**Research Department** **Bank of Israel**

# Forecasting Short run Inflation Using Mixed Frequency Data (MIDAS)

**Sigal Ribon* and Tanya Suhoy****

Occasional Paper 2011.02
December 2011

Research Department, Bank of Israel – http://www.boi.org.il
* Sigal Ribon – Phone: 972-2-6552610; E-mail: sigal.ribon@boi.org.il
** Tanya Suhoy – Phone: 972-2-6552620; E-mail: Tanya.suhoy@boi.org.il



חטיבת המחקר, בנק ישראל ת"ד 780 ירושלים 91007
Research Department, Bank of Israel, POB 780, 91007 Jerusalem, Israel

# חיזוי האינפלציה לטווח הקצר באמצעות מידע בתדירויות שונות (MIDAS)

**טניה סוחוי וסיגל ריבון**

## תקציר

אנו מציגות מודל עם נתונים בתדירויות שונות (MIDAS) לחיזוי השינוי במדד המחירים לצרכן לטווחים של חודש וחודשיים. המודל המוצג מאפשר לנו לכלול נתונים פיננסיים ומחירי סחורות עולמיים בתדירות יומית על ידי התאמת התפלגות ביתא גמישה לפיגורים היומיים של המשתנים המסבירים. בהינתן אורך הפיגורים, ניתן לאמוד את הפרמטרים האופטימליים של התפלגויות אלו בו-זמנית עם אמידת המקדמים של הרגרסיה. אנו בוחנות גם מודל בייסיאני גמיש יותר, המאפשר להעריך את אורך הפיגורים הסביר בהתבסס על תדירות הופעתם בדגימה בשיטת Gibbs. נמצא שניסוח ה-MIDAS המוצע משפר את יכולת החיזוי יחסית למודל עם התפלגות אחידה של הפיגורים היומיים וכן יחסית למודל עם משתנים מסבירים בתדירות חודשית בלבד. עוד נמצא כי העיתוי הטוב ביותר לעריכת תחזית לטווח של חודש הוא השבוע השלישי של החודש. הנתונים לשבועיים הראשונים של החודש ומדד המחירים של החודש הקודם, הידוע בעיתוי זה תורמים לשיפור התחזית. הוספת השבועיים האחרונים של החודש לא תורמת לביצועי המודל.

# Forecasting Short Run Inflation Using Mixed Frequency Data (MIDAS) *

Sigal Ribon and Tanya Suhoy

**Abstract**

We present a MIxed DAta Sampling model for one-month- and two-months-ahead forecasts for the monthy changes in the Israel's CPI. This model enables us to incorporate daily financial and commodity price data in a monthly model by imposing a flexible Beta function on the lag distribution of the daily explanatory variables. Given the lag length, the parameters of those distributions can be optimized simultaneously with the regression coefficients. We also consider a more flexible Bayesian model, enabling to evaluate inter alia the most likely lag lengths, based on the frequency of their appearance in the Gibbs sample. We find that the proposed MIDAS specification improves the forecast ability, measured by the RMSFE (Root Mean Square Forecast Error) and MAFE (Mean Absolute Forecast Error), relative to a model with uniformly distributed daily lags and a model with only monthly frequency data. We also find that the preferred timing to perform the one-month-ahead forecast is on the third week of the forecasted month. The first two weeks of daily data and information about the previous month's CPI, both contribute to the improvement of the forecast accuracy. The addition of the two last weeks of the month does not contribute to the performance of the model.

KEYWORDS: Inflation, forecasting, MIDAS, Beta pdf, daily data.
JEL classification: C53, E31, E37.

## 1 Introduction

Forecasting short-term inflation is an important ingredient in the assessment of the inflationary environment for the implementation of monetary policy. Central banks, among them the Bank of Israel, use different monthly models for forecasting short-term inflation, or more accurately, the rate of price changes in the next few months. Among the models used by the Bank of Israel are: a statistical model (Suhoy and Rotberger, 2006), based on separate forecasts of the rate of change of the ten main groups of the total

*We thank Guy Segal and Avichai Sorezcky for performing simulations of alternative models. We also thank Avichai Sorezcky and the participants of the Bank of Israel Research Department seminar for their helpful comments and suggestions.

CPI; an econometric model (Ilek, 2006), taking into account the effect of macroeconomic variables, a small Bayesian VAR model (Segal, 2010) and a single AR equation (Sorezcky, 2009).

Some of these existing models, and in particular the statistical model, rely heavily on the ability to forecast the change in the housing component of the CPI which constitutes about a quarter of total CPI. However, the ability to forecast this component has deteriorated significantly in recent years along with the de-dollarisation of the housing services pricing mechanism. As long as prices were linked to the dollar, knowing the rate of change in the exchange rate made it possible to construct a relatively accurate forecast of this component, and hence of the overall CPI. As the tendency to set prices in dollars declined, so did the accuracy of the statistical forecast. This development serves as a major incentive for the present project, which explores a different channel for improving the ability to forecast short-term price changes.

The aim of this paper is to offer an additional instrument for forecasting short-term CPI rate of change using an approach known as MIDAS - Mixed DAta Sampling, which enables to combine data of different frequencies, and in particular the use of daily frequency data in addition to monthly frequency data commonly used for forecasting the monthly change in the CPI. The daily data enables us to match data relating to parts of calendar months, e.g., last 45 days or last two weeks, and the monthly CPI. It also allows to assign different weights to daily lags of the data.

The MIDAS procedure has become rather popular in recent years. Although there are alternative procedures that allow us to incorporate high frequency data in forecasting lower frequency variables, one of the major advantages of the MIDAS procedure is its simplicity in terms of specification, estimation and computation of forecasts. The first to introduce this method were Ghysels, Santa-Clara and Valkanov (2002), who proposed the integration of mixed frequency data within a unified model, while preserving a parsimonious specification by assuming a predetermined distribution for the lagged effects of the higher frequency data.[1] Originally this technique focused on volatility predictions (see for example Chen and Ghysels, 2011). A number of authors employed this technique in two major areas. The first is improving quarterly GDP forecasts by incorporating higher frequency data. This was done by Kuzin, Marcellino and Schumacher (2009) for the euro zone and by others cited by Armesto, Engermann and Owyang (2010). Suhoy (2010) employed the MIDAS technique for nowcasting quarterly private consumption in the Israeli economy, using a set of monthly indicators. The other strand of literature exploits daily data, usually from the financial markets for different analyses. Financial data may contain potential, additional information to enhance our forecasting ability because it contains a forward-looking perception of the economic environment and is typically observed in real time with a negligible measurement error. Armesto, Engermann and Owyang (2010) mention a number of studies in this area. We will mention Tay (2007) who uses daily stock returns in order to forecast quarterly real output growth, Ghysels and Wright (2009) who included daily interest rates in a MIDAS model in order

[1] Andreou, Ghysels and Kourtellos (2010a, 2010b) review other types of regression models that involve data sampled at different frequencies.

to predict the outcome of a quarterly expectations survey, and Andreou, Ghysels and Kourtellos (2010b) who test for the improvement embodied in MIDAS factor models and find that there are substantial gains in forecasting inflation and real activity using daily financial data within Factor-MIDAS models. They also show that MIDAS models can efficiently incorporate leading information from daily predictors (nowcasting). Kotze (2005) uses daily asset prices to forecast the monthly rate of inflation, incorporating multiple regressors, including interest rates, spreads, stock prices and exchange rates. He finds that the use of high frequency asset prices does not improve the results of traditional models with aggregate data. Another recent paper by Monteforte and Moretti (2010) incorporates daily financial data in a MIDAS model together with a monthly core inflation index derived from a dynamic factor model, for daily forecasts of euro zone inflation. They find that the inclusion of daily variables helps to reduce forecast errors relative to models with only monthly data. They also find that their daily forecasts are not more accurate than those extracted from daily quotes of future contracts, but are less volatile.

We estimate our MIDAS model using two different approaches. The first is partially restricted so that the lengths of explanatory lags (model space) are assigned based on empirical regularities. The second approach is flexible, assuming posterior distributions of daily lengths have to be evaluated, together with the other parameters. In the first case – when the lag lengths are predetermined – we use a Newton-Raphson optimization of Beta-distributed weights and regression coefficients. To proceed with a flexible specification, we apply Gibbs sampling. After the posterior distribution of daily lengths is obtained, it enables us to construct a Beta-mixed distribution of daily lags, which appears to be more plausible. Thus, the daily weights of each explanatory variable are constructed as a mixture of three Beta distributions with different lengths with the highest posterior probabilities.

Our results show that the MIDAS model, incorporating daily data with Beta distributed weights of daily lags with varying lengths, improves the forecasting performance as measured by the RMSFE and MAFE[2] relative to uniformly distributed lagged daily data and to a benchmark monthly model. We also find that the preferred timing to perform the one-month-ahead forecast is in the third week of the forecasted month, when the CPI of the previous month is already known and two weeks of daily data are available. We find that the two weeks' additional daily data are important at least as much as the information about the previous month's CPI which becomes available in the middle of the month. Comparing our results to other available monthly models in the Bank of Israel, we find that the MIDAS model performs as well as some of the models and significantly better than others. The model we present joins the suite of monthly models currently in use in the Research Department of the Bank of Israel, and as such is expected to contribute to our ability to understand the inflation environment and project its short-term development.

