

Drechsel, Dirk; Neuwirth, Stefan

**Working Paper**

## Taming volatile high frequency data with long lag structure: An optimal filtering approach for forecasting

KOF Working Papers, No. 407

**Provided in Cooperation with:**

KOF Swiss Economic Institute, ETH Zurich

*Suggested Citation:* Drechsel, Dirk; Neuwirth, Stefan (2016) : Taming volatile high frequency data with long lag structure: An optimal filtering approach for forecasting, KOF Working Papers, No. 407, ETH Zurich, KOF Swiss Economic Institute, Zurich, <https://doi.org/10.3929/ethz-a-010667032>

This Version is available at:

<https://hdl.handle.net/10419/148971>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**KOF** Swiss Economic Institute

# Taming volatile high frequency data with long lag structure: An optimal filtering approach for forecasting

Dirk Drechsel and Stefan Neuwirth

KOF Working Papers, No. 407, June 2016

# KOF

ETH Zurich  
KOF Swiss Economic Institute  
LEE G 116  
Leonhardstrasse 21  
8092 Zurich, Switzerland

Phone +41 44 632 42 39  
Fax +41 44 632 12 18  
[www.kof.ethz.ch](http://www.kof.ethz.ch)  
[kof@kof.ethz.ch](mailto:kof@kof.ethz.ch)

# Taming volatile high frequency data with long lag structure: An optimal filtering approach for forecasting

Dirk Drechsel, Stefan Neuwirth\*

June 14, 2016

Preliminary Version

## Abstract

We propose a Bayesian optimal filtering setup for improving out-of-sample forecasting performance when using volatile high frequency data with long lag structure for forecasting low-frequency data. We test this setup by using real-time Swiss construction investment and construction permit data. We compare our approach to different filtering techniques and show that our proposed filter outperforms various commonly used filtering techniques in terms of extracting the more relevant signal of the indicator series for forecasting.

*JEL classifications:*

*Keywords:* Forecasting, construction, Switzerland, Bayesian, mixed data frequencies

---

\*We are grateful to Klaus Abberger, Yngve Abrahamsen, Alexander Rathke, Samad Sarferaz and Jan-Egbert Sturm for helpful comments and discussions. We further thank Florian Eckert and Salomon Faure for valuable research assistance. Author's address: KOF Swiss Economic Institute, ETH Zurich, Leonhardstrasse 21 LEE, CH-8092 Zurich, Switzerland. Corresponding authors' e-mail: neuwirth@kof.ethz.ch.

# 1 Introduction

Policy makers need up-to-date information on the state of the economy in order to implement policy actions. Often publication lags complicate this, calling for an inclusion of readily available up-to-date high frequency data to be incorporated into forecasting models in order to produce better nowcasts. Bridge models and (restricted and unrestricted) Mixed Data Sampling (MIDAS) models tackle this issue by combining high-frequency and low-frequency data for estimation and forecasting.

In this application we investigate how restricted and unrestricted MIDAS setups and bridge equations perform if the required lag length of the high frequency variable is long and the data very volatile. This can pose problems for the U-MIDAS as many parameters will have to be estimated. But also the restricted MIDAS approach could have problems to form a proper weighting scheme for very volatile data. This also holds for bridge equations, where the volatile structure of the data will likely introduce noise into the autoregressive forecasts of the high-frequency variable. To circumvent these problems we propose the use of a flexible Bayesian Beta filter for bridge equations. While still using a standard bridge equation for forecasting the low-frequency variable, the aggregation and forecast of the high-frequency variable is enhanced with our Bayesian beta filter which is able to filter a usable signal from the data. We test this compared to unfiltered data as well as smoothed data by using established filtering techniques.

Particularly, we investigate the forecast of construction investments (low frequency) by construction permits (high frequency) in Switzerland. In order to utilize the series of permits as a predictor for low frequency investments we employ a Bayesian beta filter. Using real-time data with vintages from January 2005 - December 2014 and data starting in Q1 1993 for construction investment and January 1993 for construction permits respectively, we calculate out-of-sample forecasts. We show that our filtering approach for bridge models clearly beats a U-MIDAS, a standard restricted MIDAS setup as well as traditional bridge models and competing autoregressive (AR) models in terms of out-of-sample performance both when using unfiltered data as well as when using standard smoothing approaches.

## Related literature

The problem of how to incorporate data with different frequency sampling in econometric models has been addressed in the literature in the last decade.

A large number of studies have been published looking at the benefits of employing both high and low frequency data simultaneously in the context of single-equation approaches. One of those are bridge equations, which have been used for quite some time and are common in policy organizations due to their simple method and transparency. The general idea behind bridge equations is to explain a low-frequency variable by time-aggregated low-frequency lags of a high frequency variable. First, forecasts of the high frequency variable are generated by using an additional model, normally an autoregressive process, which are then time-aggregated to the lower frequency. The estimation of both equations can be easily done by ordinary least squares (OLS). Forecasts are then done iteratively by using the previously obtained forecasts. Early applications of bridge equations in the literature can be found for example in Ingenito & Trehan (1996) or Baffigi et al. (2004) as well at central banks like ECB (2008) or Bundesbank (2013).

Another single equation approach to handle time series with different frequencies that is also able to address the problem that arises when accounting a long lag structure is mixed data sampling (MIDAS) proposed by Ghysels et al. (2004), building on Almon (1965). In this approach the high-frequency variable is not time-aggregated but directly related to the low-frequency variable. As this can lead to a high number of parameters to be estimated lag polynomials can be used to decrease the necessary number of parameters and then be estimated by non-linear least squares (Ghysels et al. (2007)). Early applications of this method were mostly with financial data where sampling differences are quite big when using daily data, for example in Ghysels et al. (2006). More recently MIDAS has also been used on macroeconomic data for example in Clements & Galvão (2008) and Clements & Galvão (2009) or Armesto et al. (2010) and Andreou et al. (2011). More recently Foroni et al. (2012) have shown, that if differences in frequencies are small, for instance for a mixture of quarterly and monthly data, an unrestricted MIDAS setup (U-MIDAS) is equivalent or even superior compared with standard MIDAS setups. An unrestricted MIDAS setup requires less computational and modelling efforts compared with standard MIDAS setups.

Schumacher (2014) proposed an iterative MIDAS (MIDAS-IT) as a combination between those two approaches. It differs from bridge equations by using the MIDAS weighting scheme on the right-hand side of the equation instead of a time-aggregated high frequency variable. The high frequency variable has still to be forecasted by using a separate model. The difference to the MIDAS approach is the iterative forecasting method instead of a direct forecast. In his application he finds no systematic advantage of any

of the three methods proposed methods.

A not finally clarified question when handling volatile high-frequency data in mixed frequency models is if MIDAS models or bridge equations could benefit from the inclusion of filtering techniques to tackle volatile high-frequency data. Established methods for dealing with volatile data are for instance a one-sided simple moving average or the Hodrick-Prescott-Filter (Hodrick & Prescott (1997)). The decomposition of the data into trend and cyclical components is one of the workhorse filters in economics, but especially for forecasting it leads to some problems due to being a two-sided filter (Baxter & King (1999)), but some workarounds have been suggested in the literature, mainly extending the current edge of the data with forecasts (European Commission (1995)). This filtered data can be used in the standard mixed-frequency models as a benchmark for comparison with more advanced filtering techniques (in this application a Bayesian beta filtering approach).

