Forecasting inflation rates using daily data: 
A nonparametric MIDAS approach

Jörg Breitung*  Christoph Roling
University of Cologne  University of Bonn
18th December 2014

Abstract
In this paper a nonparametric approach for estimating mixed-frequency fore-
cast equations is proposed. In contrast to the popular MIDAS approach that 
employs an (exponential) Almon or Beta lag distribution, we adopt a penal-
ized least-squares estimator that imposes some degree of smoothness to the 
lag distribution. This estimator is related to nonparametric estimation proce-
dures based on cubic splines and resembles the popular Hodrick-Prescott fil-
tering technique for estimating a smooth trend function. Monte Carlo exper-
iments suggest that the nonparametric estimator may provide more reliable 
and flexible approximations to the actual lag distribution than the conven-
tional parametric MIDAS approach based on exponential lag polynomials. 
Parametric and nonparametric methods are applied to assess the predictive 
power of various daily indicators for forecasting monthly inflation rates. It 
turns out that the commodity price index is a useful predictor for inflations 
rates 20–30 days ahead with a hump-shaped lag distribution.

Key words
Forecasting, mixed frequency, nonparametric regression

JEL classification
C12, C32

*Corresponding author: Institute of Econometrics, University of Cologne, Albert-Magnus-Platz, 50923 Bonn, Germany. E-mail address: breitung@statistik.uni-koeln.de.
1 Introduction

Due to increased accessibility of high-frequency data it is becoming more and more attractive to employ variables observed at different sampling frequencies for economic forecasting (see Ghysels et al. (2006), Monteforte and Moretti (2008), Andreou et al. (2011), Modugno (2013), and Banbura et al. (2013) for a comprehensive review of the literature). In such exercises various high-frequency observations (30 days, say) of the predictor are associated with a single low-frequency observation (one month). This raises the problem of how to exploit the potentially large number of intra-period observations to forecast the low frequency process. The simplest approach just averages the corresponding high-frequency observations to obtain an aggregated predictor for the dependent low-frequency variable. Accordingly, the intra-period observations are equally weighted when forming the predictor. A more appealing approach is to adapt a more flexible weighting scheme. For example, when using daily financial data to forecast monthly variables like inflation rates we may want to assign larger weights to more recent observations. Using least-squares methods for estimating the weighting scheme may however yield unreliable estimates due to a large number of estimated parameters and possible multicollinearity. The MIDAS approach suggested by Ghysels et al. (2006) is a convenient and elegant parametric approach to overcome these problems.

An important drawback of such a parametric approach is that the shape of the lag distribution is governed by a (more or less) arbitrary class of functions such as exponential polynomials or the beta function. In empirical applications, parsimonious specifications of parametric functions may fail to yield an accurate approximation of the lag distribution. For example, using exponential polynomials or the beta function implies that the regressors affect the dependent variable in the same direction, ruling out the case that the predictive regressor affects the target variable positively for some range of lags, whereas the effect becomes negative at some other lags. This happens, for example if the long-run effect is smaller than the short-run effect. Such scenarios are ruled out in the standard parametric MIDAS framework.

It is therefore desirable to employ nonparametric techniques that circumvent the choice of a particular class of parametric functions when fitting the lag distribution to high-frequency data. In this paper a nonparametric approach is suggested that merely imposes some degree of smoothness to the lag distribution.
This approach is similar to fitting cubic splines for approximating an unknown functional form. Related approaches are popular in many areas of applied statistics. For example, the widely used Hodrick-Prescott filter for an unknown trend (e.g. Hodrick and Prescott (1997)) or the flexible least-squares approach (e.g. Kalaba and Tesfatsion (1989) and Lütkepohl (1993)) is based on similar principles.

In Section 2 we sketch the parametric MIDAS approach proposed by Ghysels et al. (2006) and motivate our nonparametric framework. More details of the suggested smoothed least-squares (SLS) estimator are considered in Section 3 and some first results on the relative performance of the nonparametric estimator are presented in Section 4. In the empirical application of Section 5 we investigate the predictive power of daily time series (such as commodity prices, stock prices and interest rates) for monthly inflation rates in Germany. It turns out that the daily commodity price index (or daily oil prices in Rotterdam) of the previous month explains more than 40 percent of the variance of monthly changes in the German inflation rates. Section 6 concludes.

2 Mixed frequency estimation

Consider the regression model combining the low-frequency variable $y_t$ and the high-frequency variable $x_{t,j}$, where $t = 1, \ldots, T$ is the low-frequency time index (monthly, say) and $j$ is the intra-period high-frequency (daily) observation with $j = 1, \ldots, n_t$ ($n_t \in \{28, \ldots, 31\}$ in our example). For convenience we assume that the time index $s$ runs in the opposite direction, that is, the pair $(t, n_t)$ represents the first and $(t, 0)$ the final observation of the month $t$. The low and high frequency variables are linked by a linear regression of the form

$$y_{t+h} = \alpha_0 + \sum_{j=0}^{p} \beta_j x_{t,j} + u_{t+h},$$

where $u_{t+h}$ is uncorrelated with $x_{t}, \ldots, x_{t-p}$. To simplify the exposition, we confine ourselves to a single predictor variable $x_{t,j}$ so that $\beta_j$ is a scalar. Furthermore, the lag-length $p$ is assumed to be smaller than the minimum number of intra-period observations (i.e. $p < \min(n_t) = 28$ in our example with daily observations). The extension to higher lag-orders is straightforward but involves an additional notational burden. In the MIDAS approach of Ghysels et al. (2007) and Andreou et al. (2011)) the coefficients are written as $\beta_j = \alpha_1 \omega_j(\theta)$, where the
weights are generated by an exponential polynomial

$$\omega_j(\theta) = \frac{\exp (\theta_1 j + \cdots + \theta_k j^k)}{\sum_{i=0}^{p} \exp (\theta_1 i + \cdots + \theta_k i^k)}.$$

The vector of $k$ hyper-parameters $\theta = (\theta_1, \ldots, \theta_k)'$ is unknown. Alternatively, the Beta distribution may be used (e.g. Ghysels et al. (2007)). Obviously, $w_j(\theta) \in [0, 1]$, and $\sum_{j=0}^{p} w_j(\theta) = 1$. With this specification, the model becomes

$$y_{t+h} = \alpha_0 + \alpha_1 \sum_{j=0}^{p} w_j(\theta) x_{t,j} + u_{t+h}, \quad (1)$$

and the parameters $\alpha_0, \alpha_1,$ and $\theta_1, \ldots, \theta_k$ can be estimated by non-linear least squares (NLS). In many empirical applications a second order exponential Almon lag with $k = 2$ is sufficient to yield an accurate approximation of the weight function (cf. Ghysels et al. (2007)). Essentially, MIDAS regression models extend the aggregation with equal-weights to more general temporal aggregation schemes that appear to be better suited to many empirical applications, in particular for the purpose of forecasting volatility (see Ghysels et al. (2006)) or macroeconomic forecasting with high-frequency predictors (see Andreou et al. (2013) or Foroni et al. (2014)).