The rest of the paper is organized as follows: The next section describes the daily and monthly data, Section 3 presents the MIDAS model and its estimation and Section

[2] Root Mean Square Forecast Error and Mean Absolute Forecast Error

4 presents the forecast results. Section 5 concludes the paper.

# 2 The Data

## 2.1 Daily frequency explanatory variables

When assessing the inflation environment, central banks look at a wide range of indicators, including financial variables, as the exchange rate, the yield curve and break-even market inflation expectations. Financial variables have the advantage of being available immediately on a daily basis and they are not subject to revisions and updates as are some other macroeconomic indicators. Assuming perfect markets, which is reasonable to assume given the fairly deep and sophisticated financial markets, these variables reflect the public's perceptions about the "right" prices of different assets, and therefore reflect correctly their concept about future inflation.

We chose to include several daily financial and market variables:

*The exchange rate:* The exchange rate is expected to affect price changes in two ways. The first is the usual effect on prices of imported goods whose price in local terms depends on the exchange rate as well as their global price (which we also refer to). Because global price indices are measured in dollar terms, the relevant exchange rate to be included is the dollar/shekel exchange rate. The second reason for including the exchange rate goes back to the historic Israeli pricing mechanism in the housing market, which was characterized until the end of 2007 by dollar denomination of almost all transactions (buying and renting). Because of the dramatic decline in the share of dollar denominated contracts in the housing market, this effect is decreasing over time, and may be expressed by multiplying the dollar exchange rate by the share of dollar denominated contracts. These two effects of the exchange rate cannot be reflected sepaerately because they are highly correlated for most of the period. Therefore we try different specifications, among them, including the effective exchange rate - a basket of currencies weighed according to Israel's trade with its trading partners together with the adjusted dollar exchange rate multiplied by the share of dollar contracts, or specifications including only the dollar exchange rate - adjusted or as is. In addition to the results presented, we also experimented with versions that excluded the effective exchange rate and included only the dollar exchange rate, with or without multiplication by the share of dollar rent contracts. The in-sample results for these alternative specifications are very similar to those presented in the table, with a slightly lower $R^2$.

*Inflation expectations:* The exsistence of both nominal and CPI-indexed government bonds in the Israeli bond market enables us to derive break-even inflation expectations for different horizons. We include the one-year ahead market based expectations in our model. These are often refered to in the monetary policy decision process in the central bank and are expected to provide an indication about the market's perception of future inflation and therefore influence decisions concerning price updating.

*Bank of Israel's rate:* Although the Bank of Israel's rate is normally set only once a month, because there does not exist a perfect correspondance between the calendar months and the "liquidity months"[3] for which the interest rate is set, including a daily series may be beneficial. In addition, this allows forecasts from the middle of the month ($17^{th}$) to include the relevant interest rate for the different daily lags included, instead of computing some weighted average rate of the monthly interest rate.

*Commodity, food and fuel price indices:* As mentioned before, world prices affect the CPI directly through the prices of imported final goods, but also have a significant indirect effect on the prices of local goods through the prices of intermediate goods, especially energy prices.

We use different variables as indicators of world commodity prices, as will be described below. Although we do not expect local prices to follow daily fluctuations in these indices, the option to choose the right lag and range in which these variables may affect local prices may improve the ability to understand and forecast monthly price changes. The specific indices we examine are the oil price, as Cushing oil prices published by Bloomberg, and Bloomberg‘s agricultural commodity prices.[4]

We also tried to include the daily changes in the TASE share index, as a proxy for wealth or expectations about future real activity, but the effect of this variable was found to be insignificant. Another variable we tried to take account of tax changes - changes in the rate of VAT and in the lump-sum tax on gasoline. Neither had a significant effect and therefore they were dropped from the estimation. Still, information about expected changes in these taxes may be included by discretion in the process of forecasting.

Our sample consists of data from November 1999 to September 2010 (a total of 131 months). All financial variables, except for inflation expectations and the BoI interest rate are expressed as log differences. Some of the financial variables do not have actual data for days which are not trading days.[5] In order to avoid discontinuities in the data, missing values for these days were completed by interpolating the levels of the data for missing days, taking into account the intra-monthly seasonality due to the trading days effect.

We present some basic statistics for the daily variables and the monthly change in the CPI (dp) in Table 1. For the daily frequency variables, the mean represents the average over the sample of the monthly averages of the daiy rate of change in the variable (excluding the BoI interest rate and inflation expectations, which are expressed in levels). The standard deviation is that of the monthly average rate of change. We also persent the mean monthly standard deviation, which is the mean of the standard deviation of the daily changes (or levels for interest rate and expectations) within each month.

As seen in the table, there is considerable variation in the daily rate of change of the financial variables (second column from the right), and also changes in the magnitude of variation between daily changes between different months (first column from the right).

[3] A liquidity month starts on the last Thursday of the previous calendar month.

[4] We checked alternative specifications with Bloomberg's CRB index which is constituted from a large number of price indices and Bloomberg's forward food index, but found that the estimation results with the agriculture index are better.

[5] Trading days in the FOREX market do not include Saturday and Sunday. Inflation expectations, derived from the bond market, do not have data for Fridays and Saturdays.

Table 1: Mean and standard deviation of daily changes. Inflation expectations and BoI interest rate in levels; all other variables in rate of change. 1999.11-2010.9

| | Mean daily value (%) | | Std. of daily value (%) | |
|---|---|---|---|---|
| Between months | Average | Std. | Average | Std. |
| Inflation expectations (level) | 1.90 | 0.81 | 0.003 | 0.002 |
| BoI interest rate (level) | 5.00 | 2.70 | 0.066 | 0.134 |
| $ exchange rate | -0.004 | 0.064 | 0.313 | 0.154 |
| $ exchange rate * $ contract share | -0.003 | 0.038 | 0.195 | 0.091 |
| Effective exchange rate | -0.001 | 0.070 | 0.390 | 0.150 |
| Oil prices | 0.044 | 0.319 | 1.766 | 0.641 |
| Agricultural commodities | 0.018 | 0.137 | 0.725 | 0.216 |
| dp (monthly) | 0.17 | 0.504 | – | – |

The average daily rate of change in commodity prices is relatively large (first left column), and the daily standard deviation (within a month) is very substantial.

### 2.2 Monthly frequency variables

The monthly frequency data we include in the specification are the lagged monthly log differences of the CPI, which were found to improve the estimation results, and the monthly dummy variables to take account of seasonal effects and the holiday-adjustment shift variables to correct for the Jewish holidays.

## 3 The model and its estimation

### 3.1 The general framework

In this section we present the mixed frequency model we use in order to estimate and forecast the rate of CPI changes. As mentioned above, the specification allows us to link the change in the CPI which is measured monthly and information from the markets, and in particular the financial market, which is, by nature, measured in daily frequency. The main idea of the MIDAS framework is to construct monthly frequency variables using daily data together with some assumptions on the distributed lagged effect of the daily variables on the monthly aggregate, or in other words, to look at a weighted average of daily variables, with the lags weighted according to a chosen distribution. Once the monthly variable is constructed, we are back to a "regular" monthly frequency estimation. The general form of the equation will be:

(1)
$$\pi_t = \zeta_0 + \sum_{i=1}^{N} \beta_i \sum_{k=1}^{K_i} b_i\left(k; \boldsymbol{\theta}\right) L^{\frac{k}{K_i}} x_{i,t}^{k} + \gamma' \mathbf{Z}_t + \varepsilon_t \ ,$$

where $\pi_t$ is the monthly inflation rate; for each variable $i$[6], $b_i\left(k;\boldsymbol{\theta}\right)$ is a function of a small set of parameters, $\boldsymbol{\theta}$, determining the lag distribution; $K$ is the number of daily lags of variable $i$; $N$ is the number of daily variables; and $\mathbf{Z}_t$ is a set of monthly variables. Specifically, we write :

$$b_i\left(K;\boldsymbol{\theta}\right) = \frac{f\left(k;\boldsymbol{\theta}\right)}{\sum_{k=1}^{K_i} f\left(k;\boldsymbol{\theta}\right)} \quad . \tag{2}$$

We assume a specific distribution for the effect of the daily data on the monthly inflation rate and choose to concentrate on the common Beta distribution, proposed by Ghysels et al. (2002), because it is very flexible and allows for a wide range of specific lag distributions.[7] The Beta pdf specification for each variable $i$, includes three parameters: $K$, $\theta_1$ and $\theta_2$ that have to be set:

$$f_i\left(K;\theta_1,\theta_2\right) = \left(\frac{\Gamma(\theta_1)\Gamma(\theta_2)}{\Gamma(\theta_1+\theta_2)}\right)^{-1}\left(\frac{k}{K}\right)^{\theta_1-1}\left(1-\frac{k}{K}\right)^{\theta_2-1} \quad , \tag{3}$$

where $k$ is the lag and $K$ is the assumed length of the wave for variable $i$, so that $\frac{k}{K}$ is between 0 and 1. $\Gamma$ is the Gamma function: $\Gamma(s) = \int_0^\infty e^{-v}V^{s-1}dv$. The Beta density is very flexible and allows many shapes of the weighting function, depending on the two parameters $\theta_1$ and $\theta_2$ including a uniform distribution, hump, or a decreasing density function. As a particular case, the uniform distribution can be viewed as Beta(1,1); Beta with $\theta_1$=1 $\theta_2$>1 is strictly decreasing and Beta with $\theta_1$>1 and $\theta_2$=1 is strictly convex. When $\theta_1$=$\theta_2$ the density is symmetric about 1/2.