## 2 Data

Construction permits issued by municipalities offer information about upcoming construction activity as well as its volume. Since most construction activity in Switzerland requires a permit, these permissions can be utilized to forecast Swiss construction investments.<sup>1</sup>

In our analysis we use quarterly nominal, non seasonal adjusted construction investment as low frequency series and monthly construction permits as high frequency series. For the data on construction investment volumes the sample ranges from Q1 1993 to Q4 2014 and is taken from the Swiss State Secretariat for Economic Affairs (SECO).<sup>2</sup> The data on construction

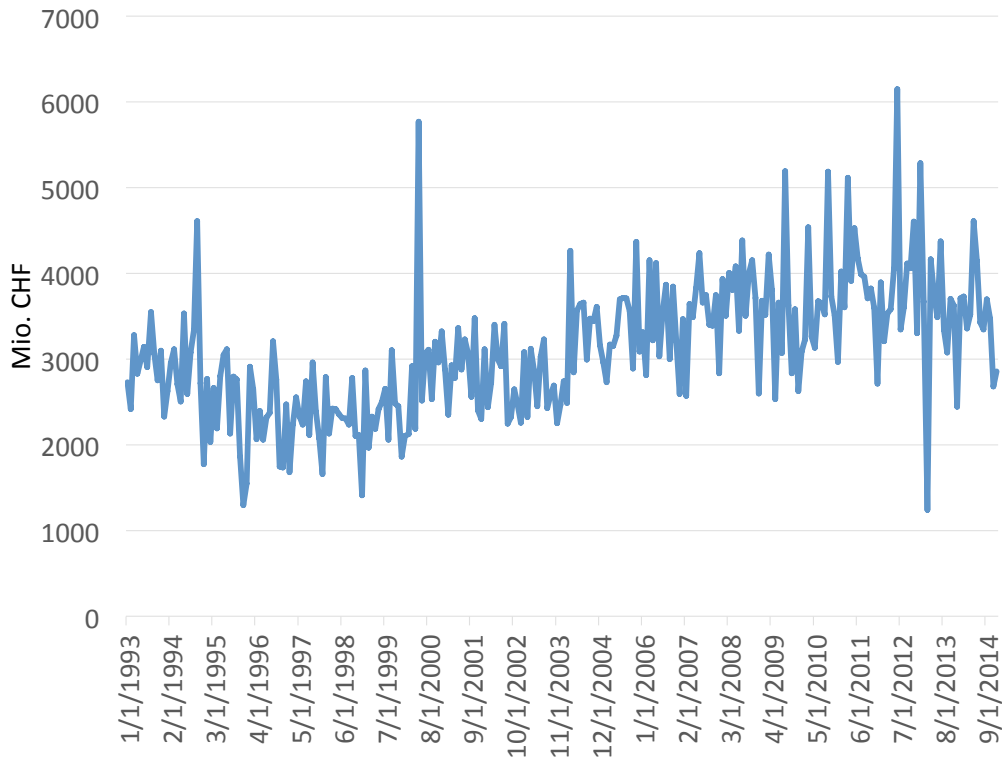
---

<sup>1</sup>Construction permit data have for instance been studied by McDonald & McMillen (2000) or Somerville (2001) for the US or by Lerbs (2012) for Germany. For Switzerland Peters & Wapf (2007) have investigated the timing of construction applications, permits, and housing starts. On average it takes 3.3 months for applications to be permitted by authorities.

<sup>2</sup>Given the share of construction investments relative to GDP in several countries forecasts of construction investment are of significant importance for decision makers analyzing the development of the aggregate economy. In contrast to other industries (e.g. trade, etc.) construction activities involve a substantial amount of local work in the production process. This substantial amount is value-adding and influences GDP strongly. The import share is lower than in other sectors. Changes in construction activity thereby are of importance for domestic business cycles and investment figures. Roundabout 9% of GDP can be attributed to construction in Switzerland. Fluctuations in construction thereby strongly influence overall investments and thereby GDP.

permits ranges from January 1993 to December 2014 and is available from Dokumedia Baublatt ([www.doku.ch](http://www.doku.ch)). Each permit contains information on the expected construction volume measured in Swiss Francs. The series used in our analysis is the sum of nominal construction permits approved by municipal authorities within Switzerland in the respective month. As can be seen in Figure 1 the data on construction permits is highly volatile with many spikes. These spikes are caused by large construction permits projects which drive up the permit series in a month by several hundred million CHF and drop out the next month again.

Figure 1: **Construction permits (in nominal Mio. CHF)**

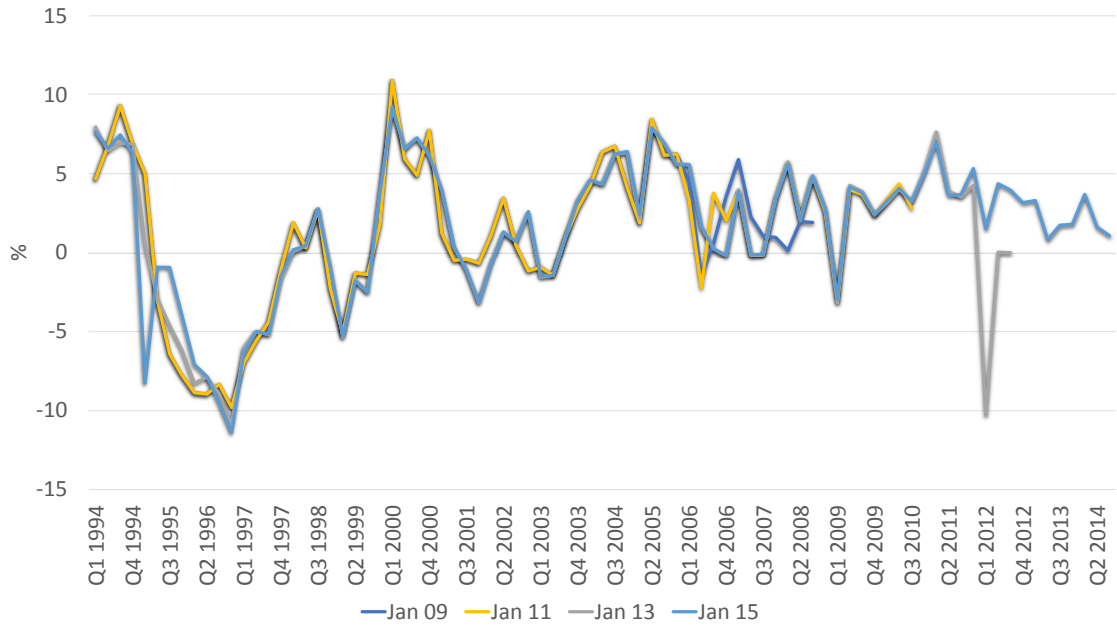


The data for construction investment in Switzerland is revised sometimes considerably, as can be seen in Figure 2. Thus, we conduct a real-time forecasting exercise for testing the forecast performance of our model. For this we construct a real-time data set for vintages starting in January 2005 until December 2014 taking both into account the state of revisions as well as the publication lag of both time series. Construction permits are normally available within one month while the publication lag for construction investment



is 3 months.<sup>3</sup> In our first vintage for January 2005 we would thus be able to use the construction permits up to December 2004 as well as construction investment until the third quarter 2004.