The MIDAS approach may also be motivated as a restricted OLS regression where the “reduced form parameters” in the OLS regression

$$y_{t+h} = \alpha_0 + \beta' x_{t,\bullet} + u_{t+h}, \quad (2)$$

with $\beta = (\beta_0, \ldots, \beta_p)'$ and $x_{t,\bullet} = (x_{t,0}, \ldots, x_{t,p})'$ are restricted by the nonlinear function $\beta = g(\alpha_1, \theta)$ with the $j$'th element $\beta_j = \alpha_1 \omega_j(\theta)$. Accordingly, the parameters are estimated by minimizing the criterion function

$$S(\alpha, \theta) = \sum_{t=1}^{T} [y_{t+h} - \alpha_0 - x_{t,\bullet}' g(\alpha_1, \theta)]^2.$$

The crucial feature of the function $g(\alpha, \theta)$ is that it represents the high-dimensional vector $\beta$ by a smooth parametric function depending on a low-dimensional vector of parameters $\theta$. In this paper we propose an alternative nonparametric approach that does not impose a particular functional form but merely assumes that the coefficient $\beta_j$ is a smooth function of $j$ in the sense that the absolute values of the
second differences

\[ \nabla^2 \beta_j = \beta_j - 2\beta_{j-1} + \beta_{j-2} \quad \text{for } j = 2, \ldots, p, \]

are small. Specifically, the coefficients \( \beta_0, \ldots, \beta_p \) are obtained by minimizing the penalized least-squares objective function

\[ \tilde{S}(a_0, \beta) = \sum_{t=1}^{T} \left( y_{t+h} - a_0 - \beta' x_{t\bullet} \right)^2 + \lambda \sum_{j=2}^{p} (\nabla^2 \beta_j)^2, \quad (3) \]

where \( \lambda \) is a pre-specified smoothing parameter. This objective function provides a trade-off between goodness of fit (which is maximized by using the unrestricted least-squares estimator) and an additional term that penalizes large fluctuations of the coefficients \( \beta_0, \ldots, \beta_p \). To some extent this approach resembles the well known Hodrick-Prescott (1997) filter for extracting a smooth trend. An important difference is, however, that the dimension of the parameter space is typically smaller than the number of observations, whereas the parameter space underlying the Hodrick-Prescott filter (represented by the number of terms in the penalty function) equals the number of observations. Our approach is also related to other nonparametric approaches such as cubic splines (e.g. Härdle (1990)) that involve an objective function based on the sum of squared residuals and a quadratic penalty function.

In the following section some properties of the estimator \( \tilde{\beta}_{\lambda} \) that results from minimizing the penalized least-squares function are analyzed and data-dependent methods for choosing the smoothing parameter are considered.

### 3 A nonparametric MIDAS approach

#### 3.1 The SLS Estimator

Consider the smoothed least-squares (SLS) estimator resulting from

\[
\tilde{\beta}_{\lambda} = \arg\min_{\beta} \left\{ (y^h - X\beta)'(y^h - X\beta) + \lambda \beta' D'D\beta \right\} \\
= (X'X + \lambda D'D)^{-1} X'y^h, \quad (4)
\]
where

\[
D = \begin{bmatrix}
1 & -2 & 1 & 0 & \cdots & 0 \\
0 & 1 & -2 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots \\
0 & \cdots & \cdots & \cdots & \cdots & 1 & -2 & 1
\end{bmatrix} \quad (p - 1) \times (p + 1),
\]

\[y^h = [y_{1+h}, \ldots, y_{T+h}]', \quad X = [x_{1,\bullet}, \ldots, x_{T,\bullet}]', \quad \text{and} \quad x_{t,\bullet} = [x_{t,0}, \ldots, x_{t,p}]'.\]

For simplicity we ignore the constant \(a_0\) in the regression.

This estimator admits different interpretations. First, it may be rewritten as a shrinkage (ridge) estimator. Define the \((p - 1) \times 1\) vector of second differences \(\gamma_2 = D\beta\). Furthermore, let \(D_0\) denote a \(2 \times (p + 1)\) matrix such that \(\tilde{D} = [D_0', D']'\) is invertible. For concreteness, let \(D_0 = [I_2, 0_{2 \times (p-1)}]\) such that \(\gamma = \tilde{D}\beta = [\gamma_1', \gamma_2']'\) with \(\gamma_1 = [\beta_0, \beta_1]'\). Minimizing (3) is equivalent to minimizing the objective function

\[
\tilde{S}(\gamma) = (y^h - \tilde{X}_1\gamma_1 - \tilde{X}_2\gamma_2)'(y^h - \tilde{X}_1\gamma_1 - \tilde{X}_2\gamma_2) + \lambda\gamma_2'\gamma_2,
\]

with respect to \(\gamma = [\gamma_1', \gamma_2']' = \tilde{D}\beta\), where \(\tilde{X} = X\tilde{D}^{-1} = [\tilde{X}_1, \tilde{X}_2]\). This reformulation of the minimization problem shows that the SLS is related to the shrinkage (ridge) estimator since the estimator of \(\gamma_2\) resulting from a minimization of (5) takes the form

\[
\tilde{\gamma}_2 = \left(\tilde{X}_2^*\tilde{X}_2^* + \lambda I_{p-1}\right)^{-1}\tilde{X}_2^*y^h,
\]

where \(\tilde{X}_2^* = M_1\tilde{X}_2\) and \(M_1 = I_T - \tilde{X}_1(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'\). Note that the unrestricted OLS estimator results from setting \(\lambda = 0\), whereas for \(\lambda \to \infty\) the coefficients \(\beta_0, \ldots, \beta_p\) lie on a straight line.

A second interpretation of the estimator emerges from assuming stochastic coefficients with \(\nabla^2 \beta_i \overset{iid}{\sim} \mathcal{N}(0, \eta^2)\). In this case \(\lambda = \sigma_n^2 / \eta^2\) and the objective function is equivalent to the log-likelihood function. In a similar manner a Bayesian interpretation is available (with a normal prior distribution for \(\nabla^2 \beta_i\)). Finally we note that the SLS estimator can be conveniently computed as an OLS estimator with \([y', 0_{p-2}']\) as the vector of dependent variables and \([X', \sqrt{\lambda}D']'\) as regressor matrix.

Another interesting interpretation is obtained from rewriting the first order
condition as

$$\tilde{\beta}_\lambda = \left[ I_{p+1} + \overline{\lambda}_T \left( \frac{1}{T} X'X \right)^{-1} D'D \right]^{-1} \hat{\beta}, \quad (7)$$

where $\hat{\beta} = (X'X)^{-1}X'y^h$ denotes the OLS estimator and $\overline{\lambda}_T = \lambda / T$. This relationship between the unrestricted and the smoothed estimator shows that $\tilde{\beta}_\lambda$ is obtained as a weighted average of all coefficients $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$. Assuming $T^{-1}X'X \overset{p}{\rightarrow} S_X$ it follows that as $\overline{\lambda}_T = \lambda / T \rightarrow 0$, the SLS estimator converges to the OLS estimator in the sense that $\sqrt{T}(\tilde{\beta}_\lambda - \hat{\beta}) = o_p(1)$.

### 3.2 Choice of the smoothing parameter

When deriving the SLS estimator we have assumed that the smoothing parameter $\lambda$ is given. For empirical applications, however, some guidance is needed for choosing this parameter. One possible approach is to adapt some cross-validation method. To this end the sample is partitioned into $k$ blocks with $\text{int}(T/k)$ observations. For a given value of the smoothing parameter $\lambda$, we discard the $j$-th block of observations, $j = 1, \ldots, k$, and obtain $\hat{\beta}_\lambda$ by using the observations in the remaining $k-1$ blocks. Let $\tilde{\beta}_\lambda^{(-j)}$ denote the SLS estimator by leaving out the $j$'th block. The CV criterion is obtained by computing the squared prediction error for all $k$ blocks. By evaluation the CV criterion for a grid of values from the interval $\lambda \in [\lambda_\ell, \lambda_u]$ the MSE optimal value of the smoothing parameter can be found. In practice, however, the CV procedure is computational demanding as the interval $[\lambda_\ell, \lambda_u]$ is typically very large.

A computationally simpler approach is to select the parameter $\lambda$ based on some information criterion. Note that the fitted values of the SLS method result as

$$\hat{y}^h(\lambda) = X (X'X + \lambda D'D)^{-1} X'y^h = P_\lambda y^h,$$

with $P_\lambda = X (X'X + \lambda D'D)^{-1} X'$. For $\lambda = 0$ we have $\text{tr}(P_\lambda) = p$, whereas for $\lambda \rightarrow \infty$ it follows that $\text{tr}(P_\lambda) = 2$. Accordingly, the trace of the matrix measures the dimension of the space spanned by the estimator $\hat{\beta}_\lambda$ and the choice of $\lambda$ is equivalent to select the “dimension” of the space spanned by estimated coefficients. Although this notion is reasonable for an integer number of $\kappa_\lambda = \text{tr}(P_\lambda)$,
this interpretation is somewhat unclear in the case of fractional values of \( \kappa_\lambda \).