Equation (1) may include as one of the monthly $\mathbf{Z}_t$ variables the inflation rate in the previous month, $\pi_{t-1}$, representing the inflation environment with an AR(1) process, and implying a backward-looking perception of the inflationary process. In addition, we also represent the inflationary environment by daily market based break-even inflation expectations,[8] as one of the $N$ daily-frequency variables that are included in the model, implying a forward-looking process of the inflation.[9] In addition, the $\mathbf{Z}_t$ vector of monthly variables includes only monthly seasonal dummy variables and an adjustment variable for shifting the Jewish holidays, based on the lunar year, relative to the Gregorian calendar. We may write Equation (1) as:

$$\pi_t = \zeta_1\pi_{t-1} + \sum_{i=1}^{N}\beta_i\sum_{k=1}^{K_i} b\left(k;\boldsymbol{\theta}\right)L^{\frac{k}{K}}x_{i,t}^{k} + \gamma'\mathbf{D}_t + \varepsilon_t \ , \tag{1a}$$

[6] For the sake of brevity we omit the index $i$ from the parameters of the distribution.

[7] The other common distribution is the exponential Almon polynomial, introduced by Almon (1965), which assumes: $f\left(k;\boldsymbol{\theta}\right) = e^{\theta_1 k+\theta_2 k^2+..+\theta_Q k^Q}$ with $Q$ and $\boldsymbol{\theta}$ predetermined.

[8] We found that the inclusion of lagged inflation, together with the daily break-even market expectations improves the fit and the forecast ability of the equation.

[9] Expected inflation itself may be backward looking to some extent.

where $x$ includes the daily variables, and $\mathbf{D}_t$ is the vector of dummy variables and the calander-adjustment variable.

In order to evaluate the improvement embedded in the inclusion of daily data, we set two benchmark models. The first includes daily data with the same lag length, $K_i$, but uniformly distributed. This will enable us to evaluate the improvement that may be attributed to the non-uniform distribution of the lagged effects of the daily variables. The second benchmark model will include the same financial variables, but as monthly averages instead of daily data. That is, we revert to the "standard" monthly model. The monthly data restricts us to a wave with a length of full months, and the lag available is that of the difference between the time of the forecast and the end of the previous month (or a number of months before that). This will allow us to check the importance of using daily data.

### 3.2 Parameter Estimation

The inclusion of daily data in the model offers a range of possible specifications of the model and its estimation. The three parameters that have to be set for the Beta distribution for each of the daily variables are the number of daily lags $K$ - the maximum lag we assume to have any effect on a given month's inflation, and $\theta_1$ and $\theta_2$ which determine the shape of the distribution of the lagged effects. Thus, we consider our MIDAS-model in two approaches: the first is partially restricted so that the lengths of explanatory lags are assigned upon some empirical knowledge. The second approach is flexible, assuming posterior distributions of lag lengths have to be evaluated, among other parameters. In the first case – when the lag lengths are predetermined – we use a Newton-Raphson optimization of the Beta-distributed weights and regression coefficients. To proceed with a flexible specification, we apply Gibbs sampling. As soon as posterior distribution of daily lengths is obtained, it enables us to construct beta-mixed distribution of daily lags, which appears to be more plausible. Thus, the daily weights of each explanatory variable are constructed as a Beta-mixture with three different lengths with the highest posterior probabilities.

Although the Bayesian technique has been widely applied since Albert and Chib's (1993) and Chib and Greenberg's (1995) seminal papers, its empirical applications in the MIDAS-field are not many. An example is Owyang (2009). Rodriguez and Puggioni (2010) suggest Bayesian approach for estimation of MIDAS- models. They argue this approach is helpful, first, in evaluation of the model space, i.e. the length of high-frequency lags of the explanatory variable, and second, in estimation of beta-distributed weights and regression coefficients. In the multivariate MIDAS where the lengths of daily lags of explanatory variables are unknown, optimization over the model space and over the parameter space is cumbersome. Alternatively, Bayesian framework takes advantage of "the promising lengths", identified by their higher posterior probabilities. Simulation of beta-distributed weights at various lag lengths attracted our particular attention. We contribute a Gibbs-sampling scheme enabling to estimate the most likely lag lengths of daily explanatory variables, the appropriate beta-parameters and regression coefficients at each time of forecast. This scheme builds on MCMC methods of parameterization of beta-distributions (Norets and Tang [2011]) and Bayesian model selection (Geweke [1994] and George and McCulloch [1993, 1997]).

The Newton-Raphson approach is described in section 3.2.1. The Bayesian approach is presented in detail in section 3.2.2. We show that both methods result in similar parametrization for some of the variables, but differ for others. In particular, the Beta-mix distributions have thicker tails and therefore larger weight on longer lags of the daily variables. Nonetheless, the regression coefficients are very similar for both methods.

### 3.2.1 The Maximum Likelihood approach

In order to get a first assessment of the set $(K, \theta_1, \theta_2)$ for each of the daily variables we checked the correlation between the monthly change in CPI and weekly lags of each of the daily explanatory variables, and chose by discretion a Beta distribution that generates a similar lag distribution. The resemblance between the theoretical Beta distributions and the empirical correlations[10] may be seen in Figures 1a and 1b. Of course, the correspondence between the two is not unique; nontheless, the general characteristics such as skweness or symmetry of the lag distribution are evident.

The correlation between the exchange rate and prices is hump-shaped and lasts about two months. The correlation of inflation expectations with prices deteriorates monotonically, fuel prices effect is maximized with some lag and other commodity price indices correlate with local prices with a relatively long lag. The Bank of Israel's interest rate correlates with prices in a close-to-uniform manner, as expected, due to the fact that it is changed only once a month.

Based on the empirical characteristics shown in figures 1a and 1b we choose the lag length of each of the explanatory variables (model space).Given the wave length, $K$, we estimate jointly the sets of $(\theta_1, \theta_2)$ for each of the daily variables, together with the aggregate coefficients ($\zeta_1, \beta_i$, and $\gamma$ in equation (1a)) - altogether 23 parameters. The optimization is achieved analytically using the Newton-Raphson algorithm to attain maximum likelihood.[11].

It should be noted that the information available and therefore the shape of the lag distribution depends on the point in time in the month we perform the forecast. As the change in the CPI for a certain month is published only on the $15^{th}$ of the following month, month $t$'s CPI may be forecast during the month itself and until the $14^{th}$ of the $t+1$ month. The parameters presented here are for estimations using information until the $17^{th}$ of the month of interest. In the next section we test for forecasts made on the $2^{nd}$ and $17^{th}$ day of month $t$, and on the $2nd$ day of the following month, when all the relevant information for the month of interest is already available. Table 2 presents the

[10] The correlation between the change in the monthly CPI and lags of the average weekly change in the daily variables was computed by using the rate of change in the CPI attributed to the middle week of each month using cubic interpolation of the monthly data to weekly data.

[11] When the day of forecast shifts to the beginning of the month, we had to restrict the Beta parameters to be positive.

shifts in the optimal Beta parameters $(\theta_1, \theta_2)$ according to the Newton-Raphson method, for different dates of prediction.