Figure 2: **Construction investments, year-on-year growth rates at different vintages**



### 3 A beta filter approach for forecasting

#### 3.1 Formal model description

In order to use a volatile high frequency variable such as construction permits to forecast a low frequency variable such as construction investments a couple of latent variables have to be estimated: The way how a single permit is mapped into a construction investment is unknown. Volumes will not

<sup>3</sup>Quarterly investment data is published by SECO at the beginning of March, June, September, December. Construction permission data is published by Documedia Baublatt on the 1st every month.

be spent entirely at the initiation of the construction project - the volumes will be spread over an extended time frame as it takes several months and up to many years to finish a construction project. Furthermore, the relationship between independent variable (permit volumes) and dependent variable (construction investment) is unknown because a) permit holders are not required to realize their permits and to build and b) most, but not all construction activity require a permit. Therefore construction permits are only a rough indicator of later realized construction investments.

The combination of unknown distributions and parameters poses problems. The estimation of distributions becomes computationally feasible if one is willing to select distributions out of the set of known statistical distributions. Such distributions are characterized by their moments which can be chosen in such a fashion that the overall fit of model and data is optimized. By using the beta distribution, the estimation strategy reduces itself to retrieve these moments and is made feasible by reducing the parameter space. This is represented by the weighting equation:

$$\tilde{X}_{HFt} = \sum_{k=0}^{n-1} \mathcal{B}(k/n; p, q) \tilde{L}_{HF}^k X_{HFt} \quad (1)$$

where  $X_{HFt}$  is a monthly high-frequency variable,  $\tilde{L}_{HF}^k$  a lag operator,  $\tilde{X}_{HFt}$  the redistributed high-frequency variable and  $n$  the number of desired lags.  $\mathcal{B}$  represents a beta distribution, depending on the number of discrete density steps  $k$  and the shape parameters  $p, q$ . The Beta distribution has the convenient property that values cannot be negative (which would be implausible in terms of construction volumes) and can be written as follows:

$$\mathcal{B}(k/n; p, q) = \frac{1}{B\left(\frac{k+1}{n}, \frac{k}{n}, p, q\right)} \left(x - \frac{k+1}{n}\right)^{p-1} \left(\frac{k}{n} - x\right)^{q-1} \quad (2)$$

Outside of the interval  $x \in [0, 1]$  the function values of  $f(x)$  are set to zero. Shape parameters  $p, q$  are strictly positive  $> 0$ . The term  $B\left(\frac{k+1}{n}, \frac{k}{n}, p, q\right)$  indicates a restricted Beta function with the upper and lower boundaries  $\frac{k+1}{n}$  and  $\frac{k}{n}$

$$B\left(\frac{k+1}{n}, \frac{k}{n}, p, q\right) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \left(\frac{k}{n} - \frac{k+1}{n}\right)^{p+q-1} \quad (3)$$

where  $\Gamma$  represents a Gamma function:

$$\Gamma(z) = \int_0^\infty u^{z-1} e^{-u} du \quad (4)$$

The redistributed high-frequency variable  $\tilde{X}_{HF_t}$  is then time-aggregated to the lower frequency and can then be plugged into the standard bridge equation as  $\tilde{X}_{HF_t}^{LF}$ :

$$\Delta Y_t = c + \sum_{i=1}^p \beta_i L^i \Delta Y_t + \gamma \Delta \tilde{X}_{HF_t}^{LF} + \epsilon_t. \quad (5)$$

$\Delta Y_t$  are observed quarterly year-over-year construction investment growth rates,  $c$  is a constant,  $p$  is the desired lag length of the autoregressive term with the coefficients  $\beta_i$  to be estimated,  $L^i$  is a lag operator,  $\gamma$  is the coefficient associated to the redistributed construction permit series and  $\tilde{X}_{HF_t}^{LF}$  are the time-aggregated quarterly redistributed construction permits at time  $t$  which are used as year-on-year growth rates, indicated by  $\Delta$ .

### 3.2 Estimation strategy

By utilizing a Gibbs sampler the bridge equation parameters  $\psi = (c, \beta_1, \dots, \beta_p, \gamma)$  are drawn from a Normal-Gamma distribution given the Beta distribution shape parameters  $p, q$ . Proposals for  $p, q$  are drawn in a Metropolis-within-Gibbs sampler step.<sup>4</sup> Random walk candidate draws for  $\theta = (p, q)$  are generated by

$$\theta^* = \theta^{(s-1)} + z \quad (6)$$

with  $s$  being the number of the current draw,  $s-1$  indicating the previous accepted draw, and  $\theta^*$  the new draw. Innovations  $z$  are determined by drawing

---

<sup>4</sup>A Bayesian Metropolis-within-Gibbs sampler setup for an estimation of a MIDAS regression has for instance been employed by Ghysels (2012), Ghysels & Owyang (2011) and Rodriguez & Puggioni (2010). In contrast to Rodriguez & Puggioni (2010), we do not impose prior values on the lag structure of beta distributions. In contrast to Ghysels (2012), we employ a Random-Walk Metropolis Hastings step, while they draw from candidate densities utilizing importance sampling. While Rodriguez & Puggioni (2010) employ Bayesian model selection to identify the model fitting the data best based on marginal likelihoods, we consider the joint distribution of prior and likelihood over all models visited by the algorithm to find the expected mean parameters values fitting the data.

random numbers from a Normal density with mean  $\theta^{(s-1)}$  and covariance-matrix  $\Sigma_\theta$ , which is chosen to yield acceptance ratios between 0.2 and 0.6 for each shape parameter. The shape parameters  $p, q$  are drawn independently of each other, but are jointly accepted or discarded in each iteration. Draws will be accepted based on their acceptance probability:

$$\alpha(\theta^{(s-1)}, \theta^*) = \min \left[ \frac{p(\theta = \theta^* | y, \psi)}{p(\theta = \theta^{(s-1)} | y, \psi)}, 1 \right] \quad (7)$$

Thus if the posterior of the new draw given the data  $y$  and the parameters  $\psi$  is higher than the posterior of the previous draw the new draw will be accepted with probability 1. If the posterior of the new draw is lower than the posterior of the previous draw the new draw will only be kept if the ratio of the new posterior and the posterior evaluated at the previous step is higher than a random number drawn from a uniform distribution between 0 and 1. I.e. even if the posterior declines and the new draw has a lower probability it will be kept if it is not too unlikely. This ensures, that the Metropolis-Hastings algorithm walks over the entire parameter space even into regions with lower probability while taking more draws in regions with high probability, i.e. where the posterior is highest.

### 3.3 Forecasting strategy

Once the beta shape parameters have been drawn, the coefficients of the bridge equation regression can be drawn. For each set of coefficients a forecast  $\hat{\mathbf{Y}}_s$  is calculated. The median forecast for each forecasting horizon is the result of

$$\hat{\mathbf{Y}} = \text{median}_{(s) \in S} \hat{\mathbf{Y}}_s(\theta^{(s)}, \psi^{(s)}) \quad (8)$$

where  $S$  is the number of draws.<sup>5</sup> Forecasts will be generated for a nowcast and up to four quarters ahead. A nowcast for the unknown  $y_t$  (due to the publication lag) is estimated by using the drawn parameters and relying on observed investment values  $y_{t-1}, \dots, y_{t-4}$  and redistributed construction permits for  $t$ .<sup>6</sup> The forecasts are done iteratively, so the forecast for  $\hat{\mathbf{Y}}_{t+1}$

---

<sup>5</sup>In our setting 200,000 draws minus 160,000 burn in, divided by 10 to adjust for autocorrelation in the draws

<sup>6</sup>In January, February, April, May, July, August, October, November the previous quarter investment values have not been published yet. Therefore also a backcast for the previous quarter has to be estimated.