The notion of the pseudo-dimension \( \kappa_\lambda \) is also useful to compare the non-parametric approach to alternative parametric procedures. Choosing the degree of exponential Almon lag distributions imply discrete values of the pseudo-dimension, whereas the nonparametric approach allows us to choose any real value in the interval \([2, \infty)\), and, therefore, we expect choosing a more appropriate level of smoothness in practice.

The information criteria balances the trade-off between goodness-of-fit and the dimension reduction of the estimation procedure: as \( \lambda \) increases, the sum of squared residuals increases relative to OLS, while the (pseudo) dimensionality \( \kappa_\lambda \) shrinks down. We adapt the framework of Hurvich, Simonoff and Tsai (1998) and employ the modified Akaike criterion given by

\[
AIC(\lambda) = \log \left( \left[ y^h - \hat{y}^h(\lambda) \right]' \left[ y^h - \hat{y}^h(\lambda) \right] \right) + \frac{2(\kappa_\lambda + 1)}{T - \kappa_\lambda - 2}.
\]

It is interesting to note that for large \( T \) the modified Akaike criterion is equivalent to the original AIC criterion suggested by Akaike (1974), where \( \kappa_\lambda \) replaces the number of parameters. The smoothing parameter results from minimizing this criterion with respect to \( \lambda \). As shown in the appendix, the minimizer \( \lambda_{\text{AIC}}^* \) results from solving the first-order condition

\[
\left( y^h' X Q_{\lambda_{\text{AIC}}^*}^{-1} (D'D) Q_{\lambda_{\text{AIC}}^*}^{-1} X' \left( y^h - \hat{y}^h(\lambda_{\text{AIC}}^*) \right) \right) (\text{SSR}(\lambda_{\text{AIC}}^*))^{-1}
= (T - 1) \text{tr} \left[ (X'X) Q_{\lambda_{\text{AIC}}^*}^{-1} (D'D) Q_{\lambda_{\text{AIC}}^*}^{-1} \right] (T - \kappa_{\lambda_{\text{AIC}}^*} - 2)^{-2},
\]

where \( Q_\lambda = (X'X + \lambda D'D) \), \( \text{SSR}(\lambda) = \left[ y^h - \hat{y}^h(\lambda) \right]' \left[ y^h - \hat{y}^h(\lambda) \right] \). Accordingly, the optimal choice of \( \lambda \) can be found by determining the root of equation (8).

### 4 Estimation Performance

#### 4.1 Design of Monte Carlo simulation

In this section, we compare the small sample properties of parametric and non-parametric variants of the MIDAS regression. In order to compare our results to studies that have been conducted previously in the literature, we generate data
according to the model of Andreou et al. (2010), given by

\[ y_{t+h} = \beta_0 + \sum_{j=0}^{p} \beta_j x_{t,j} + u_{t+h}, \quad (9) \]

\[ \beta_j = \alpha_1 \omega_j(\theta), \quad (10) \]

\[ u_{t+h} \overset{\text{i.i.d.}}{\sim} N(0, 0.125), \quad (11) \]

for \( t = 1, 2, \ldots, T \), where \( \beta_0 = 0.5 \), and \( \omega_j(\cdot) \) is a weighting function for which several specifications are presented in more detail below. The high-frequency predictor is generated by the AR(1) process

\[ x_{t,j} = c + \varrho x_{t,j-1} + \varepsilon_{j,t}, \quad \varepsilon_{j,t} \overset{\text{i.i.d.}}{\sim} N(0, 1), \]

for \( j = 0, \ldots, p \), where \( x_{t,j-p-k} = x_{t-j-k} \) for all \( k > 0 \). Accordingly, \( x_{t,j} \) denotes the \( j^{th} \) lag of the AR(1) series \( x_{t,0} \). As in Andreou et al. (2010), \( c = 0.5 \) and \( \varrho = 0.9. \)

Since \( X \) is independent of \( u^h \) with \( E(u^h u^h') = \sigma^2 u I_T \), the (trace) MSE conditional on \( X \) is obtained as

\[ E[(\tilde{\beta}_\lambda - \beta)'(\tilde{\beta}_\lambda - \beta)|X] = \beta' \Psi_T(\lambda)' \Psi_T(\lambda) \beta + \sigma^2 u [I_{p+1} - \Psi_T(\lambda)] X'X[I_{p+1} - \Psi_T(\lambda)'], \quad (12) \]

where \( \Psi_T(\lambda) = \lambda (X'X + \lambda D'D)^{-1} D'D \). Minimizing the (conditional) MSE(\( \lambda \)) with respect to \( \lambda \) yields the MSE minimal value of the smoothing parameter \( \lambda_{\text{MSE}} \). In our simulation experiments the smoothing parameter is reported relative to the sample size, that is, \( \frac{\lambda_{\text{MSE}}}{T} \). Similarly, \( \frac{\lambda_{\text{AIC}}}{T} = \frac{\lambda_{\text{AIC}}}{T} \) is obtained from minimizing the modified AIC criterion.

In model (9) – (11), we choose the sample size as \( T \in \{100, 200, 400\} \) and the number of high-frequency lags as \( p+1 \in \{20, 40, 60\} \). Moreover, in (10), we choose the scaling parameter \( \alpha_1 \in \{0.2, 0.3, 0.4\} \), corresponding to models of small, medium and large signal-to-noise-ratios, respectively. Although the implied (centered) \( R^2 \) differ a bit due to different sample sizes, weighting functions and the number of high-frequency lags, these signal-to-noise ratios roughly cor-

\[ ^1 \]In additional Monte Carlo simulations (cf. Breitung et al. (2013)) we found that the performance of the nonparametric MIDAS approach improves relative to the parametric competitor as the autocorrelation of the regressor gets larger. This may be due to the implied multicollinearity of the regressors that tends to affect the parametric estimator more severely.
respond to $0.40 \leq R^2 < 0.50$, $0.50 \leq R^2 \leq 0.70$ and $R^2 > 0.70$, respectively. The number of Monte Carlo replications is 1000.

For each of the different weighting functions given below, we then consider the ratio of the median MSE of the nonparametric approach relative to median MSE of the parametric estimation procedure. Accordingly, a ratio above one implies that the parametric estimator produces more accurate estimates of the weighting function on average, whereas a ratio below one indicates that the nonparametric method is superior. All results regarding relative estimation accuracy are presented in table 1.

The different weighting functions are presented in figure 1. The weighting functions generated for this exercise are meant to cover different types of behavior of the high-frequency weights, including fast decay (with a lot of weights approximately equal to zero after some cut-off lag), as well as modest or even slow decay (see the first and third row of figure 1).

### 4.2 Results

In our first Monte Carlo experiment we consider a lag distribution that fits into the parametric MIDAS framework. The weight function is exponentially declining with

$$
\omega_j(\theta) = \frac{\exp(\theta_1 j)}{\sum_{i=0}^{p} \exp(\theta_1 i)}, \quad j = 0, \ldots, p.
$$

(13)

where we follow Andreou et al. (2010) and set $\theta_1 = 7 \cdot 10^{-4}$ and $\theta_2 = -6 \cdot 10^{-3}$. The parametric NLS estimator is computed using the exponential Almon lag with $k = 2$ parameters. The top row of figure 1 depicts the lag distributions and the top row of table 1 shows the MSE ratios.