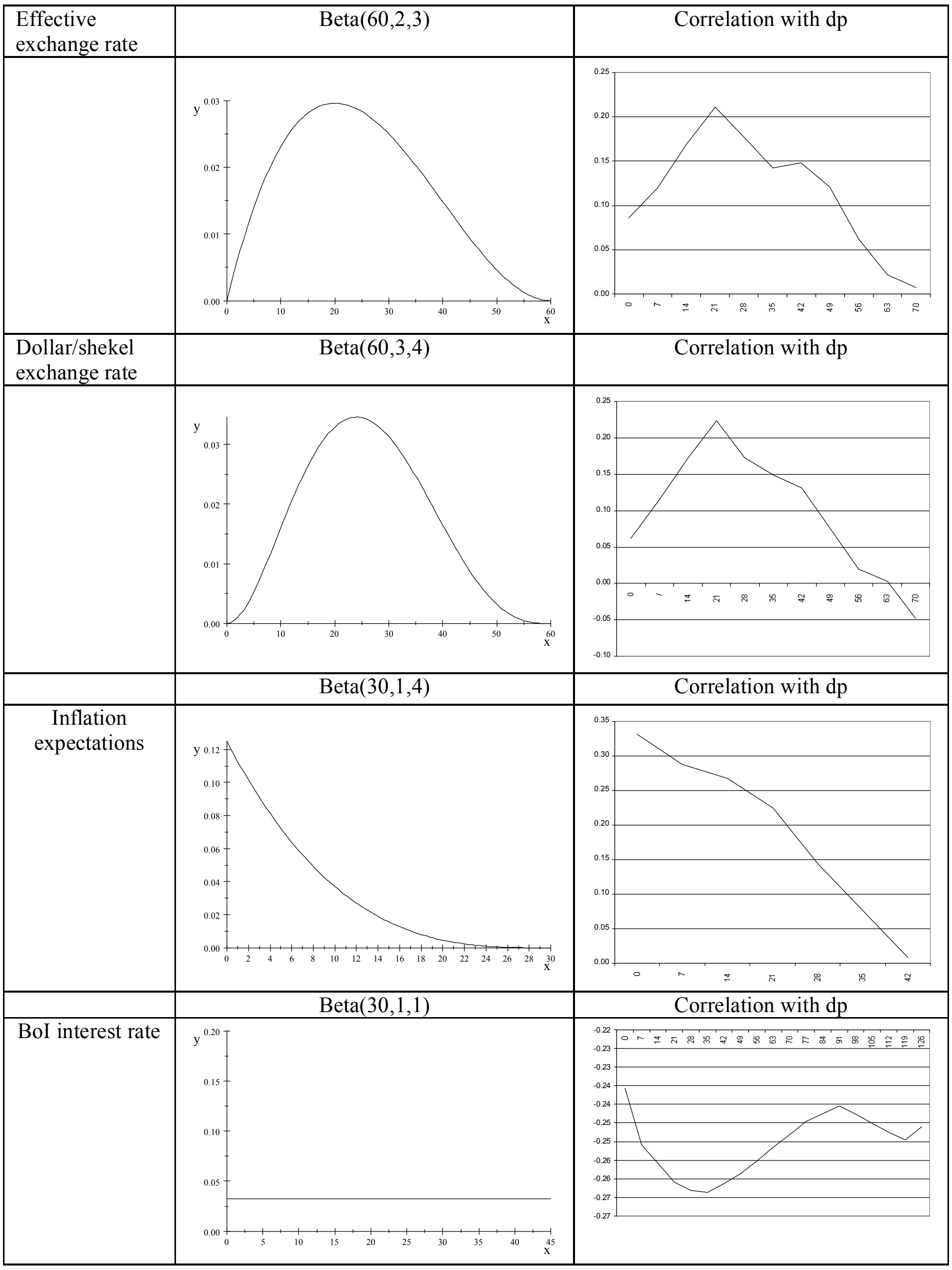



Figure 1a: the correlation between lagged explanatory variables and inflation, and the corresponding Beta distribution, for the 17th of the month of interest; local variables.

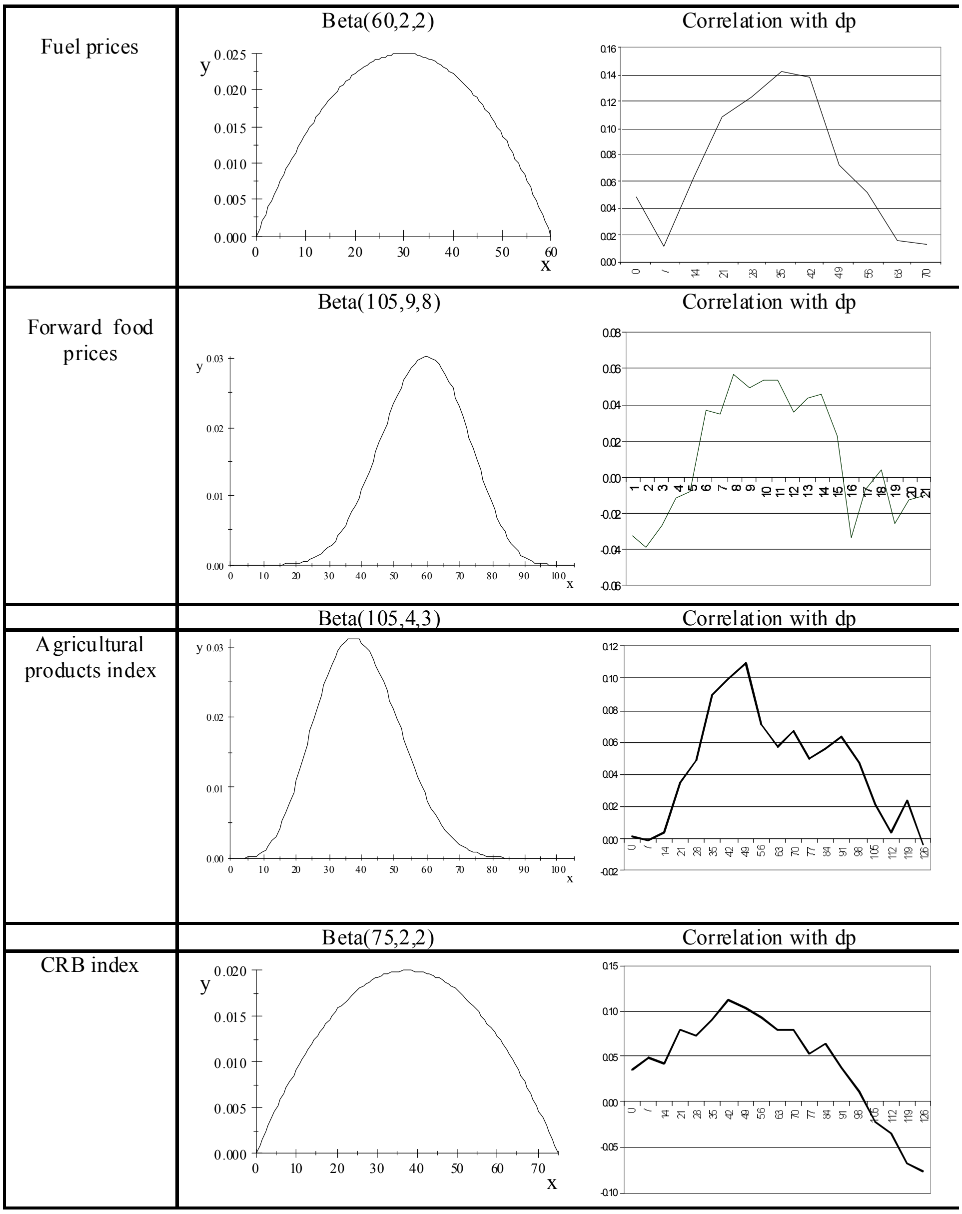



Figure 1b: The correlation between lagged explanatory variables and inflation, and the corresponding Beta distribution, for the 17th of the month of interest; global variables.

Table 2: Initial and optimal Beta parameters for different dates of forecasting

| Explanatory variables at daily frequency | K | Initial $(\theta_1, \theta_2)$ | Optimal parameters for specification with agricultural commodities | | |
|---|---|---|---|---|---|
| | | starting on 17 of the month in prediction | starting on 2 of the month in prediction[1)] | starting on 17 of the month in prediction[2)] | starting on 2 of the next month[2)] |
| Effective exchange rate, | 60 | (2, 3) | (3.8, 5.8) | (1.8, 2.5) | (1.8, 1.1) |
| $ exchange *contract ratio | 45 | (3, 4) | (1.7, 8.5) | (9.3, 17.4) | (21.5, 10.9) |
| fuel prices | 60 | (2, 2) | (6.7, 10.7) | (10.6, 6.2) | (11.6, 2.6) |
| Inflation expectations | 30 | (1, 4) | (0.1*, 19.2) | (6.9, 42.8) | (0.1*,42.6) |
| Agricultural commodities | 130 | (6, 10) | (2.9, 3.5) | (3.5, 3.8) | (5.5, 3.0) |

Notes:
* Constrained by the lower bound=0.1.
[1)] Running with an intercept (assuming AR(1) is unknown).
[2)] Running with AR(1) and no intercept.

### 3.2.2 Estimation according to the Bayesian approach

We proceed now to the flexible MIDAS-specification, which allows us to infer lengths $K_i$ from their posterior distributions, as all other parameters of the model. The estimation entails Gibbs-sampling with embedded Metropolis steps. The most likely lag lengths are identified by their more frequent appearance in the Gibbs sample and the parameters of beta-polynomials are derived conditionally on these lengths. Our sampling scheme is organized as follows.

We initialize uniform priors for daily weights of explanatory variables covering data from the past month until the time of forecast; that is, the polynomial weights are $Beta(45, 1, 1)$ if the forecast is made on the $15^{th}$day of the month. Given this, initial regression coefficients $(\widehat{\beta}_0)$ and variance $(\Sigma_0)$ can be easily obtained via the OLS.

Thus, we enter an iterative process, where each draw is conditional on the previous one. Each iteration simulates all parameters of equation (1) drawn in a particular sequence: the regression coefficients $(\widehat{\beta})$, the variance $(\sigma_\varepsilon^2)$, the lag lengths $(K)$ and the parameters of beta polynomials $(\theta)$. Note, that secondary order of draws has been imposed on lag lengths and beta-polynomials, with respect to multiple explanatory variables. The likelihood function, used within the Metropolis steps is:

$$L(\pi \mid \beta, K, \theta) = \exp\left[-\frac{1}{2\sigma_\varepsilon^2}\left(\pi - Z\widehat{\beta}\right)'\left(\pi - Z\widehat{\beta}\right)\right] \tag{4}$$

where $Z$ includes all explanatory variables, i.e. an autoregressive term, beta-weighted daily lags $(Xb)$, seasonal dummies and calendar shifts$(D)$ and $\widehat{\beta} = \{\alpha_1, \beta, \gamma\}$ includes all slopes.