uses the drawn parameters with both the nowcast  $\hat{\mathbf{Y}}_t$  as well as the past observed values  $y_{t-1}$ ,  $y_{t-2}$ ,  $y_{t-3}$  and the redistributed construction permits for period  $t + 1$ :

$$\Delta Y_{t+1} = c + \sum_{i=1}^p \beta_i L^i \Delta Y_{t+1} + \gamma \Delta \tilde{X}_{HF_{t+1}}^{LF}. \quad (9)$$

In contrast to bridge models or the MIDAS-IT approach without leads by Schumacher (2014), which generate high frequency variable forecasts by employing a simple AR model, the forecasts by the beta filtering approach for the high frequency variable  $\tilde{X}_{HF_t}$  are done by rescaling the truncated density function to correct for missing probability mass and re-weighting the remaining high-frequency observations:

$$\tilde{X}_{HF_{t+h}} = \frac{1}{\sum_{k=h}^{n-1} \mathcal{B}(k/n; p, q)} \sum_{k=h}^{n-1} (\mathcal{B}(k/n; p, q) \tilde{L}_{HF}^k X_{HF_t}). \quad (10)$$

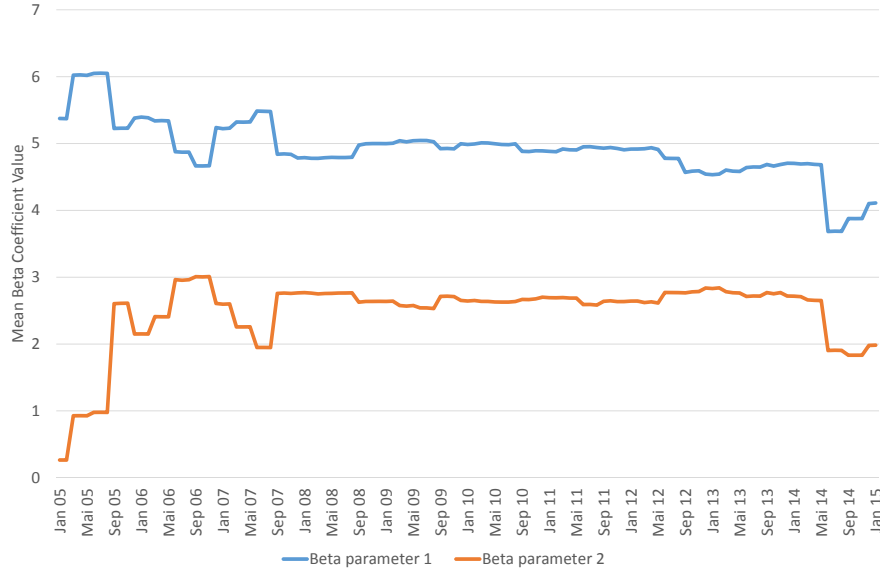
## 4 Real-time parameter estimates

In order to check if the estimation results are stable over time we analyse the distribution and estimation parameters of the real-time estimation. The mean coefficients for the shape parameters  $p, q$  over all vintages can be seen in figure 3. The shape parameters were mostly stable over all vintages between 4.5 and 5 for the first parameter  $p$  and between 2.5 and 2.8 for the second parameter  $q$ .

In June 2014 the beta shape parameters dropped considerably, probably to a very high first estimate of construction investment for the first quarter 2014 of 12.4% year-on-year growth which was shortly afterwards revised down to 3.7% during an extensive revision of the whole history of construction investment. This led to a shift of the distribution mean to the right, i.e. extending the time until construction permits are actually effective for construction investment.

The mean shape coefficients can be displayed as Beta distribution for example with  $p = 5$  and  $q = 2.7$  in figure 4. The estimated beta distribution gives little weight to the first observations but rises rapidly thereafter, reaching a climax after five years and quickly reducing its weight structure thereafter. Such a weight structure might be the consequence of the long time it takes for construction to start after the actual permit was issued. Furthermore, the lag structure of large projects is very long, implying a

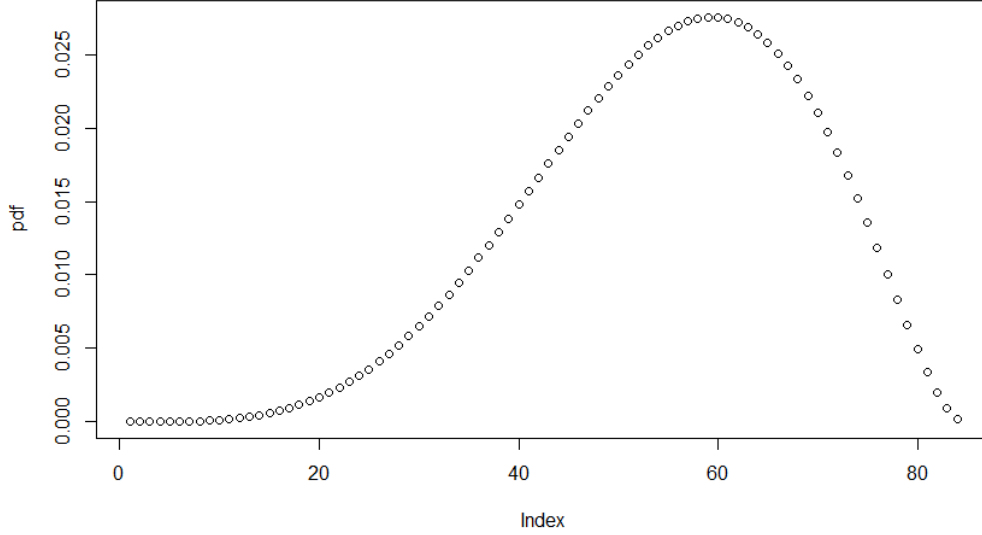
Figure 3: Real-time coefficients of beta distribution



longer duration of construction projects.

Using the beta-filtered real-time series to depict the estimated coefficients for the bridge equation shows a varying picture over the real-time vintages (see figure 5). This figure shows that the constant as well as the coefficient for the redistributed permit data are more or less stable over all vintages. The autoregressive terms of the bridge equation exhibit more variation, which could be attributed to mayor revisions or surprising releases of construction investments. Before June 2012 the coefficients for Lag 3 and Lag 4 are basically not statistically distinguishable from zero. The story changes for vintages until September 2014. In June 2012 the first release of negative construction investment growth rates was published for Q1/2012 (a drop by more than -10%). Until September 2014 negative growth rates for construction investments were published by the statistical office (see also figure 2). In October 2014 the construction investment growth rate for Q1/2012 was revised to +1.5% - an upward revision by more than 11 percentage points compared to the first estimate which now indicates more a boom than a slump. This could also explain the slight decrease in the coefficient for the redistributed

Figure 4: **Beta distribution**



permits and the sudden increase of the coefficient in October 2014.