Next, we also examine a hump-shaped pattern given by

$$
\omega_j(\theta) = \frac{\exp((\theta_1 j - \theta_2 j^2))}{\sum_{i=0}^{p} \exp((\theta_1 i - \theta_2 i^2))}.
$$

(14)

We determine the parameters such that the weighting function reaches a maximum at $j = 5$, $j = 10$, and $j = 15$ when 20, 40 and 60 lags are employed, respectively. To this end, we choose $\theta_1 = 8 \cdot 10^{-2}$ and $\theta_2 = \theta_1/10, \theta_2 = \theta_1/20,$
and \( \theta_2 = \theta_1 / 30 \), respectively. The parametric NLS estimator is computed using the exponential Almon lag with \( k = 2 \) parameters. The second row of figure 1 presents the lag distributions and the MSE ratios are given in the second row of table 1.

As another example of a monotonically declining weight function we consider the linear function:

\[
\omega_j(a_0, a_1) = \frac{a_0 + a_1 j}{a_0(p + 1) + a_1(p + 1)(p + 2)/2} \tag{15}
\]

The weights are normalized to sum up to unity. The third row in figure 1 presents the shapes of the lag distributions and the third row of table 1 reports the MSE ratios in this case.

Finally, to highlight the flexibility of the nonparametric approach, we consider a cyclical weight function in which the weights change sign.

\[
\omega_j(c_1, c_2) = \frac{c_1}{p + 1} \left[ \sin \left( c_2 + \frac{j2\pi}{p} \right) \right], \tag{16}
\]

where \( c_2 = 0.01 \), and \( c_1 = 5 \), \( c_1 = 2.5 \), and \( c_1 = 5/3 \) for the cases of 20, 40 and 60 lags, respectively, ensuring that these weights sum up to one. The bottom row of figure 1 show the implied lag distribution of this specification, while the MSE ratios are displayed in the bottom row of table 1.

The results of table 1 can be summarized as follows. First, the nonparametric approach can improve in-sample estimation accuracy substantially relative to the parametric approach. This observation applies in particular to the linear and the sign-changing weights, but to some extent also to the exponentially declining weights. For the humped-shaped weights, the parametric estimator yields more accurate estimates in most cases. Moreover, the nonparametric approach tends to perform better for low or medium signal-to-noise ratios.

Regarding the exponentially declining weights, the performance of the parametric approach improves as the sample size \( T \) increases, which is to be expected, as the parametric model for the weights is correctly specified. The same pattern applies to the humped-shaped case, in which the parametric estimator clearly outperforms the nonparametric estimator as the sample size grows. Turning to

---

\(^2\text{For the sign-changing weights, we also experimented with the 3 parameter version of the exponential Almon lag, but it did not yield an improvement in terms of MSE relative to the specification with two parameters. See also Breitung et al. (2013).}\)
the sign-changing weights, the nonparametric approach yields sizeable gains in estimation accuracy for this weighting function. This result is not too surprising as the parametric approach is misspecified and cannot produce negative weights. Similarly, the case of linear weights may be favorable for the nonparametric approach, as the SLS estimator attains a linear function as the degree of smoothing increases. The point is, however, that in a given estimation problem, the true weights may not be described by a presupposed (even though flexible) parametric family of models, and the nonparametric approach may be very beneficial in these cases.

Second, selecting the smoothing parameter by the AIC criterion leads to MSE ratios that are relatively close to the ratios implied by the MSE minimal (infeasible) value of the smoothing parameter in some cases, but may fall short of this benchmark in other cases considered here. Hence, although the AIC criterion suggests reasonable (albeit not optimal) values for the smoothing parameter, experimentation with different values for the smoothing parameters may be necessary to select the smoothing parameter in practice.

Until now we have focused on estimating the weights $\beta_j$ in (10). We now turn to evaluate the accuracy of out-of-sample forecasts made by the nonparametric and the parametric approach in this simple model. To this end, we partition the whole sample into an estimation sample, comprising observations $1, 2, \ldots, T^e$, and a forecasting sample consisting of observations $T^e + 1, \ldots, T$. In this exercise we set $T_e = T/2$.

The estimation sample is used to obtain a baseline estimate of the weights $\beta_j$ from the sample

$$\hat{y}_{t+h} = \sum_{j=0}^{p} \widehat{\beta}_j x_{t,j} + \hat{u}_{t+h}, \text{ for } t = 1, \ldots, T^e.$$ Given the baseline estimates, the one-step ahead forecast of the dependent variable is

$$\hat{y}_{T^e+1|T^e} = \sum_{j=0}^{p} \widehat{\beta}_j x_{T^e,j}.$$ We then include the next period in the estimation sample while dropping the first period, re-estimate the model and obtain the next one-step ahead forecast. Proceeding in this fashion, we obtain a series of one-step ahead forecast errors
employing a moving window of $T_e$ observations in each step.\(^3\) It is important to note that forecasts relying on the parametric and nonparametric MIDAS regressions are based on the same information set $\mathcal{I}_t = \{x_{t,0}, \ldots, x_{t,p}\}$. Thus, it is not surprising that both forecasting methods perform similarly. Let $\hat{y}_{t+h|t}^{(1)} = f_{\hat{\theta}}(\mathcal{I}_t)$ and $\hat{y}_{t+h|t}^{(2)} = f_{\hat{\lambda}}(\mathcal{I}_t)$ denote the parametric and nonparametric forecasts based on the high-frequency predictor. The decomposition of the forecast error yields

$$y_{t+h} - \hat{y}_{t+h|t}^{(j)} = E(y_{t+h} | \mathcal{I}_t) - \hat{y}_{t+h|t}^{(j)} + u_{t+h}.$$ 

In our out-of-sample forecast exercise the MSE is dominated by the forecast error, whereas the differences between $\hat{y}_{t+h|t}^{(1)}$ and $\hat{y}_{t+h|t}^{(2)}$ is typically small relative to the forecast error. Therefore it is not surprising that the MSE ratios of the two forecast methods are generally close to unity. Notable exceptions are the results for the hump-shaped weight function with $p + 1 = 60$ and sign-changing weights with $p + 1 = 20$, where the forecasts of the non-parametric MIDAS procedure clearly outperforms the parametric approach.

\section{Forecasting Monthly Inflation Rates Using Daily Indicators}

\subsection{Daily predictors}

Frequent updates of inflation forecasts are becoming more and more important for central banks and financial analysts. Such updates can be easily obtained by employing daily time series. Nowadays a variety of financial time series is available on a daily basis and their role in macroeconomic now- and forecasting is considered; see Andreou et al. (2013) and Banbura et al. (2013) for examples of recent applications. Modugno (2013) analyzes the predictive content of daily world market prices of raw materials as well as some financial indicators such as interest rates, stock price indices and exchange rates. From the set of monthly, weekly and daily indicators, she extracts leading factors by using the state-space representation of the mixed-frequency dynamic factor model.

\(^3\)Due to the recursive nature of the exercise and the multiple non-linear optimization algorithms involved, the number of Monte Carlo replications is reduced to 250.
In our empirical application we focus on the most reliable daily indicator for Euro area inflation. In a preliminary analysis Breitung et al. (2013) analyze yield spreads, inflation-indexed bonds, commodity prices and stock price indices as daily indicators for monthly inflation. The choice of these variable was based on the literature on forecasting inflation. Most of the variables are related to inflation expectations. Kozicki (1997) argues that yield spreads reflect expectations on the direction of future inflation changes. The US Treasury began issuing inflation-linked bonds with its treasury inflation-protected securities (TIPS) since 1997. From subtracting yields of inflation-linked bonds from nominal interest rates, market expectations of future inflation can be derived. Stock and Watson (2003) employ overnight interest rates, stock price indices and exchange rates in their analysis of inflation forecasts.

In a comprehensive comparison of alternative predictors Stock and Watson (1999) and Modugno (2013) found that the commodity prices index is the most reliable leading indicator among a variety of candidate predictors. Breitung et al. (2013) analyze the predictive content of 15 daily indicators for the German inflation rate and found that predictive (MIDAS) regressions using a broad commodity price index and crude oil prices yield a regression $R^2$ of more than 0.3, whereas all other predictive regressions exhibit an $R^2$ of less than 0.10. As argued by Bruneau et al. (2007) energy prices do play a dominant role in inflation forecasting, whereas exchange rates are poor predictors of inflation changes. In a similar vein, empirical results suggest that interests rates, inflation-indexed bonds, bond yields and yield spreads do not reveal any considerable predictive power.