Below are steps during the $(t+1)^{th}$ iteration of the sampler.

1. Drawing regression coefficients $\widehat{\beta}^{t+1}$conditional on the data and all parameters drawn at the $t^{th}$ iteration, i.e. $\widehat{\beta}^{(t+1)} \mid \pi, X, \sigma_\varepsilon^{(t)}, K^{(t)}, \theta^{(t)}$ :

$$\widehat{\beta} \sim N\left((\Sigma_0^{-1} + \sigma_\varepsilon^2 Z'Z)^{-1}(\widehat{\beta}_0\Sigma_0^{-1} + \sigma_\varepsilon^2 Z'Z)^{-1}, (\Sigma_0^{-1} + \sigma_\varepsilon^2 Z'Z)\right) \tag{5}$$

2. Drawing variance $\sigma_\varepsilon^{2(t+1)} \mid \pi, X, \widehat{\beta}^{(t+1)}, K^{(t)}, \theta^{(t)}$ from inverted Gamma:

$$\frac{1}{\sigma_\varepsilon^2} \sim Gamma\left(\frac{1}{T}, \frac{\varepsilon'\varepsilon}{2}\right) \tag{6}$$

where $T$ is the sample size (in months) and $\varepsilon'\varepsilon$ is the sum of squared residuals.

3. The length of the i$^{th}$ explanatory variable, $K_i^{(t+1)} \mid \pi, X, \widehat{\beta}^{(t+1)}, \sigma_\varepsilon^{2(t+1)}, K_1^{(t+1)}, \theta_1^{(t+1)}$, .. .., $K_{i-1}^{(t+1)}, \theta_{i-1}^{(t+1)}, K_{i+1}^{(t+1)}, \theta_{i+1}^{(t+1)}, K_N^{(t+1)}, \theta_N^{(t+1)}$ has been simulated as a contingent selection (Geweke, 1994), i.e. each specific lag is allowed only if all shorter lags of the same variable also enter. The lengths have been drawn sequentially (with a step of 5 or 10 days) over an interval, predetermined for each explanatory variable: for exchange rates – from 30 to 70 days; for inflation expectations – from 5 to 30 days; for agricultural commodities and fuel prices – from 30 to 120 days. The candidate length $K_i^*$considered by the corresponding Metropolis step is accepted with a transition probability , defined as follows:
(7)
$$\alpha_k = \min\left\{\frac{L(\pi \mid K_i^*, X, \beta^{(t+1)}, \sigma_\varepsilon^{(t+1)}, \theta_1^{(t+1)}, ., \theta_i^{(t+1)}, ., \theta_N^{(t+1)}, K_1^{(t+1)}, K_{i-1}^{(t+1)}, K_{i+1}^{(t+1)}, ., K_N^{(t+1)})}{L(\pi \mid K_i^{(t)}, X, \beta^{(t+1)}, \sigma_\varepsilon^{(t+1)}, \theta_1^{(t+1)}, ., \theta_i^{(t+1)}, ., \theta_N^{(t+1)}, K_1^{(t+1)}, K_{i-1}^{(t+1)}, K_{i+1}^{(t+1)}, ., K_N^{(t+1)})}, 1\right\},$$

4. Parameters of the beta distribution for the i$^{th}$ explanatory variable, i.e. $\theta_i^{(t+1)} \mid \pi, X, \widehat{\beta}^{(t+1)}, \sigma_\varepsilon^{2(t+1)}, K_1^{(t+1)}, \theta_1^{(t+1)}, ..., K_{i-1}^{(t+1)}, \theta_{i-1}^{(t+1)}, K_{i+1}^{(t+1)}, \theta_{i+1}^{(t+1)}, K_N^{(t+1)}, \theta_N^{(t+1)}$ have been derived conditionally on the regression coefficients and the variance, drawn during the current iteration, as well as on the beta-parameters of remaining explanatory variables, drawn previously, depending on the predetermined order of draws.[12] We follow Norets and Tang (2011) which pick parameters for beta-distributions accordingly with the method-of-moments estimators. Let the explanatory variable $x$ have $K$ daily lags (assigned at step 3), and change over the interval $[l, h]$ with the mean $ave(x)$ and the variance $var(x)$.[13] The method-of-moments estimates of beta-parameters $\theta = \{\theta_1, \theta_2\}$ are:

$$\theta_1 = ms \text{ and } \theta_2 = (1-m)s\ . \tag{8}$$

[12]This order does not affect posterior distributions, since a large portion of first draws has been removed.

[13]Here and henceforth the index $i$ of the explanatory varibles is omitted, for simplicity.

where $m = \frac{ave(x)-l}{h-l}$, $s = \frac{m(1-m)}{K\ var(m)} - 1$, $var(m) = \frac{var(x)}{K(h-l)^2}$. The values of $s^{(t+1)}$ and $m^{(t+1)}$ are drawn from $Gamma(u,v)$ and $Beta(g_1, g_2)$,respectively, while the transition densities $q_s$ and $q_m$ within the Metropolis step are defined as:

$$q_s \sim Gamma\left(u^{(t)} + u_0 - 1, \left(\frac{1}{v^{(t)}} + \frac{1}{v_0}\right)^{-1}\right) \tag{9}$$

$$q_m \sim Beta\left(g_1^{(t)} + g_1^{(0)} - 1, g_2^{(t)} + g_2^{(0)} - 1\right) \tag{10}$$

where $g_1^{(t)} = \left[\frac{m^{(t)}(1-m^{(t)})}{var(m)} - 1\right] m^{(t)}$, $g_2^{(t)} = \left[\frac{m^{(t)}(1-m^{(t)})}{var(m)} - 1\right] (1 - m^{(t)})$ ,
$u^{(t)} = \frac{(s^{(t)})^2}{var(s^{(t)})}$ , $v^{(t)} = \frac{var(s^{(t)})}{s^{(t)}}$, $var(s^{(t)}) = \frac{d^4 - \frac{(var(x))^2}{(h-l)^2}}{K\frac{(var(x))^4}{(h-l)^8}}(m^{(t)})^2(1-m^{(t)})^2$ ,$d^4 = \frac{\Sigma(x-ave(x))^4}{(h-l)^4 K}$
and the initial values are: $g_1^{(0)} = g_2^{(0)} = 2$, $u^{(0)} = 1.2$ and $v^{(0)} = 100$ , tuned to achieve reasonable acceptance rate. The candidate draw $s^*$ is accepted with probability

$$\alpha_s = \min\left\{\frac{L(\pi \mid s^*, m^{(t+1)}, X, \beta^{(t+1)}, \sigma_\varepsilon^{(t+1)}, K^{(t)})q_m(s^{(t)} \mid s^{(t+1)}, X, \beta^{(t+1)})}{L(\pi \mid s^{(t)}, m^{(t)}, X, \beta^{(t+1)}, \sigma_\varepsilon^{(t+1)}, K^{(t)})q_m(s^* \mid s^{(t+1)}, X, \beta^{(t+1)})}, 1\right\}, \tag{11}$$

The candidate draw $m^*$ is accepted with probability

$$\alpha_m = \min\left\{\frac{L(\pi \mid m^*, s^{(t+1)}, X, \beta^{(t+1)}, \sigma_\varepsilon^{(t+1)}, K^{(t)})q_m(m^{(t)} \mid s^{(t+1)}, X, \beta^{(t+1)})}{L(\pi \mid m^{(t)}, s^{(t)}, X, \beta^{(t+1)}, \sigma_\varepsilon^{(t+1)}, K^{(t)})q_m(m^* \mid s^{(t+1)}, X, \beta^{(t+1)})}, 1\right\}, \tag{12}$$

Further expansion of expressions for $\alpha_s$ and $\alpha_m$ should be done regarding the multivariate structure of the model. With regard to the i$^{th}$ explanatory variable, additional conditioning is required on parameters $K_1^{(t+1)}, \theta_1^{(t+1)}, ...., K_{i-1}^{(t+1)}, \theta_{i-1}^{(t+1)}, K_{i+1}^{(t+1)}$ $, \theta_{i+1}^{(t+1)}, K_N^{(t+1)}, \theta_N^{(t+1)}$ drawn at the current and previous iterations.