## 5 Out-of-sample tests

### 5.1 Benchmark model setup

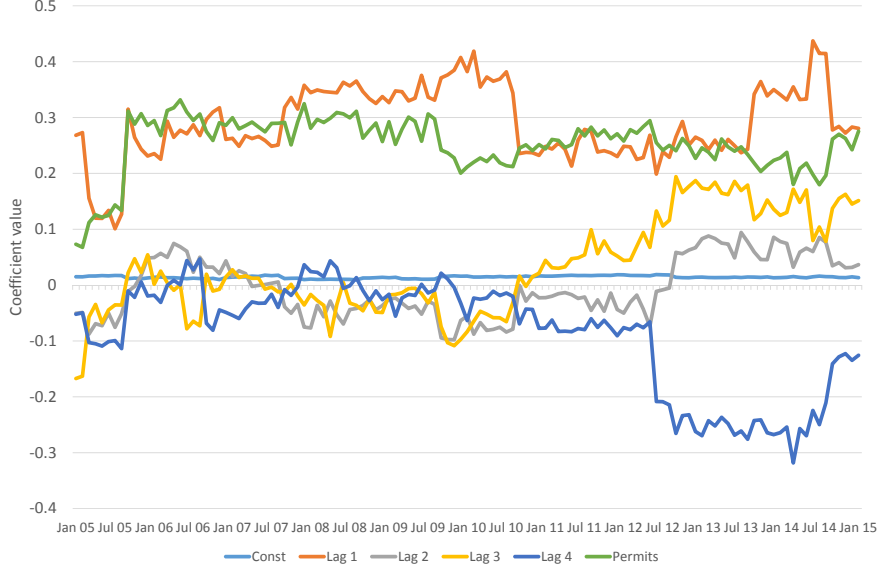
We employ four different benchmark models for the forecast comparison with the bridged MIDAS model. Namely we use a simple autoregressive model, a bridge model, a MIDAS regression with an Almon lag polynomial for parameter reduction and an unrestricted MIDAS regression to compute forecasts, which will be covered in more detail in the following sub-chapters.

#### 5.1.1 Autoregressive model

We use a simple autoregressive model as benchmark model for the construction investment. An autoregressive model is specified using the formula

$$Y_t = c + \sum_{i=1}^p \beta_i L^i Y_t + \epsilon_t,$$

Figure 5: **Real-time mean coefficients of the bridge equation**



where  $L^i$  indicates a lag function ranging from 1 to the desired lag length  $p$  and  $\beta_i$  the parameter for each  $i$ th lag of variable  $Y_t$ .  $\epsilon_t$  is an error term. Forecasts of  $\hat{Y}_{t+1}$ ,  $\hat{Y}_{t+2}, \dots, \hat{Y}_{t+h}$  are done iteratively by plugging in the previously forecast values, i.e  $\hat{Y}_{t+2} = c + \beta_1 \hat{Y}_{t+1} + \beta_2 Y_t + \dots + \beta_p Y_{t-p+2}$ .

### 5.1.2 Bridge model

Additionally we also use a simple bridge model, which enables us to enhance the autoregressive model with monthly information. This approach is separated into three steps. Firstly, we specify a model using the own lags of the quarterly variable  $Y_t$  as well as the time aggregated quarterly values of the monthly variable  $X_t^q$ . This is defined as follows:

$$Y_t = c + \sum_{i=1}^p \beta_i L^i Y_t + \sum_{j=0}^n \gamma_j L^j X_t^q + \epsilon_t$$

In a second step, using only a monthly variable and an autoregressive model, forecasts of the monthly variables are computed, as specified in the previous subsection. The monthly variable and the respective forecasts



are time-aggregated to quarterly frequency and plugged into the above specification in order to compute forecasts of the quarterly variable. The forecasts are again computed iteratively.

By aggregating the high-frequency variable to the lower frequency potential information is lost because time averaging assumes that each high-frequency observation  $X_t$  inside a low frequency period receives the same weight. It could be possible that for instance the first observation of  $X_t$  should get a higher weight in the time aggregation than the others. Thus, potential information is lost due to this process.

### 5.1.3 MIDAS

Ghysels et al. (2004) introduced the MIDAS approach. This approach circumvents the problem of time averaging by approximating the parameters of each high-frequency observation of the high frequency variable  $X_t$  with a polynomial function  $\theta(k; \omega)$ . As benchmark model we use a MIDAS model with a non-exponential Almon lag polynomial, which is quite flexible but at the same time still maintains parsimony. Additionally the non-linear weighting function can be transformed back into linear form, which allows the model to be easily estimated by using OLS. The approximation of the high-frequency observations is done for  $K = n * m$  lags, which depend both on the number of high frequency periods inside the low frequency period  $m$  as well as the number of low frequency period lags  $n$ . For instance, if  $X_t$  were a monthly variable and  $Y_t$  a quarterly variable, then  $m = 3$ . The model is specified with respect to the desired forecast horizon  $h$ . The aim is to find the specification that would have predicted  $Y_t$   $h$  quarters before. Thus the MIDAS equation for growth  $Y_{t+h}$  in period  $t + h$  would be defined as:

$$Y_{t+h} = c + \sum_{i=0}^p \beta_i L^i Y_t + \gamma \sum_{k=0}^K \theta(k; \omega) L^{k/3} X_{t+l} + \epsilon_{t+h}$$

$L^{k/3}$  is a lag operator for monthly variables which is defined as  $X_{t-1/3} = L^{1/3} X_t$  and  $l$  is defined as the lead of the high-frequency variable on the low-frequency variable. Thus, the forecasts are done directly by plugging the most recent data into the formula using the estimated parameters for each forecast horizon  $h$ .

#### 5.1.4 U-MIDAS

The unrestricted MIDAS approach was promoted by Forni et al. (2012). Instead of approximating the parameters of each high-frequency observation of the high frequency variable  $X_t$ , this approach estimates the weights as unrestricted parameters. This allows for even more flexibility than the MIDAS approach. But in contrast U-MIDAS approach does not maintain parsimony. Depending on the frequency of the high-frequency variable, this approach can easily be over-parametrized. But when frequency differences are small as with macroeconomic data the results of the U-MIDAS are similar or even slightly superior to the MIDAS approach. The functional form of the U-MIDAS is defined as:

$$Y_{t+h} = c + \sum_{i=0}^p \beta_i L^i Y_t + \sum_{k=0}^K \gamma_k L^{k/3} X_{t+l} + \epsilon_{t+h}$$

## 5.2 Benchmark results

Our aim is to forecast the quarterly year-on-year growth rate of nominal construction investments up to one year ahead using the construction permits and investment data that were available at the moment of the forecast. In order to test the forecast performance of our model we conduct a real-time experiment using Vintages from January 2005 until December 2014. For each of the vintages we conduct a nowcast for the quarter of the vintage as well as one to four step-ahead forecasts.

In order to evaluate the forecast performance of our model we take several steps. First we compare those forecasts with a simple autoregressive model, which uses the same amount of information as the bridge model with our redistributed data, namely 4 quarterly lags. In a second step we produce forecasts for Swiss construction investment using Swiss building permit data as additional information, but in its original form. For this we use different benchmark models, namely a bridge model, a standard MIDAS setup and an unrestricted MIDAS (U-MIDAS) as explained in section 5.1. All models have the same quarterly information as our model, but as the bridge model and especially the U-MIDAS tend to become over-parametrized quite fast, we also restrict the lagged information of the high frequency variable to one year. The Almon-MIDAS can use more lagged information due to the reduction of the parameter space by using an Almon lag polynomial. Thus

we use the same lag length as in our baseline model (84 monthly lags). In a third step we use a Hodrick-Prescott filter for smoothing the high-frequency data, both in its simple form and with an adjustment for endpoint problems. In a final step we compare the forecast performance of our setup with a moving average for the high-frequency data.