In our empirical investigation we therefore focus on the commodity prices index as a daily predictor of inflation. The results for the oil price index are broadly similar. In our empirical application we therefore analyze the predictive power of a daily commodity price index employing parametric and nonparametric MIDAS frameworks.

5.2 Data and preliminary analysis

The target variable is the seasonally adjusted Euro area HICP overall price index for the Euro area as provided by the European Central Bank. The monthly in-
flation rate is computed as \( y_t = 100 \left[ \log(P_t) - \log(P_{t-1}) \right] \), where \( P_t \) denotes the price index, see figure 2. The time span ranges from 2000m1 until 2013m12. The univariate time series is well represented by an AR(1) process, yielding a pretty low \( R^2 \) of 0.07, cf. Table 3 for the full-sample estimation results.

Our daily predictor is the HWWI total commodity price index (Code S275VR). This price index is derived from world market prices of various commodities and provided by the Hamburgisches WeltWirtschaftsInstitut (HWWI) on a daily (Euro) basis. For our preliminary analysis the aggregated monthly indicator is computed as the change rate from the first to the final available day of the respective month. Table 3 presents the estimation results of the full-sample predictive regressions including the lagged commodity price index as an additional regressor. It turn out that this predictor improves the fit up to an \( R^2 \) of 0.425, suggesting that changes of the commodity price index is indeed a powerful predictor for monthly inflation rates. Furthermore, the lagged inflation rate becomes insignificant and, therefore, we drop the lagged dependent variable in our predictive regressions. Finally we observe that any further lag of changes in the commodity price index is insignificant and does not improve the fit.

An important practical problem is that the number of daily observations in a month is not constant: months have different lengths (from 28 up to 31 days) and some observations are missing due to weekends, holidays, publication lags and so on. In fact the number of daily observations ranges from 17 to 23. As there is no silver bullet to fix this problem, previous research has addressed it in different ways. Andreou et al. (2013) in their work on the predictive ability of financial time series fixed the number of trading days in a month to 22. In a similar vein, Hamilton (2008) assumes 21 business days in a month when analyzing measures of the policy impact at daily frequencies. Dias and Embrechts (2004) linearly interpolate missing observations of foreign exchange spot rates for US dollar against Deutsche Mark. Hence, there are different possible strategies to deal with missing daily observations and which approach prevails is an empirical matter. In the daily time series we use, there are not enough observations in each month to fix the length equal to 22 or 23 days and cutting all months to 17 days leads to an unacceptable loss of information. The simplest way to deal with this problem is to ignore the missing observations and include a fixed number of lags in the regression so that the actual time span may be longer than the number of lags. We also experimented with linearly interpolating missing observations in the daily series and use imputed observations in the regression. Overall, the
results are not much affected by the alternative ways to deal with missing observations and, therefore, we only present results where the gaps between daily observations are ignored.\footnote{In our preliminary work (Breitung et al. (2013)) we present results based on a linear interpolation of the missing values.}

Figure 3 depicts the adjusted $R^2$ as a function of the number of daily lags $p$. It turns out that the goodness of fit converges to a value of roughly 0.55 after 30 lags. This implies that including 30 lags should be sufficient to capture the predictive content of the daily change of commodity prices.

5.3 In-sample results of MIDAS regressions

Table 4 presents the estimation results for the parametric MIDAS regression using the HWWI index as the predictor:

$$y_{t+h} = \alpha_0 + \alpha_1 \left( \sum_{j=0}^{p} \omega_j(\theta) x_{t,j} \right) + u_{t+h}$$

where

$$\omega_j(\theta) = \frac{\exp(\theta_1 j + \theta_2 j^2)}{\sum_{i=0}^{p} \exp(\theta_1 i + \theta_2 i^2)}.$$ 

The estimated parameters for different forecast horizons $h$ and lag lengths $p$ are reported along with their $t$-statistics. Interestingly, the $R^2$ of the daily HWWI series exceeds 0.3 only for $h = 1$ and $p \in \{30, 50\}$, which suggests that the HWWI index is a useful predictor only for the next month and $h > 10$. A plausible reason for this is that prices need 20 up to 30 days to adjust to changes in commodity (or energy) prices. Related research (e.g. Kilian 2008) finds a similar response lag.

To test the hypothesis that the lag distribution is flat (i.e. the weights are identical for all 30 lags), the regression is performed by using the simple average of 30 days, yielding an $R^2$ of 0.3421 for $h = 1$. The $F$-statistic of the joint hypothesis $\theta_1 = \theta_2 = 0$ takes a value of 3.689 which is significant with a $p$-value less than 0.01.

The lag distributions resulting from different estimation methods are presented in figures 4a - 4c for the case of 20, 30 and 50 daily lags. Unrestricted OLS estimation of the weights already suggest a hump-shaped weight function,
and the parametric and nonparametric estimators smooth out these erratic unrestricted estimates. Moreover, in the preferred specification with 30 daily regressors, the lag distribution estimated by SLS follows the unrestricted estimates slightly more closely than the parametric MIDAS estimates, with a maximum around 5 lags, while the parametric estimator suggests a maximum around 9 lags.

5.4 Out-of-sample forecasting comparison

Following Kuzin et al. (2009), direct forecasts are performed based on a horizon-specific model (see also Marcellino et al., (2006)). For this forecasting approach we regress the future values of the dependent variable (denoted by $y_{t+h}$) on current or past values of the regressors. Consequently, all parameters of the model are re-estimated for each forecast horizon.

We now turn to forecasting monthly and quarterly inflation in the Euro area. The samples are split into an estimation sample $t = 1, \ldots, T^e$ and a forecasting sample $t = T^e + 1, \ldots, T$, where $n_f = T - T^e$ denotes the number of forecasts. The estimation sample runs from January 2000 to December 2006 and the forecasting exercise covers January 2007 until December 2013. Forecasts are obtained recursively with a moving estimation window. Given the estimated lag distributions obtained from the estimation sample, the first one-period-ahead forecast is produced. Then the first observation from the estimation sample is dropped while the end of the sample is extended to include the observations in the next period. The MIDAS regressions are re-estimated and inflation in the next period is forecasted. This forecasting scheme continues until the penultimate period in the sample is reached and inflation is forecasted for the last period in the sample.

We evaluate forecasts according to the root mean squared forecast error

$$\text{RMSE} = \sqrt{\frac{1}{n_f} \sum_{t=T^e+1}^{T^e+n_f} (y_{t+h} - \hat{y}_{t+h|t})^2}.$$ 

Table 5 shows the relative forecasting accuracy of the nonparametric (SLS) MIDAS relative to the parametric approach (NLS) for various choices of the lag length. As a benchmark, we also present the RMSE of the forecast based on the mean of the regressors $\bar{x} = (p + 1)^{-1} \sum_{j=0}^{p} x_{t,j}$. Starting with the one-step ahead forecasts, the nonparametric approach produces (slightly) more accurate
forecasts than the benchmark and the parametric (NLS) MIDAS regression. For $h = 2$ all forecasts are uninformative provided that the RMSE is close to the sample standard deviation $\hat{\sigma}_y^2 = 0.1953$ of the target variable.