From a total of 1000 iterations the first 500 have been discarded and the subsequent 500 are used to obtain expected values of the parameters and their standard errors. Concerning posterior distributions of lag lengths, we identify the most likely ones as having the highest frequencies in the Gibbs sample. Thus, the posterior distributions of $\{s, m\}$ and corresponding expected values of $\theta$ are derived conditionally on these lengths. In the out-of-sample experiments, we apply a mixture of three beta-polynomials, computed upon the three most likely lag lengths of each explanatory variable.Then, the one-step-ahead forecast equation takes form:

$$\widehat{\pi}_t = \overline{\zeta}_1 \pi_{t-1} + \sum_{i=1}^{N} \overline{\beta} \frac{\sum_{j=1}^{3} \lambda_{ij} \sum_{k=1}^{K_{ij}} b(k; \overline{\theta}_{ij}) L^{\frac{k}{K_{ij}}} x_{i,t}^{(k)}}{(\lambda_1 + \lambda_2 + \lambda_3)} + \overline{\gamma} D_t \tag{13}$$

where $\overline{\zeta}_1$, $\overline{\beta}$, and $\overline{\gamma}$ are expected slopes. $K_{i1}$, $K_{i2}$, $K_{i3}$ are the most likely are lengths of the $i^{th}$ explanatory variable, having posterior probabilities $\lambda_1$, $\lambda_2$ and $\lambda_3$ and respectively. $\theta_{i1}$, $\theta_{i2}$, $\theta_{i3}$ are three pairs of beta-parameters, derived upon lengths $K_{i1}$, $K_{i2}$, $K_{i3}$ from the corresponding posterior distributions $\{s_{i1}, m_{i1}\}$, $\{s_{i2}, m_{i2}\}$and $\{s_{i3}, m_{i3}\}$.

## 3.3 Comparing methods

We compare the results of the estimation according to the two alternative approaches - the Bayesian approach and the Newton-Raphson procedure with the initial distribution based on empirical regularities.[14]

### 3.3.1 The Beta parameters

Table 3 presents the $(K, \theta_1, \theta_2)$ parameters of the daily distributions of each of the variables. For the Bayesian approach we present the beta-mixture, described in Section 3.2.2. The estimated parameters, the posterior probabilities of the lag length and the corresponding $\theta$ parameters, derived based on equation (8), are presented in Table A.2 in the Appendix.

Table 3 presents the Beta distributions according to both methods and shows that in some cases the Bayesian method and the Newton-Raphson method result in visually similar Beta distributions, but in other cases the Beta-mix procedure results in thicker tails and more weight on longer daily lags, possibly in a double-humped shape. The exchange rate - both the effective and the dollar exchange rate augmented for the share of dollar rent contracts - have the maximal effect within 20 to 30 days prior to the estimation date on the $17^{th}$ of the relevant month. The effect of the inflation expectations becomes much smoother, closer to uniform, according to the Beta-mix estimation. Fuel prices affect prices both immediately, probably directly through gazoline prices, and with a longer lag, representing indirect effects on energy prices like electricity. The agricultural commodities price index has a prolonged effect of 2 to 3 months prior to the month in interest, according to both methods.

### 3.3.2 The regression coefficients

Next, we compare the monthly estimation coeffiicients resulting from the two alternative methods and for the benchmark models. Table 4 shows the regression coefficients attained using the Newton-Raphson method, and the expected value of the coefficient, according to the distribution attained using the Bayesian method. Looking at the figures it is apparent that the values of the coefficients are generally very similar. The standard errors according to the Newton-Raphson optimization show that all variables have a significant effect on the estimated monthly change in the CPI.

[14]The daily distribution of the Bank of Israel interest rate is assumed to be uniform and is not optimized, as is the distribution of the other variables.

Table 3: Daily distribution parameters: According to Bayesian Mixed beta and Newton-Raphson methods, estimation on the 17th

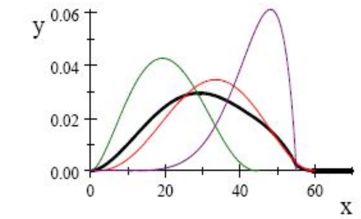
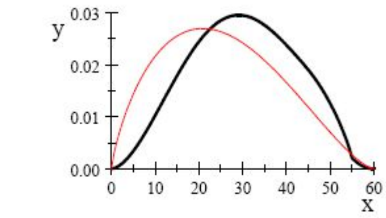
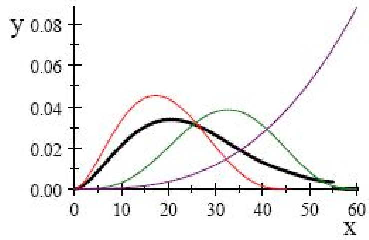
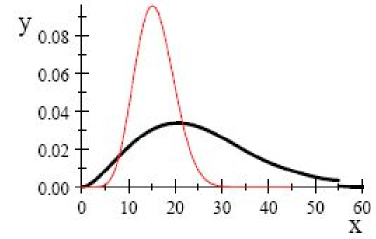
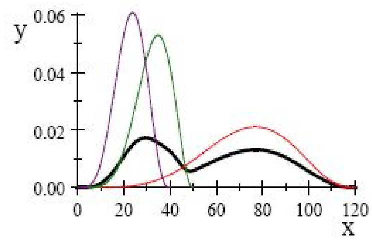
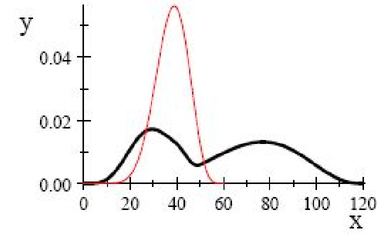
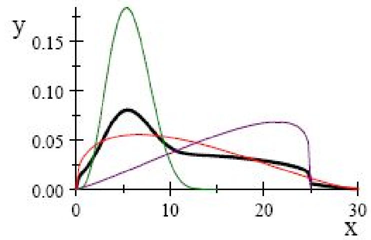
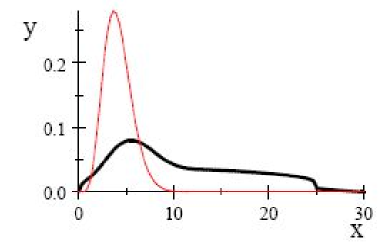
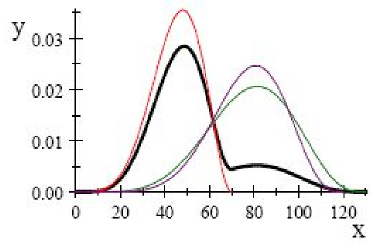
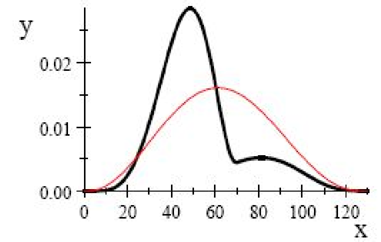
<table>
<tr><td></td><td>Beta Mix (Thick black)<br>and source distributions</td><td>Beta Mix (Thick black)<br>and NR estimation</td></tr>
<tr><td>NER</td><td>
</td><td>
</td></tr>
<tr><td>DOL</td><td>
</td><td>
</td></tr>
<tr><td>FUEL</td><td>
</td><td>
</td></tr>
<tr><td>INF_E</td><td>
</td><td>
</td></tr>
<tr><td>AGR</td><td>
</td><td>
</td></tr>
<tr><td colspan="3">NER=Effective exchange rate; DOL=Dollar/Shekel exchange rate;<br>FUEL=Fuel prices; INF_E=Inflalation expectations; AGR=agrc. prices</td></tr>
</table>

Table 4: Estimation results for Beta model and benchmark models.

| | Monthly data | UNIFORM | Newton-Raphson | Bayesian Beta-mix |
|---|---|---|---|---|
| | Coeff. (std. error) | Coeff. (std. error) | | Coeff. (std. error) |
| dp(-1) | 0.34 (0.08) | 0.25 (0.06) | 0.26 (0.05) | 0.24 (0.06) |
| Effective nominal exchange rate | 0.06 (0.02) | 2.26 (0.66) | 3.01 (0.44) | 2.07 (0.54) |
| Effective nominal exchange rate (-1) | -0.08 (0.02) | | | |
| Dollar exchange rate * contract ratio | 0.06 (0.03) | 3.71 (1.09) | 1.85 (0.50) | 3.53 (0.97) |
| (Dollar exchange rate * contract ratio)(-1) | 0.01 (0.03) | | | |
| Fuel price ($) | 0.014 (0.003) | 0.20 (0.11) | 0.26 (0.06) | 0.25 (0.12) |
| Fuel price ($)(-1) | 0.004 (0.004) | | | |
| Inflation expectations | 0.06 (0.03) | 0.10 (0.03) | 0.11 (0.02) | 0.10 (0.02) |
| Agricultural commodities index | 0.007 (0.007) | 1.12 (0.40) | 0.82 (0.27) | 0.77 (0.25) |
| Agricultural commodities index(-1) | 0.001 (0.007) | | | |
| Agricultural commodities index(-2) | -0.001 (0.007) | | | |
| BoI interest rate | -0.00 (0.00) | -0.03 (0.00) | -0.03 (0.01) | -0.02 (0.01) |
| Seas. dummies and holiday shift | YES | YES | YES | YES |
| Adjusted $R^2$ | 0.65 | 0.69 | 0.77 | 0.73 |
| RMSE in sample | 0.30 | 0.28 | 0.24 | 0.26 |

The tables presented display similar in-sample results using the two alternative methods. The Bayesian algorithm results in Beta distributions similar to these attained by the Newton-Raphson method, although the initial weights in the Bayesian approach are assigned uniformly, far from the empirical distribution set as initial values for the Newton-Raphson method. The coefficients for the aggregated data also display considerable similarities. According to these results, and considering the relative simplicity of the Newton-Raphson method relative to the Bayesian procedure, in particular for recurring out-of-sample exercises, we proceed with the Newton-Raphson method for further analysis of the forecasting characteristics of our model.

