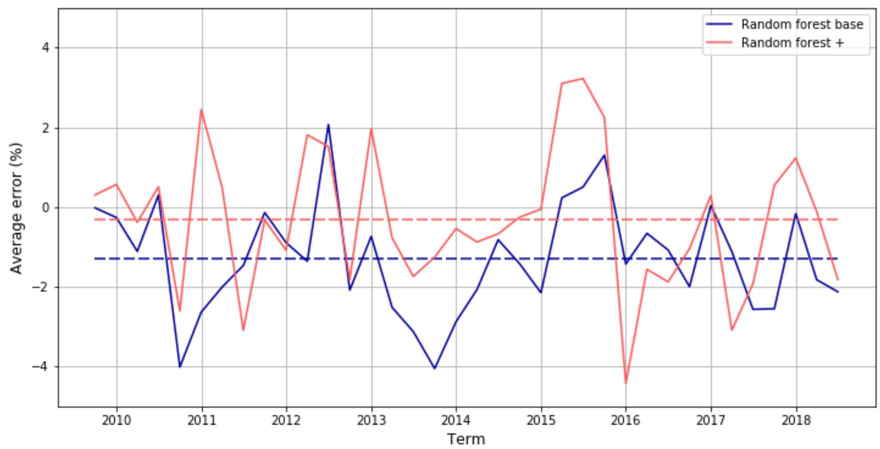

Figure 1: Average error since 2010 of out-of-sample analysis (described in previous articles in this series). Dashed line shows average levels. *base model* includes *NPV* and *poolfactor* and *base model+* includes many other variables. Source: Nordea

[1]Since the prepayment model minimizes the *square* error and there is asymmetric distribution of the error this does not indicate a model calibration issue.

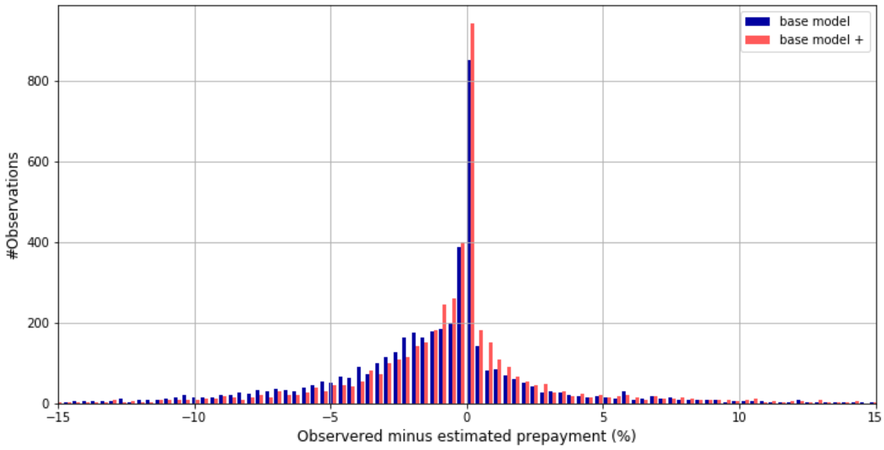

Figure 2: Histogram of errors since 2010 of out-of-sample analysis. *base model* includes NPV and *pool factor* and *base model +* includes many other variables. Source: Nordea

2.2 No silver bullets - at least until mortgage institution share more data

In our search for the silver bullet that explains why estimates deviate from realized prepayments from quarter to quarter, we conclude that we can find a combination of variables that on average makes the estimate significantly less biased. There is however still some noise left. We believe that mortgage issuers have the information that will make it possible to reduce this noise significantly, such as information on campaigns and customer contacts regarding prepayment opportunity, development in margin fees on different loan products etc. This information is not public available, but we can only encourage the mortgage institutions to share this data with the market, in the favour of bond investors whom would get more predictable returns.

3 Systematic usage of preliminary prepayments

Luckily there is another other way to detect whether the upcoming prepayments will be lower or higher than the model prediction. This is done by using the weekly information in the preliminary prepayments. In this section we show how this information can improve prediction of prepayments.