6 Conclusion

Combining variables sampled at different frequencies involves the problem of estimating a rich autoregressive lag distribution, where the effective sample size is governed by the low frequency variable. The MIDAS regressions represent a simple, parsimonious, and flexible class of time series models that allow the left-hand and right-hand side variables of time series regressions to be sampled at different frequencies. In this framework the lag distribution is represented by a class of parametric functions such as exponential Almon lag polynomials. A drawback of this approach is that the shape of the lag distribution is restricted by a small number of parameters. To allow for a more flexible representation of the lag distribution we only impose a smoothness condition yielding some smoothed version of the unrestricted autoregressive coefficients. The resulting SLS estimator can be seen as a shrinkage estimator that penalizes deviations from a (local) linear functional form. Monte Carlo simulations suggest that the SLS estimator performs similarly to the parametric MIDAS regression if the lag distribution is included in the parametric class of specification. In other cases the SLS estimator may yield a substantially lower MSE than the usual parametric MIDAS regressions. To assess the predictive power of daily indicators for monthly (Euro area) inflation rates, we apply parametric and nonparametric methods to a sample of 14 years. It turns out that the HWWI commodity price index is a useful one-month ahead predictor for inflation rates. The estimated lag-distribution covers around 30 days and is hump-shaped with a maximum at 5–10 days.

Appendix A: First order conditions for the minimum of the AIC criterion

We consider the problem

$$\min_{\lambda \geq 0} \text{AIC}(\lambda),$$
where

\[
AIC(\lambda) = \log \left( (y - \hat{y}(\lambda))' [y - \hat{y}(\lambda)] \right) + \frac{2(\kappa_\lambda + 1)}{T - \kappa_\lambda - 2} .
\]

Assuming a minimum exists in the interior of \( \mathbb{R}_+ \), we consider

\[
\frac{d}{d\lambda} AIC(\lambda) = \frac{d}{d\lambda} \left( \log \left[ (\hat{y}^h - \hat{y}_\lambda^h)' (\hat{y}^h - \hat{y}_\lambda^h) \right] + \frac{2(\kappa_\lambda + 1)}{T - \kappa_\lambda - 2} \right),
\]

where \( \kappa_\lambda = \text{tr} \left[ X (X'X + \lambda D'D)^{-1} X' \right] \). First,

\[
\frac{d}{d\lambda} \left( \log \left[ (\hat{y}^h - \hat{y}_\lambda^h)' (\hat{y}^h - \hat{y}_\lambda^h) \right] \right) = \frac{1}{\text{SSR}_\lambda} \frac{d}{d\lambda} \left( (\hat{y}^h - \hat{y}_\lambda^h)' (\hat{y}^h - \hat{y}_\lambda^h) \right)
\]

with \( \text{SSR}_\lambda = (\hat{y}^h - \hat{y}_\lambda^h)' (\hat{y}^h - \hat{y}_\lambda^h) \). We establish two intermediate results: first,

\[
\frac{d}{d\lambda} P_\lambda = \frac{d}{d\lambda} \left( X (X'X + \lambda D'D)^{-1} X' \right)
\]

\[
= X \left( \frac{d}{d\lambda} (X'X + \lambda D'D)^{-1} \right) X'
\]

\[
= X \left( - (X'X + \lambda D'D)^{-1} (D'D) (X'X + \lambda D'D)^{-1} \right) X'
\]

\[
= -XQ_\lambda^{-1} (D'D) Q_\lambda^{-1} X' \quad (17)
\]

with \( Q_\lambda = (X'X + \lambda D'D) \). Second,

\[
\frac{d}{d\lambda} (P'_\lambda P_\lambda) = \left( \frac{d}{d\lambda} (P'_\lambda) \right) P_\lambda + P'_\lambda \left( \frac{d}{d\lambda} (P_\lambda) \right)
\]

\[
= \left( \frac{d}{d\lambda} (P_\lambda) \right)' P_\lambda + P'_\lambda \left( \frac{d}{d\lambda} (P_\lambda) \right)
\]

\[
= \left( -XQ_\lambda^{-1} (D'D) Q_\lambda^{-1} X' \right) P_\lambda + P'_\lambda \left( -XQ_\lambda^{-1} (D'D) Q_\lambda^{-1} X' \right) \quad (18)
\]

Using (17) and (18),

\[
\frac{d}{d\lambda} \left( (\hat{y}^h - \hat{y}_\lambda^h)' (\hat{y}^h - \hat{y}_\lambda^h) \right) = \frac{d}{d\lambda} \left( -2y^h'P_\lambda y^h + y^h'P'_\lambda P_\lambda y^h \right)
\]

\[
= -2y^h' \left( \frac{d}{d\lambda} (P_\lambda) \right) y^h + y^h' \left( \frac{d}{d\lambda} (P'_\lambda P_\lambda) \right) y^h
\]

\[
= 2y^h'XQ_\lambda^{-1} (D'D) Q_\lambda^{-1} X' (\hat{y}^h - \hat{y}_\lambda^h) \quad (19)
\]
Next,
\[
\frac{d}{d\lambda} \left( \frac{2 (\kappa_{\lambda} + 1)}{T - \kappa_{\lambda} - 2} \right) = \frac{2 \frac{d}{d\lambda} (\kappa_{\lambda}) (T - \kappa_{\lambda} - 2) + 2 (\kappa_{\lambda} + 1) \left( \frac{d}{d\lambda} (\kappa_{\lambda}) \right)}{(T - \kappa_{\lambda} - 2)^2}
\]

and
\[
\frac{d}{d\lambda} (\kappa_{\lambda}) = \frac{d}{d\lambda} \left( \text{tr} [P_{\lambda}] \right)
= \text{tr} \left[ \frac{d}{d\lambda} (P_{\lambda}) \right]
= -\text{tr} \left[ (X'X)^{-1} (D'D) Q^{-1}_{\lambda} \right]
\]

Inserting (21) into (20) yields
\[
\frac{d}{d\lambda} \left( \frac{2 (\kappa_{\lambda} + 1)}{T - \kappa_{\lambda} - 2} \right) = -\frac{2 (T - 1) \text{tr} \left[ (X'X)^{-1} (D'D) Q^{-1}_{\lambda} \right]}{(T - \kappa_{\lambda} - 2)^2}
\]

Taken together, a necessary condition for the minimizer\( \lambda_{AIC}^* \) is
\[
\left( \text{SSR}_{AIC}^{-1} \right) \left( \frac{y^h' X Q^{-1}_{AIC} (D'D) Q_{\lambda}^* X' \left( y^h - \hat{y}^h_{AIC} \right)}{T - 1} \right) \text{tr} \left[ (X'X)^{-1} (D'D) Q_{AIC}^{-1} \right] = \frac{T - 1}{(T - \kappa_{AIC}^* - 2)^2} \text{tr} \left[ (X'X)^{-1} (D'D) Q_{AIC}^{-1} \right]
\]

This condition can be solved numerically. A sufficient condition for a minimizer is obtained by checking the second-order derivative at\( \lambda_{AIC}^* \). Using the same rules for matrix differential calculus as above, we find
\[
\frac{d^2}{d\lambda^2} \text{AIC} (\lambda) = \text{SSR}_{\lambda}^{-2} \left\{ \left( 2 y^h' X Q_{\lambda}^{-1} (D'D) Q_{\lambda}^{-1} X' y^h - 4 y^h' X Q_{\lambda}^{-1} (D'D) Q_{\lambda}^{-1} X' \left( y^h - \hat{y}^h_{\lambda} \right) \right) \text{SSR}_{\lambda} - \right\}
\]
\[
\left\{ \text{tr} \left[ (X'X)^{-1} (D'D) Q_{\lambda}^{-1} (D'D) Q_{\lambda}^{-1} \right] \right\} + 4 (T - 1) (T - \kappa_{\lambda} - 2)^{-3}.
\]
\[
\left\{ \text{tr} \left[ (X'X)^{-1} (D'D) Q_{\lambda}^{-1} (D'D) Q_{\lambda}^{-1} \right] \right\} (T - \kappa_{\lambda} - 2)
\]
\[
+ \left( \text{tr} \left[ (X'X)^{-1} (D'D) Q_{\lambda}^{-1} \right] \right)^2
\]

20
References


Figure 1: From top to bottom: exponentially declining, hump-shaped, linear and cyclical, sign-changing weights for $p + 1 \in \{20, 40, 60\}$