#### 3.3.3 Alternative specifications

The preferred specification was estimated for the log difference of the "original" CPI, adding seasonal dummies and holiday shift variables as explanatory variables. Alternatively, we also estimated (using the Newton-Raphson procedure) a version for the seasonally adjusted CPI, dropping the seasonal dummy variables from the estimation. According to the results for this estimation, shown in Table A.1 in the Appendix, it is apparent that the economic factors contribute a major share of the variation in the rate of change of the CPI, and seasonal factors make a smaller contribution to the explanatory power. Table A.1 shows that the adjusted $R^2$ remains relatively high, at about 50 percent, when estimating the log difference of the seasonally adjusted CPI with the daily financial data. As the explanatory daily variables are not seasonally adjusted, although there may be seasonal effects in their dynamics, we prefer the estimations relating to the log difference of the original CPI which includes seasonal dummy variables and holiday shifts among the explanatory variables.

#### 3.3.4 Benchmark models

The first measurement is a uniform distribution of the daily changes in the variables over different lag lengths, as specified in the table. The second is a model estimated using variables based on monthly data, i.e., the rate of change is calculated between average monthly levels of the variable of interest. This alternative is close to, but not identical with the first benchmark case (the average rate of change is not equal to the rate of change of the average). All estimations are assumed to be conducted on the $17^{th}$ of the relevant month, and include the lagged inflation rate. As seen in the table, the regression coefficients are similar for all methods. The preferred MIDAS specifications yield considerably higher $R^2$ than the other two specifications. The specification with the monthly data includes monthly lags which are comparable to the wave length in the Beta specifications. Therefore, agrcultural prices, for example, appear with the last full month available and 2 additional lags - 3 months in total, similar to a wave length of 130 days. The coefficients' magnitude is proportional to the length of the wave. We will explore the out-of-sample properties of these models in the next section.

# 4 Forecast Results

Based on the estimation results we test for the properties of the one-month and two-months-ahead forecasts attained using the MIDAS approach. In order to compare the models we proceed by estimating each of them starting from November 1999 to date $T$, and then forecasting, based on this estimation, the next month's unknown value of the change in CPI. We employ this procedure for $T$ ranging from December 2007 to August 2010. This means we have 33 one-month-ahead forecasts from January 2008 to September 2010. This period was selected because it was found to be the period for which the statistical model produced the largest RMSFE. For each subsample we re-optimize simultaneously the Beta parameteres together with the regression coefficients using the two alternative procedures - the Newton-Raphson algoritm and the Bayesian approach.

## 4.1 Comparing specifications

After compiling the series of forecasts, for each alternaive we compute the RMSFE and the MAFE. As mentioned above, because the MIDAS estimation is based on daily data, in order to forecast the rate of change in the CPI in month $T + 1$, which is published only on the $15^{th}$ day of the next month, the daily data we have available depends on the timing of the forecast. Therefore, we check for the quality of the MIDAS-based forecasts for forecasts formed on the $2^{nd}$ of the current month, when we still do not have the information about the CPI of the past month; on the $17^{th}$ of the month of interest, when we have two weeks of daily information and data for the past month's CPI, and on the $2^{nd}$ day of the following month, when all relevant daily data for the forecast month is already available. Table 5 presents the RMSFE and MAFE for the MIDAS model for alternative forecast dates, with or without the AR term (rows 0 to 4) and for the uniform distribution and monthly data benchmark models for different dates of forecasting (rows 5 to 10).

We show that the forecast based on the Beta-mixture procedure performs somewhat better than the Newton-Raphson approach and that he best performing model according to these tests is the MIDAS model for forecasts executed on the $17^{th}$ of the month of interest (rows 0 and 1 in the table). The one-month-ahead forecasts together with the actual values, are presented in Figure 2.

Next we compare the results for different forecasting dates. Moving from the $2^{nd}$ of month $T$ to the $17^{th}$ of that month, two additional types of information are made available. The first is additional daily information for the two weeks of the month, and the second is the CPI of month $T - 1$ available on the $15^{th}$ of each month. We check the significance of these contributions by comparing the difference in the quality of forecasts between the $2^{nd}$ and the $17^{th}$, both without the AR component (rows 2 and 3), exhibiting the additional daily information, and then comparing the forecasts made on the $17^{th}$ with or without the AR component, representing the contribution of the information about the lagged CPI (rows 3 and 1). The RMSFE and MAFE improve on

Table 5: RMSFE and MAFE for different specifications of the MIDAS model and for monthly averages, out-of-sample one month ahead forecasts forecast 2008.1-2010.9

| | **Model** | **Time of forecast** | **AR** | **RMSFE** | **MAFE** |
|---|---|---|---|---|---|
| 0 | MIDAS - Beta mixture | $17^{th}$ of month $T$ | yes | **.258** | **.213** |
| 1 | MIDAS - Newton Raphson | $17^{th}$ of month $T$ | yes | **.275** | **.219** |
| 2 | | $2^{nd}$ of month $T$ | no | .330 | .260 |
| 3 | | $17^{nd}$ of month $T$ | no | .290 | .243 |
| 4 | | $2^{nd}$ of month $T+1$ | yes | .280 | .222 |
| 5 | Uniform dist. | $2^{nd}$ of month $T$ | no | .342 | .289 |
| 6 | | $17^{th}$ of month $T$ | yes | .325 | .271 |
| 7 | | $2^{nd}$ of month $T+1$ | yes | .314 | .234 |
| 8 | Monthly data | $2^{nd}$ of month $T$ | no | .361 | .287 |
| 9 | | $17^{th}$ of month $T$ | yes | .318 | .267 |
| 10 | | $2^{nd}$ of month $T+1$ | yes | .283 | .220 |

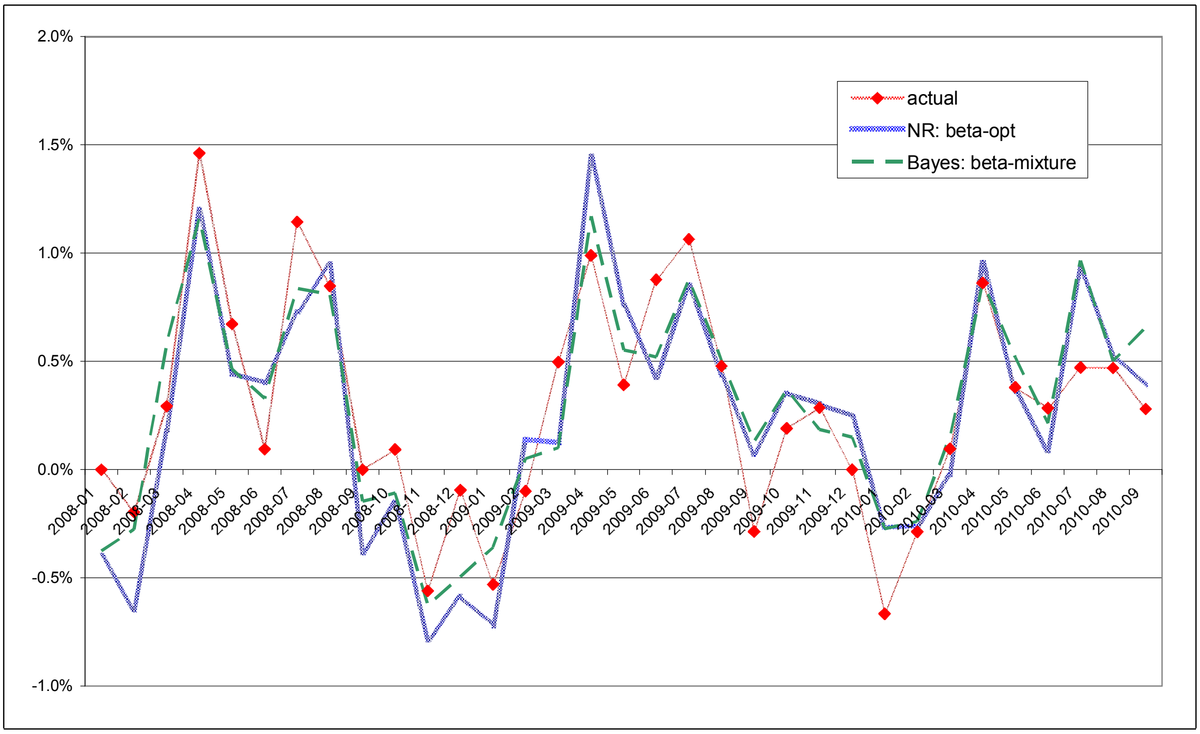
both stages, expressing the contribution of both components to a more accurate forecast. Waiting until the beginning of the following month and gaining additional daily data for the second half of the month does not improve the forecast (rows 4 and 1), meaning that the information embedded in the first half of the month is the most relevant for forecasting price changes in the same month. For the two benchmark models we find that the best performance is achieved by waiting until the beginning of month $T + 1$ (rows 7 and 10).