3.1 Building a model for preliminary prepayments

In figure 3 we show the relationship between preliminary prepayments and final prepayments depending on the number of days to notification date. We notice that

- The ratio between preliminary prepayments and the final prepayment is a non-linear, monotonically increasing function of the time to the notification date. There is a lot noise around the average, meaning that a model only using this information will result in estimates with some noise.
- The difference in the development per quarter. A few weeks prior to term, the ratio is highest for the October terms where the notification date is in the end of the summer holiday (people prepay before the summer holiday it seems). The ratio is lowest for

the April terms where borrowers apparently wait till after Christmas before they start to consider prepaying their loans, while it then accelerates faster than the July term.

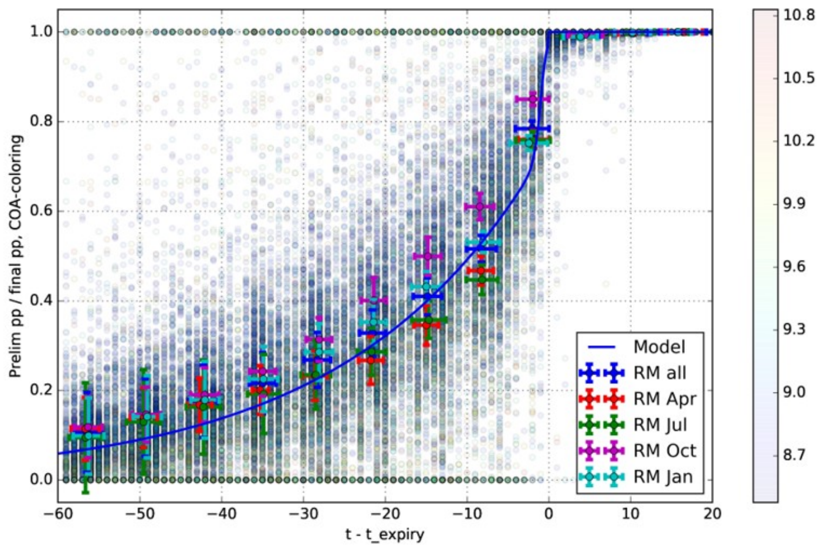

Figure 3: Relation between final prepayments and preliminary prepayments. Colored crosses marks standard deviation each quarters. Source: Nordea

To utilize this information we build an add-on to the prepayment model where we gradually increase the importance of the preliminary prepayments as we approach the notification date. The output of this model on out-of-sample data can be seen in figure 4.
We find that from three to two weeks before notification date the usage of preliminary prepayment really improves the prepayment estimate.