To compare the forecast performance of the different models we calculate the relative root mean squared forecast error (RMSFE) for each model and forecast horizon  $h$  which is defined as:

$$\Delta RMSFE_h = 100 * \left( \frac{RMSFE_h^{Beta\ filtered} - RMSFE_h^{Benchmark}}{RMSFE_h^{Benchmark}} \right).$$

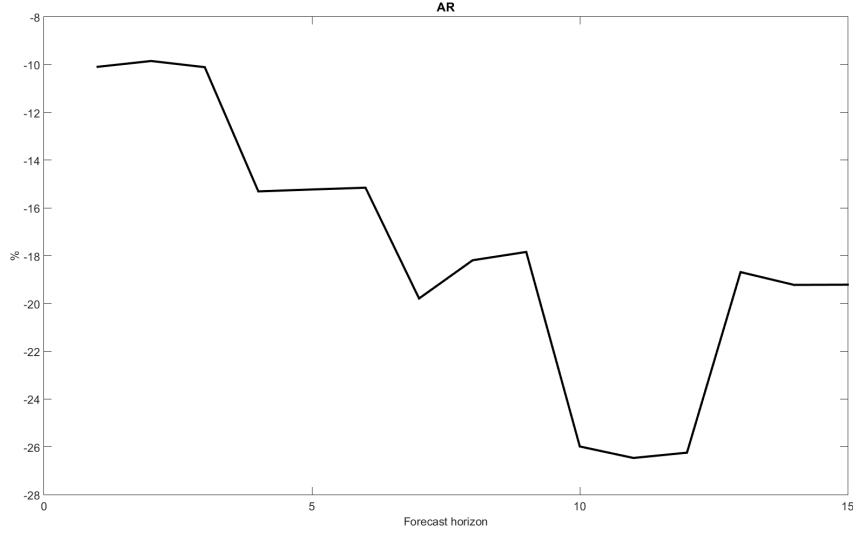
We refer to  $\Delta RMSFE_h$  as the relative change in the RMSFE. The *more negative*  $\Delta RMSFE_h$  is, the *better* performs our setup relative to the respective benchmark model in terms of predictive power.

The following results are robust for different lag specifications. When using more monthly lags for the bridge model and the U-MIDAS the models will be over-parametrized and their forecast performance thus decrease further. Fewer monthly lags do not improve the forecast performance significantly. The results are also robust when using 72 or 96 monthly lags for beta-filtered data, moving average and Almon MIDAS. Additionally the results don't change significantly when using different growth rates for the high frequency variable, i.e. month-on-month growth instead of year-on-year growth.

### 5.2.1 Autoregressive process

In order to test if the inclusion of the high-frequency building permit data actually delivers any additional information we first test the forecast performance of a bridge model including the beta-filtered data compared to a simple autoregressive process with 4 lags (AR(4)). As can be seen in figure 6 the filtered data improves upon the AR(4) model for the whole forecast horizon. The relative RMSFE actually decreases for a longer forecast horizon, which means the forecast errors of the model including the beta-filtered data are even smaller compared to the AR(4) for forecasts up to one year ahead.

Figure 6: **Relative RMSFE to autoregressive model**

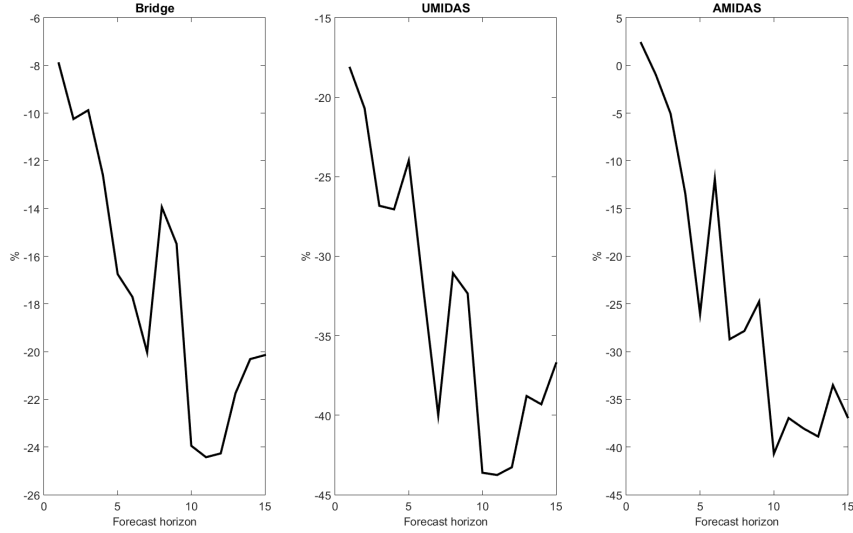


### 5.2.2 Unfiltered Data

In this section we test if the filtering of the data using our method leads to an improved forecast performance compared to using the original data. For this we look at three standard mixed-frequency forecasting models. As indicated above bridge models and the unrestricted MIDAS are not able to cope with long lags which is why we restrict the lags for those models to 4 quarterly and 12 monthly lags. The Almon MIDAS in contrast uses the same amount of information as the bridge model with the filtered data, i.e. 84 monthly lags. As can be seen in figure 7 the addition of the original permit data seems to give at least some additional information compared to AR(4) model in figure 6. The relative RMSFE of the bridge model, which includes the permit data, aggregated to quarterly frequency, yields a slightly lower forecast error than the AR(4) model, i.e. the relative RMSFE is slightly better when both are compared to the bridge model using the beta-filtered data. In contrast, both the unrestricted MIDAS model as well as the Almon MIDAS model do not seem to be able to get a good signal from the original data. While the relative RMSFE of the UMIDAS model compared to the bridge model using the beta-filtered data is constantly below zero, the relative RMSFE of the Almon MIDAS model is at least for the nowcast horizon close to zero but drops rapidly afterwards. In conclusion, the addition of the original permit data can lead to small increases in forecast performance as long as the model is able to get a useful

signal from the data.

Figure 7: **Rel. RMSFE to benchmark models with unfiltered data**



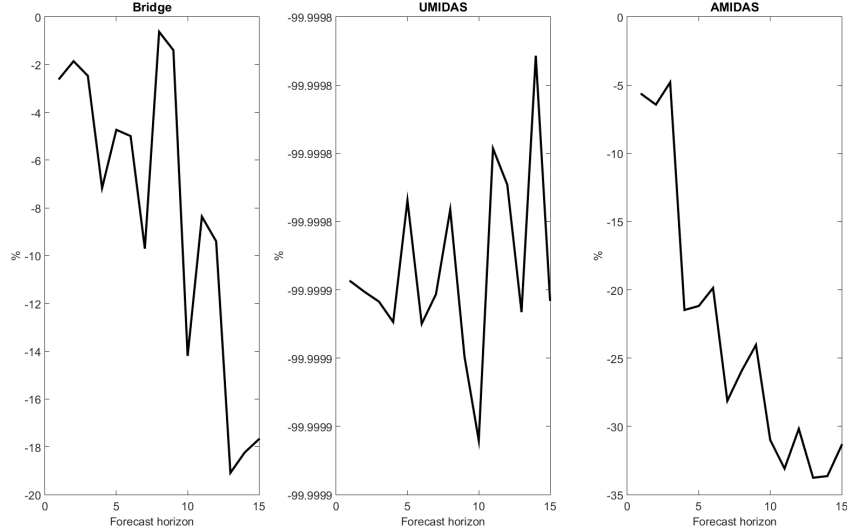
### 5.2.3 HP-filtered Data

A method to extract a useful signal from volatile data is filtering the data. One of the most used filters is the Hodrick-Prescott Filter (Hodrick & Prescott (1997)). This allows a decomposition of the data into a trend and a cyclical component. For the use of the HP-filter with monthly data we use a  $\lambda$  of 129,600 as suggested by Ravn & Uhlig (2002). We use the trend component as high frequency variable for forecasting.