Note: The weighting functions (13) - (16) employed in (9)-(11) are presented here for $T = 100$, where for the exponentially declining weights, $\theta_1 = 7 \cdot 10^{-4}$, $\theta_2 = -6 \cdot 10^{-3}$, while for the humped-shaped weights $\theta_1 = 8 \cdot 10^{-2}$ and $\theta_2 = \theta_1/10$, $\theta_2 = \theta_1/20$ and $\theta_2 = \theta_1/30$ for $p + 1 = 20$, $p + 1 = 40$, $p + 1 = 60$, respectively, yielding a maximum at the lag 5, 10 and 15, respectively.
Table 1: In-sample MSE-ratios

<table>
<thead>
<tr>
<th>High-frequency lags</th>
<th>p+1 = 20</th>
<th>p+1 = 40</th>
<th>p+1 = 60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_1 = 0.2$</td>
<td>$a_1 = 0.3$</td>
<td>$a_1 = 0.4$</td>
</tr>
<tr>
<td>exp. declining weights</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{\text{MSE}}$</td>
<td>100</td>
<td>0.26</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.29</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.36</td>
<td>0.47</td>
</tr>
<tr>
<td>$\lambda_{\text{AIC}}$</td>
<td>100</td>
<td>0.38</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.43</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.51</td>
<td>0.60</td>
</tr>
<tr>
<td>hump-shaped weights</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{\text{MSE}}$</td>
<td>100</td>
<td>0.41</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.56</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.75</td>
<td>0.98</td>
</tr>
<tr>
<td>$\lambda_{\text{AIC}}$</td>
<td>100</td>
<td>0.47</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.62</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.92</td>
<td>1.45</td>
</tr>
<tr>
<td>linear weights</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{\text{MSE}}$</td>
<td>100</td>
<td>0.24</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.25</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.29</td>
<td>0.28</td>
</tr>
<tr>
<td>$\lambda_{\text{AIC}}$</td>
<td>100</td>
<td>0.36</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.44</td>
<td>0.42</td>
</tr>
<tr>
<td>sign-changing weights</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{\text{MSE}}$</td>
<td>100</td>
<td>0.10</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>$\lambda_{\text{AIC}}$</td>
<td>100</td>
<td>0.53</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.05</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Note: The entries are the in-sample MSE ratios of the nonparametric SLS estimator relative to the parametric NLS estimator with an exponential Almon lag based on two parameters. Data are generated as in (9) – (11), where the weights are generated by (13) – (16). The nonparametric estimator is computed using the MSE minimal choice of the smoothing parameter ($\lambda_{\text{MSE}}$) and by using the modified AIC criterion ($\lambda_{\text{AIC}}$). The number of replications is 1,000.
Table 2: Out-of-sample forecast comparison

<table>
<thead>
<tr>
<th>High-frequency lags</th>
<th>p+1 = 20</th>
<th>p+1 = 40</th>
<th>p+1 = 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_1 = 0.2 )</td>
<td>1.04</td>
<td>1.03</td>
<td>1.18</td>
</tr>
<tr>
<td>( \alpha_1 = 0.3 )</td>
<td>1.04</td>
<td>1.03</td>
<td>1.16</td>
</tr>
<tr>
<td>( \alpha_1 = 0.4 )</td>
<td>1.04</td>
<td>1.04</td>
<td>1.21</td>
</tr>
</tbody>
</table>

**exp. declining weights**

<table>
<thead>
<tr>
<th>( \lambda_{MSE} )</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>100</th>
<th>200</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.04</td>
<td>1.07</td>
<td>1.09</td>
<td>1.04</td>
<td>1.07</td>
<td>1.09</td>
<td>1.04</td>
<td>1.07</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>1.04</td>
<td>1.06</td>
<td>1.09</td>
<td>1.04</td>
<td>1.06</td>
<td>1.09</td>
<td>1.04</td>
<td>1.06</td>
<td>1.09</td>
</tr>
<tr>
<td>MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.03</td>
<td>1.06</td>
<td>1.08</td>
<td>1.03</td>
<td>1.06</td>
<td>1.08</td>
<td>1.03</td>
<td>1.06</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>1.03</td>
<td>1.07</td>
<td>1.08</td>
<td>1.03</td>
<td>1.07</td>
<td>1.08</td>
<td>1.03</td>
<td>1.07</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>1.04</td>
<td>1.07</td>
<td>1.08</td>
<td>1.04</td>
<td>1.07</td>
<td>1.08</td>
<td>1.04</td>
<td>1.07</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>1.18</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
</tr>
</tbody>
</table>

**hump-shaped weights**

<table>
<thead>
<tr>
<th>( \lambda_{MSE} )</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>100</th>
<th>200</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.04</td>
<td>1.07</td>
<td>1.09</td>
<td>1.04</td>
<td>1.07</td>
<td>1.09</td>
<td>1.04</td>
<td>1.07</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>1.04</td>
<td>1.07</td>
<td>1.08</td>
<td>1.04</td>
<td>1.07</td>
<td>1.08</td>
<td>1.04</td>
<td>1.07</td>
<td>1.08</td>
</tr>
<tr>
<td>MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>1.03</td>
<td>1.05</td>
<td>0.98</td>
<td>1.03</td>
<td>1.05</td>
<td>0.98</td>
<td>1.03</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>0.91</td>
<td>0.96</td>
<td>1.01</td>
<td>0.91</td>
<td>0.96</td>
<td>1.01</td>
<td>0.91</td>
<td>0.96</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>1.03</td>
<td>1.06</td>
<td>0.98</td>
<td>1.03</td>
<td>1.06</td>
<td>0.98</td>
<td>1.03</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>1.11</td>
<td>0.94</td>
<td>0.97</td>
<td>1.11</td>
<td>0.94</td>
<td>0.97</td>
<td>1.11</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>0.88</td>
<td>0.78</td>
<td>0.77</td>
<td>0.88</td>
<td>0.78</td>
<td>0.77</td>
<td>0.88</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>0.89</td>
<td>0.95</td>
<td>0.98</td>
<td>0.89</td>
<td>0.95</td>
<td>0.98</td>
<td>0.89</td>
<td>0.95</td>
</tr>
</tbody>
</table>

**linear weights**

<table>
<thead>
<tr>
<th>( \lambda_{MSE} )</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>100</th>
<th>200</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.05</td>
<td>1.06</td>
<td>1.08</td>
<td>1.05</td>
<td>1.06</td>
<td>1.08</td>
<td>1.05</td>
<td>1.06</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.05</td>
<td>1.05</td>
<td>1.00</td>
<td>1.05</td>
<td>1.05</td>
<td>1.00</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>1.02</td>
<td>1.05</td>
<td>1.05</td>
<td>1.02</td>
<td>1.05</td>
<td>1.05</td>
<td>1.02</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>1.01</td>
<td>1.04</td>
<td>1.04</td>
<td>1.01</td>
<td>1.04</td>
<td>1.04</td>
<td>1.01</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>1.21</td>
<td>1.04</td>
<td>1.04</td>
<td>1.21</td>
<td>1.04</td>
<td>1.04</td>
<td>1.21</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>1.20</td>
<td>1.04</td>
<td>1.04</td>
<td>1.20</td>
<td>1.04</td>
<td>1.04</td>
<td>1.20</td>
<td>1.04</td>
<td>1.04</td>
</tr>
</tbody>
</table>