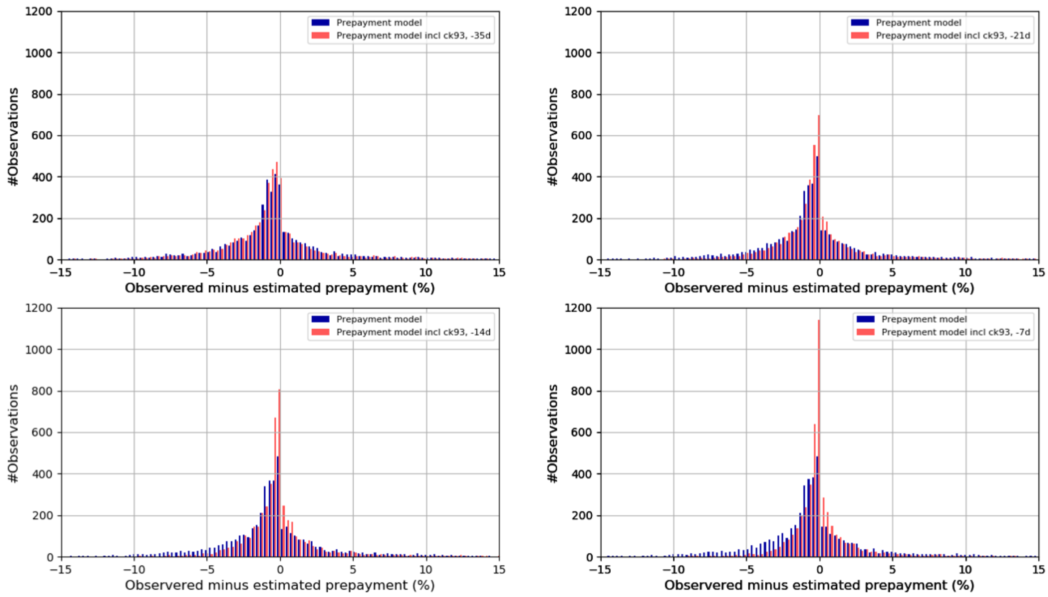

Figure 4: Histogram of out-of-sample analysis using the prepayment model with and without the usage of preliminary prepayments. The four charts shows the estimation 35, 21, 14 and 7 days before the notification date. Source: Nordea

3.2 Future callable bond model will use preliminary prepayments for upcoming term

We have shown above that the usage of preliminary prepayments can improve the estimation of prepayments for the upcoming term. As a consequence of this our expectation is to include this in the callable bond model update that will be released in 2019.

4 Random forest model with preliminary prepayments and many variables removes bias

One of the strengths of the random forest technique is that it can find subtle relationships between many variables. So one could hope that when combining the findings in section 2.1 with preliminary prepayments we would get an improved model.
Before doing that we show how the random forest is better at predicting compared to a classic prepayment model with preliminary prepayment, especially some time ahead of the notification date (figure 5).

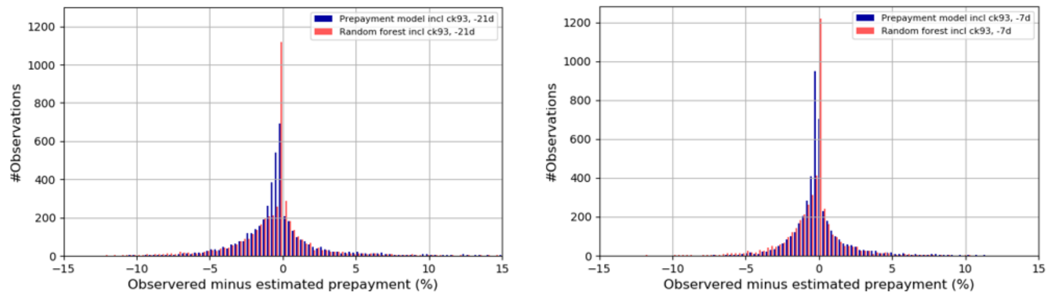

Figure 5: Histogram of out-of-sample analysis using the prepayment model and random forest using of preliminary prepayments 21 and 7 days before notification date. Source: Nordea

With regards to the overall bias when including preliminary prepayments as far as 45 days ahead of the notification date we find that it is reduced to half (figure 6). This means that random forest is a very good method for using preliminary prepayments long time ahead of the notification date.
By combining the many variables described in section 2.1 and using preliminary prepayments as far 45 days before notification date we can completely remove the bias observed over the last 9 years as well as reducing the noise in the individual quarters (figure 7). This shows the power of machine learning for these kind of problems where the combination of many variables, that otherwise would be close to impossible to model, improves estimates significantly.