As can be seen in figure 8 the forecast performance of the bridge model and the Almon MIDAS improve slightly when compared to the usage of the unfiltered data in the previous subsection. The relative RMSFE of those two benchmark models is nevertheless still clearly below zero and thus performs worse than the bridge model using the beta-filtered data. The UMIDAS seems to have severe problems dealing with the HP-filtered data. This is due to very high collinearity due to the low variation in the trend component of the filtered data between monthly observations.

The HP-filter has well known endpoint problems ((Baxter & King (1999))). This stems from the fact that the smoothed series at the beginning and the

Figure 8: **Rel. RMSFE to benchmark models with HP-filtered data**



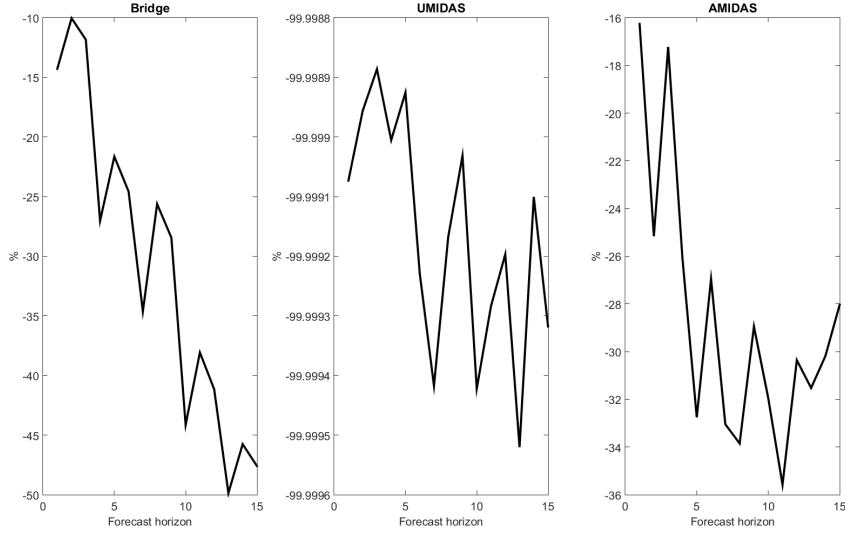
end of the time series tends to be close to the observed data. A general workaround is to use forecasts for several observations ahead in order to produce a better smoothing at the current edge of the data. In our case the mediocre results of the HP-filtered data could be related to this problem. Thus, we use a simple autoregressive process to forecast the high frequency data in order to get more reliable smoothing results. The outcome can be seen in figure 9. The forecast performance actually worsens when compared to the HP-filter without correction for endpoint problems. The relative RMSFE is more negative for bridge models and the Almon MIDAS. It seems that autoregressive forecasts when using very volatile data is not advantageous.

#### 5.2.4 Moving average

Another standard method for smoothing volatile data is to use a moving average of the data, which is closest to our method. But instead of having a separate weight for each high frequency observation a moving average gives all past observations the same weight. The basic analysis uses seven years or 84 months of past data, thus each monthly observation would get a weight of  $1/84$ . The forecast is done in the same manner as with the beta filter by re-weighted the remaining observations.

The result can be seen in figure 10 which shows the relative RMSFE of a bridge model using the moving average data and the bridge model using

Figure 9: **Rel. RMSFE to benchmark models with adjusted HP-filtered data**



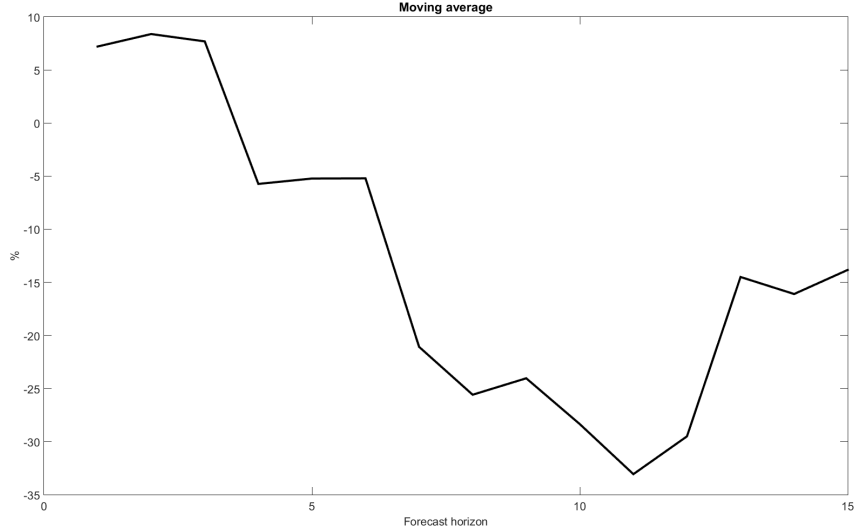
the beta filtered data. When nowcasting using moving average data does seem to give slightly smaller forecast errors, for longer forecast horizons the model using the beta-filtered data clearly outperforms the moving-average data. It seems that the weighting scheme of the beta-filtering enables us to make better forecasts by correctly re-weighting the past observations.

## 6 Conclusion

To study the usefulness of a Bayesian beta filtering approach for a setup with a long lag structure of a volatile high frequency variable to forecast a low frequency we apply beta filtering approach on construction permit and investment data. Quarterly estimates of construction activity are subject to a substantial publication lag, thus, timely available construction permits can give policy makers an early indication on the state of the construction sector.

As construction permits can contain information for several years ahead and display a volatile structure, they can pose a problem for traditional mixed-frequency approaches even when using standard filtering methods. To deal with the special structure of the data we construct a Bayesian beta filtering setup which allows us to use both long lags and is at the same time

Figure 10: **Relative RMSFE to moving average**



able to filter a usable signal from the data.

To test the forecast performance of our model we conduct a real-time experiment using vintages from January 2005 until December 2014. We compare the out-of-sample forecast performance of our model with different benchmark models, namely a simple AR-process, a bridge model, an unrestricted MIDAS and a restricted MIDAS using an Almon polynomial for the reduction of the parameter space. The beta filtering approach clearly improves the forecast accuracy upon an AR-model from 10% for shorter forecast horizons to more than 25% for longer horizons. While the inclusion of permit data in standard mixed-frequency models leads to small improvements in the forecast performance compared to the AR-model, they are still clearly outperformed by our beta filtering approach, especially for longer forecasting horizons. This result also holds when pre-filtering the data with a HP-filter, although the forecast performance of the benchmark models slightly increases. Also the application of a one-sided moving average yields a weaker forecast performance, especially for longer forecast horizons. The specific structure of construction permit data seems to contain mostly noise for traditional models while the Bayesian beta filter can still use the contained information.