**sign-changing weights**

<table>
<thead>
<tr>
<th>( \lambda_{MSE} )</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>100</th>
<th>200</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.82</td>
<td>0.86</td>
<td>0.86</td>
<td>0.82</td>
<td>0.86</td>
<td>0.86</td>
<td>0.82</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.72</td>
<td>0.72</td>
<td>0.70</td>
<td>0.72</td>
<td>0.72</td>
<td>0.70</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>0.61</td>
<td>0.62</td>
<td>0.63</td>
<td>0.61</td>
<td>0.62</td>
<td>0.63</td>
<td>0.61</td>
<td>0.62</td>
<td>0.63</td>
</tr>
<tr>
<td>MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.04</td>
<td>1.07</td>
<td>1.08</td>
<td>1.04</td>
<td>1.07</td>
<td>1.08</td>
<td>1.04</td>
<td>1.07</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>1.06</td>
<td>1.08</td>
<td>1.11</td>
<td>1.06</td>
<td>1.08</td>
<td>1.11</td>
<td>1.06</td>
<td>1.08</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>1.07</td>
<td>1.09</td>
<td>1.10</td>
<td>1.07</td>
<td>1.09</td>
<td>1.10</td>
<td>1.07</td>
<td>1.09</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>1.11</td>
<td>1.04</td>
<td>1.04</td>
<td>1.11</td>
<td>1.04</td>
<td>1.04</td>
<td>1.11</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>1.15</td>
<td>1.16</td>
<td>1.16</td>
<td>1.15</td>
<td>1.16</td>
<td>1.16</td>
<td>1.15</td>
<td>1.16</td>
<td>1.16</td>
</tr>
</tbody>
</table>

**Note:** The entries are the MSE ratios of the nonparametric SLS estimator relative to the parametric NLS estimator with an exponential Almon lag based on two parameters for exp. declining and humped-shaped weights and three parameters for the sign-changing weights. Data are generated as in (9) – (11), where the weights are generated by (13) – (16). The nonparametric estimator is computed using the MSE minimal choice of the smoothing parameter \( \lambda_{MSE} \) and by using the modified AIC criterion \( \lambda_{AIC} \). The number of replications is 250.
Figure 2: monthly Euro Area HPCI inflation rate

Figure 3: adjusted $R^2$ as a function of the lag length $p$
Figure 4a: Estimated lag distributions with $p + 1 = 20$

Figure 4b: Estimated lag distributions with $p + 1 = 30$

Figure 4c: Estimated lag distributions with $p + 1 = 50$

Note: OLS (dashed line), nonparametric SLS (straight line) and parametric NLS (dashed-dotted line) estimates for $h = 1$ and the full sample based on 168 monthly observations.
Table 3: Preliminary analysis: in-sample regressions

<table>
<thead>
<tr>
<th>regressor</th>
<th>$y_t$: monthly inflation rate (EU area)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
</tr>
<tr>
<td>$y_{t-1}$</td>
<td>0.265</td>
</tr>
<tr>
<td></td>
<td>(3.53)</td>
</tr>
<tr>
<td>HWWA$_{t-1}$</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
</tr>
<tr>
<td>HWWA$_{t-2}$</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
</tr>
<tr>
<td>const</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>(6.85)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.070</td>
</tr>
<tr>
<td>LM(6)</td>
<td>0.599</td>
</tr>
<tr>
<td>LM(12)</td>
<td>0.522</td>
</tr>
</tbody>
</table>

Note: $t$-statistics presented in parentheses. LM($k$) reports the $p$-value of the LM (Breusch-Godfrey) test for autocorrelation up to the lag order $k$. 
### Table 4: In-sample estimation of MIDAS regression

\[
y_{t+h} = \alpha_0 + \alpha_1 \sum_{j=0}^{p} \omega_j (\theta_j) x_{t-j} + u_{t+h}
\]

\[
y_{t+h} = \alpha_0 + \alpha_1 \sum_{j=0}^{p} \omega_j (\theta_j) x_{t-j} + u_{t+h}
\]

<table>
<thead>
<tr>
<th>h = 1</th>
<th>p + 1 = 50</th>
<th>p + 1 = 30</th>
<th>p + 1 = 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_0)</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>(16.13)</td>
<td>(16.09)</td>
<td>13.90</td>
</tr>
<tr>
<td>(\alpha_1)</td>
<td>0.358</td>
<td>0.359</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>(9.86)</td>
<td>(9.94)</td>
<td>(5.57)</td>
</tr>
<tr>
<td>(\theta_1)</td>
<td>0.158</td>
<td>0.151</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td>(2.31)</td>
<td>(2.25)</td>
<td>(1.15)</td>
</tr>
<tr>
<td>(\theta_2)</td>
<td>-0.010</td>
<td>-0.009</td>
<td>-0.030</td>
</tr>
<tr>
<td></td>
<td>(-2.90)</td>
<td>(-2.85)</td>
<td>(-1.02)</td>
</tr>
<tr>
<td>R² NLS</td>
<td>0.443</td>
<td>0.444</td>
<td>0.202</td>
</tr>
<tr>
<td>R² SLS</td>
<td>0.451</td>
<td>0.479</td>
<td>0.199</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>h = 2</th>
<th>p + 1 = 50</th>
<th>p + 1 = 30</th>
<th>p + 1 = 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_0)</td>
<td>0.0017</td>
<td>0.0016</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>(12.42)</td>
<td>(12.34)</td>
<td>(12.49)</td>
</tr>
<tr>
<td>(\alpha_1)</td>
<td>0.059</td>
<td>0.069</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(1.08)</td>
<td>(1.38)</td>
<td>(-0.53)</td>
</tr>
<tr>
<td>(\theta_1)</td>
<td>0.254</td>
<td>0.309</td>
<td>-2.045</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
<td>(0.43)</td>
<td>(-0.31)</td>
</tr>
<tr>
<td>(\theta_2)</td>
<td>-0.007</td>
<td>-0.009</td>
<td>0.205</td>
</tr>
<tr>
<td></td>
<td>(-0.42)</td>
<td>(-0.43)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>R² NLS</td>
<td>0.019</td>
<td>0.002</td>
<td>0.006</td>
</tr>
<tr>
<td>R² SLS</td>
<td>0.051</td>
<td>0.007</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Note:** Estimates based on 168 monthly observations. \(t\)-values are reported in parenthesis. R² NLS is the R² from the parametric MIDAS regression, while R² SLS refers to the R² from the non-parametric approach, where the smoothing parameter is selected by the AIC criterion.
Table 5: Out-of-sample forecast comparison

<table>
<thead>
<tr>
<th></th>
<th>$p + 1 = 20$</th>
<th>$p + 1 = 30$</th>
<th>$p + 1 = 40$</th>
<th>$p + 1 = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE($\bar{x}$)</td>
<td>0.1516</td>
<td>0.1667</td>
<td>0.1735</td>
<td>0.1736</td>
</tr>
<tr>
<td>RMSE(NLS)</td>
<td>0.1511</td>
<td>0.1541</td>
<td>0.1543</td>
<td>0.1621</td>
</tr>
<tr>
<td>RMSE(SLS)</td>
<td>0.1509</td>
<td>0.1521</td>
<td>0.1512</td>
<td>0.1584</td>
</tr>
<tr>
<td>$h = 2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE($\bar{x}$)</td>
<td>0.1955</td>
<td>0.1953</td>
<td>0.1952</td>
<td>0.1982</td>
</tr>
<tr>
<td>RMSE(NLS)</td>
<td>0.1982</td>
<td>0.1962</td>
<td>0.1971</td>
<td>0.1983</td>
</tr>
<tr>
<td>RMSE(SLS)</td>
<td>0.1935</td>
<td>0.1966</td>
<td>0.1969</td>
<td>0.1948</td>
</tr>
</tbody>
</table>

Note: RMSE($\bar{x}$) indicates the root mean squared error using the mean of $p + 1$ daily observations as the predictor. RMSE(NLS) and RMSE(SLS) denote the root mean squared error of the parametric MIDAS based on an exponential Almon lag distribution with $k = 2$ and the nonparametric MIDAS estimator, where the smoothing parameter $\lambda$ is determined by the AIC criterion. Initial estimation sample uses $T_e = 84$ ($T_e = 83$) months. Forecasts are obtained with a rolling window of size $T_e$ observations by successively dropping the first period in the sample and incorporating the following period.