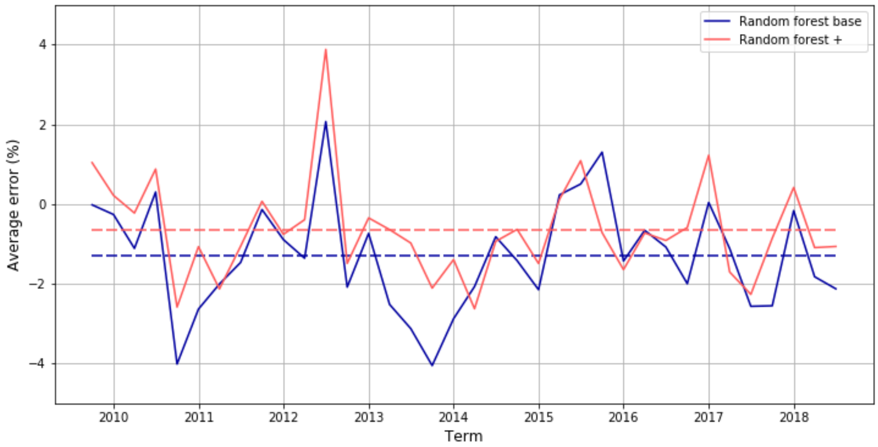

Figure 6: Average error per term using random forest using *NPV* and *pool factor* without preliminary prepayments (*Random forest base*) and with preliminary prepayments 45 days before notification date (*Random forest base +*). Source: Nordea

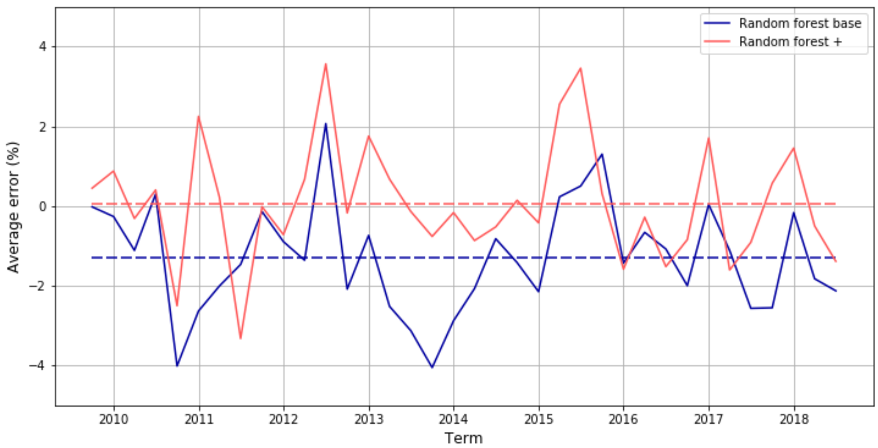

Figure 7: Average error per term using random forest without preliminary prepayments and with preliminary prepayments 45 days before notification date and with the other variables described in section 2.1. Source: Nordea

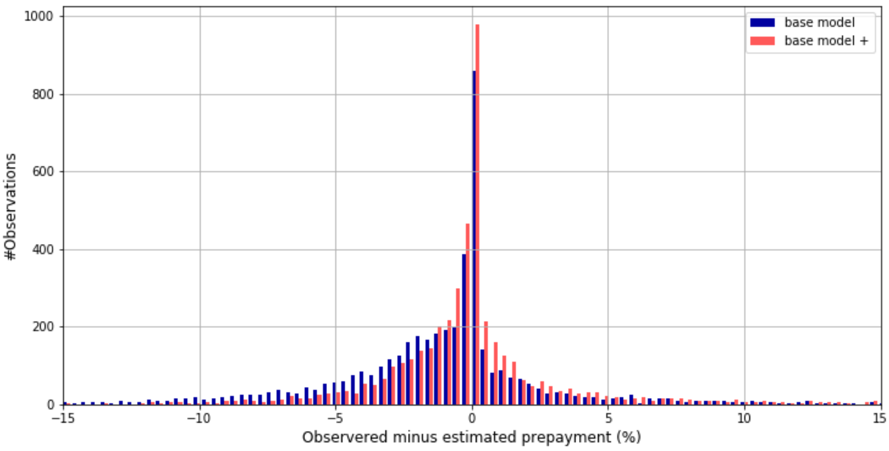

Figure 8: Histogram of out-of-sample analysis using random forest using NPV and *pool factor* (*Random forest base*) and on also including preliminary prepayments 45 days before notification date and with the other variables described in section 2.1. Source: Nordea