## References

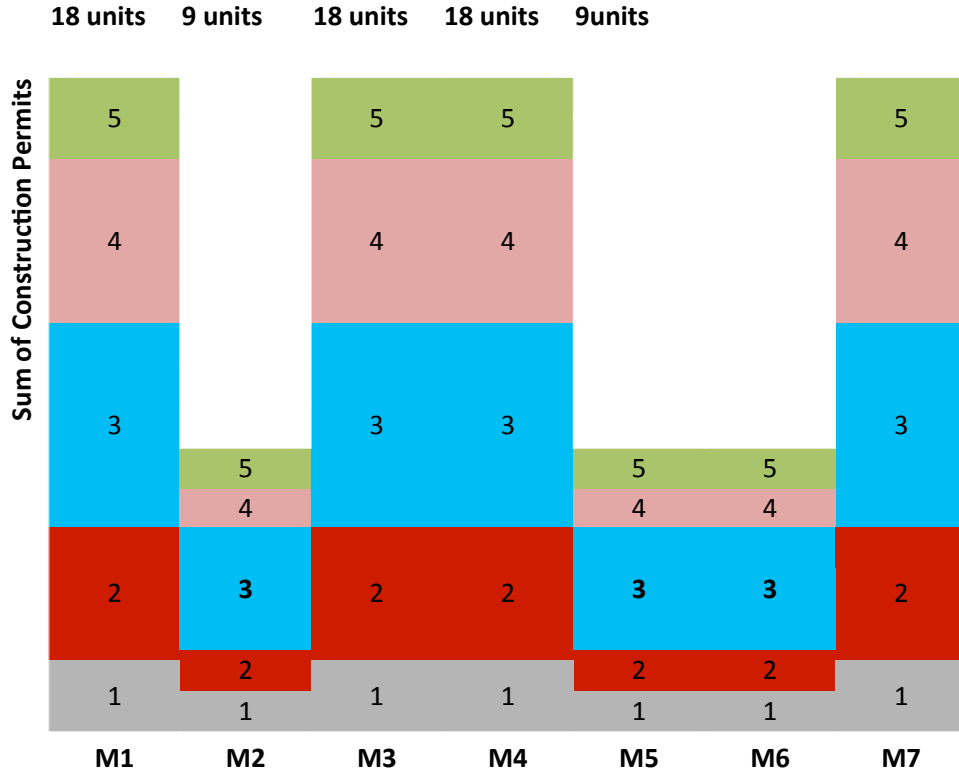
- Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, (pp. 178–196).
- Andreou, E., Ghysels, E., & Kourtellis, A. (2011). Forecasting with mixed-frequency data. In *In Oxford Handbook of Economic Forecasting*, (pp. 225–245).
- Armesto, M. T., Engemann, K. M., & Owyang, M. T. (2010). Forecasting with Mixed Frequencies. *Federal Reserve Bank of St. Louis Review*, 92(6), 521–536.
- Baffigi, A., Golinelli, R., & Parigi, G. (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting*, 20(3), 447–460.
- Baxter, M., & King, R. G. (1999). Measuring Business Cycles: Approximate Band-Pass Filters For Economic Time Series. *The Review of Economics and Statistics*, 81(4), 575–593.
- Bundesbank (2013). Forecasting models in short-term business cycle analysis - a workshop report. *Deutsche Bundesbank Monthly Report September 2013*, (pp. 69–83).
- Clements, M. P., & Galvão, A. B. (2008). Macroeconomic Forecasting With Mixed-Frequency Data. *Journal of Business & Economic Statistics*, 26, 546–554.
- Clements, M. P., & Galvão, A. B. (2009). Forecasting US output growth using leading indicators: an appraisal using MIDAS models. *Journal of Applied Econometrics*, 24(7), 1187–1206.
- ECB (2008). Short-term forecasts of economic activity in the euro area. *ECB Monthly Bulletin April 2008*, (pp. 69–74).
- European Commission (1995). The commission services method for cyclical adjustment of government budget balances. *DG II*, 401.
- Foroni, C., Marcellino, M., & Schumacher, C. (2012). U-MIDAS: MIDAS regressions with unrestricted lag polynomials. CEPR Discussion Papers 8828.
- Ghysels, E. (2012). Macroeconomics and the Reality of Mixed Frequency Data. *University of North Carolina Working Paper*, (pp. 1–52).

- Ghysels, E., & Owyang, M. T. (2011). Taylor Monetary Policy Rules with Financial Market Expectations. *Federal Reserve Bank of Saint-Louis Discussion Paper*.
- Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). The MIDAS Touch: Mixed Data Sampling Regression Models. *CIRANO Working Papers*, 20, 1–33.
- Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1-2), 59–95.
- Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, 26(1), 1–48.
- Hodrick, R., & Prescott, E. C. (1997). Postwar u.s. business cycles: An empirical investigation". *Journal of Money, Credit, and Banking*, 29, 1–16.
- Ingenito, R., & Trehan, B. (1996). Using monthly data to predict quarterly output. *Economic Review*, (pp. 3–11).
- Lerbs, O. (2012). House Prices, Housing Development Costs, and the Supply of New Single-Family Housing in German Counties and Cities. In *19th Annual European Real Estate Society Conference*, (pp. 1–31).
- McDonald, J. F., & McMillen, D. P. (2000). Residential Building Permits in Urban Counties: 1990–1997. *Journal of Housing Economics*, 9(3), 175–186.
- Peters, M., & Wapf, B. (2007). Befragung von Schweizer Baugesuchstellern. *Die Volkswirtschaft*, 2007(05).
- Ravn, M. O., & Uhlig, H. (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *The Review of Economics and Statistics*, 84(2), 371–375.
- Rodriguez, A., & Puggioni, G. (2010). Mixed Frequency Models: Bayesian Approaches to Estimation and Prediction. *International Journal of Forecasting*, 26(2), 293–311.
- Schumacher, C. (2014). MIDAS and bridge equations. Deutsche Bundesbank Discussion Papers 26/2014.
- Somerville, C. T. (2001). Permits, Starts, and Completions: Structural Relationships Versus Real Options. *Real Estate Economics*, 29(1), 161–190.

## Appendix 1: Intuition of the bridged MIDAS approach

Consider an arbitrary set of construction permits over time, which have been summed up for each month  $M$  (see figure 11).

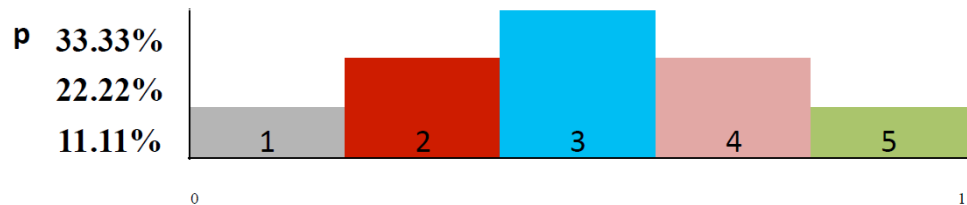
Figure 11: Example construction permits



These observed construction permit sums are in a next step multiplied with an estimated distribution. The example distribution in figure 12 peaks in the center, has a total duration of five months, i.e. all construction permits are being used for construction investments within five months, and it is bell-shaped, i.e. little construction activity takes place right after the permit issue and at the end of the five month period.

The result of the multiplication is a series of effective construction permits, meaning that based on the estimated distribution, the observed construction permits for each month are redistributed according to the volumes which will actually be used for construction in a specific month. So for instance, as can be seen in figure 13 only 11% of the permit volume will be used for construction in the first month (M1). The next month redistributed values

Figure 12: **Example distribution**



(M2) contain the sum of 11% of (observed) M2 permit numbers and 22% of (observed) M1 permit numbers. Accordingly the fifth month (M5) of the redistributed series is the sum of 11% of (observed) M5 permit numbers and 22% of (observed) M4 permit numbers, 33% of (observed) M3 permit numbers, 22% of (observed) M2 permit numbers and 11% of (observed) M1 permit numbers.

Estimation takes place only when the full information is available (99% of the distribution). Referring to the example, only redistributed series starting from (redistributed) M5 would be used for the estimation. The forecasts for the redistributed series lack the information for the current month. Therefore the values will be rescaled to 100%.

Figure 13: **Example redistributed construction permits**