Comparing the three models, given the forecasting date, we find that the Beta-distributed MIDAS is always superior to the model with uniformly distributed daily data and the monthly model. The performance of the MIDAS model is similar to that of the monthly data model only when the forecast is executed at the beginning of the next month (rows 4 and 10). The forecast based on the uniform distribution, using the preferred specification and timing of the Beta model, perform worse than any of the MIDAS models, suggesting that the non-uniform lag distribution of the daily data contributes to the goodness-of-fit of the forecast. Comparing the uniform distribution of daily data with the forecasts based on monthly data, we find that the daily uniform model does not outperform the monthly model (rows 5-7 vs. 8-10). This is not surprising due to the fact that the difference between the average of daily changes (uniform daily data) and the average monthly change is not substantial.

## 4.2 Comparison with other models

The Research Department in the Bank of Israel refers to a number of models for short-term forecasts, mostly one month ahead, of the change in the CPI. The first is a statistical model (see Suhoy and Rotberger, 2006) which is based on estimating the trend, seasonal factors and additional exogenous effects for each of the main components of the CPI. Another existing model is a simple AR equation (Sorezcky, 2009). A third monthly model

Figure 2: Actual (red diamonds) and the one-month-ahead forecasted according to Newton-Raphson (blue) and Bayesian method (dashed green) monthly change in CPI, 2008.1-2010.9.



for short run inflation forecasting is an econometric model (Ilek, 2006) incorporating macroeconomic variables. The fourth model is a small BVAR model (Segal, 2010), which forecasts total CPI as a sum of two components - the housing component in the CPI and the CPI excluding this component. In addition, the Bank refers in its decisions to the average forecast made by local forecasters, most of them representing the economic units in large commercial banks or other financial institutions.

Table 6: RMSFE and MAFE for alternative models, out-of-sample forecast 2008.1-2010.9

| Model | RMSFE | MAFE |
|---|---|---|
| Statistical model | .473 | .372 |
| BVAR model | .375 | .295 |
| AR equation | .327 | .253 |
| Econometric model | .274 | .216 |
| Forecasters average | .259 | .191 |
| MIDAS - Newton Raphson | **.275** | **.219** |
| MIDAS - Beta mixture | **.258** | **.213** |

In Table 6 we present the RMSFE and MAFE for these models and for the MIDAS model. While the forecasts of the statistical model, the BVAR model and the AR equation were computed based solely on the output of the models, the comparison with the econometric model and the forecasters average was done based on the published forecasts at the time. These forecasts implicitly include the forcasters' discretion, including adjustments due to external information and professional judgement. Therefore, this comparison is biased in favor of these models. Even so, according to the table, the MIDAS model, and in particular its Beta-mixture version, preforms at least as well as all other models. It does much better than the statistical model, the BVAR model and the AR equation. According to the RMSFE, it is also better than the econometric model.

For two of the alternative models - the statistical model and the AR equation - we compared the quality of the one-month-ahead forecast for a later period, starting from January 2010 to March 2011. The results in Table 7 show that for the later period all models do better and that the MIDAS model is still better than the other two models that were checked.

Table 7: RMSFE and MAFE for alternative models, out-of-sample forecast 2010.1-2011.3

| Model | RMSFE | MAFE |
|---|---|---|
| Statistical model | 0.285 | 0.190 |
| AR equation | 0.290 | 0.219 |
| MIDAS Newton Raphson | **0.224** | **0.180** |

## 4.3 Significance of improvement

Although the comparison of the RMSFE of the models shows the superiority of the MIDAS forecast on the $17^{th}$, we checked the improvement in forecasting accuracy more carefully using the test proposed by Harvey, Leybourne and Newborne (1997) for comparison of forecasts by observation. The results are presented in Table 8.[15] The first panel of the table shows that the MIDAS model[16] is significantly better than the uniform distribution model - asserting the advantage of the distributed lags of daily data over the simple uniform distribution. The MIDAS model also yields better forecasts than those using the monthly data, though not significantly better. The second panel analyses the contribution of the daily data and the known CPI to the improved performance of forecasts performed on the $17^{th}$ vs. forecasts made on the $2^{nd}$ of the same month. Part IIA shows that moving from the $2^{nd}$ to the $17^{th}$ significantly improves the forecast error. The next two parts of the table, IIB and IIC, investigate the contribution of two types of additional information. None of the two components contributes substantially more than the other and both of them werre found to be insignificant. Nonetheless, the contribution of the additional daily information is "almost" significant, while the additional CPI information is not significant. This is important for the MIDAS modelling, affirming the importance of the daily information.

Table 8: Test for significance of improvement in RMSFE, 2008.1-2010.9.

| | **Test value** | **Probability** |
|---|---|---|
| I. $\mathbf{H_0}$: MIDAS* on the $17^{\mathbf{th}}$ is better than benchmark models | | |
| Uniform distribution (on $17^{th}$ of month $t$) | -1.93 | 0.05 |
| Monthly data | -1.15 | 0.26 |
| | | |
| IIA. $\mathbf{H_0}$: MIDAS on the $17^{\mathbf{th}}$ *with AR* is better than MIDAS on the $2^{\mathbf{th}}$ *without AR* | | |
| | -1.73 | 0.09 |
| IIB. $H_{\mathbf{0}}$*:* MIDAS on the $17^{\mathbf{th}}$ *without AR* is better than MIDAS on the $2^{\mathbf{th}}$ *without AR* | | |
| | -1.18 | 0.25 |
| IIC. $H_{\mathbf{0}}$*:* MIDAS on the $17^{\mathbf{th}}$ *with AR* is better than MIDAS on the $17^{\mathbf{th}}$ *without AR* | | |
| | -0.21 | 0.84 |
| III. $\mathbf{H_0}$: MIDAS on the $17^{\mathbf{th}}$ is better than other models | | |
| Statistical model | -3.16 | 0.00 |
| BVAR model | -1.95 | 0.06 |
| AR model | -0.92 | 0.36 |
| econometric model | 0.24 | 0.81 |
| Forecasters average | 0.57 | 0.58 |
| * In this table we refer to MIDAS estimated using the Newton Raphson method. | | |

[15] We did not perform the test for the later period 2010.1-2011.3, presented in Table 7.

[16] We compare other models to the Newton-Raphson version of the MIDAS model. As shown in Table 6, the Bayesian version performs better.

Panel III of the table compares the MIDAS model and the other models presented in Table 6. The advantage of the preferred specification of the MIDAS is evident.[17] It should be mentioned that in the period tested in this table, the statistical model did not include any macroeconomic variables as part of the concept that differentiates this model as a statistical model that estimates each of the main components of the CPI independently, from existing models in the Bank that investigate the aggregate CPI with macroeconomic explanatory variables. Table 7 shows the improvement in the statistical model's performance in the later period due to the inclusion of some macroeconomic variables such as the inflation expectations and the rent to house price ratio in the model. The MIDAS forecast has a significantly smaller error than the BVAR model and does better than the AR model, although not significantly. It does not do better than the econometric model or the average forecast of the outside forecasters.

### 4.4 Two months ahead forecast

We compare the MIDAS forecasting properties with those of the other models, for 2-months-ahead forecasts. The results are presented in Tables 9 and 10. The MIDAS preforms significantly better that the statistical model, better, but significantly, than the AR equation and the econometric model, but the forecasters' average outperforms the MIDAS' forecasts.

Table 9: RMSFE and MAFE for alternative models, out-of-sample forecast, two months ahead, 2008.1-2010.9

| Model | RMSFE | MAFE |
|---|---|---|
| Statistical model | .42 | .35 |
| AR equation | .46 | .35 |
| Econometric model | .46 | .35 |
| Forecasters average | .24 | .18 |
| MIDAS - Newton Raphson | **.28** | **.22** |
| MIDAS - Beta mixture | **.26** | **.21** |

[17] Both the uniform specification and the monthly specification of the MIDAS equation perform significantly better that the statistical model.

Table 10: Test for significance of improvement in RMSFE, 2 months ahead forecast, 2008.1-2010.9.

| | **Test value** | **Probability** |
|---|---|---|
| $H_0$: MIDAS* on the $17^{th}$ is better than other models | | |
| Statistical model | -1.68 | 0.10 |
| AR model | -0.41 | 0.68 |
| Econometric model | -0.38 | 0.72 |
| Forecasters average | 3.02 | 0.00 |
| * In this table we refer to MIDAS estimated using the Newton Raphson method. | | |

# 5 Concluding Remarks

The paper offers an additional method for forecasting short-term inflation, in particular forecasting the one-month-ahead change in the CPI. We use the approach suggested by Ghysels, Santa-Clara and Valkanov (2002), and followed by many others, known as MIDAS - MIxed DAta Sampling - which allows us to mix data of different frequencies in a relatively simple manner in terms of specification and estimation. We estimate a monthly model which incorporates daily data for the exchange rate and the interest rate, market-based inflation expectations and commodity price indices.

Our results show that the MIDAS model, incorporating daily explanatory variables with Beta-distributed lags of varying lengths, improves the forecasting performance as measured by the RMSFE and MAFE, relative to a benchmark monthly average model. We also find that the preferred timing to perform the forecast is in the third week of the forecast month. The first two weeks of daily data and information about the previous month's CPI, both contribute to the improvement of the forecast accuracy. The addition of the two last weeks of the month does not contribute to the performance of the model.

Comparing our results with those of other available monthly models in the Bank of Israel, we find that the MIDAS model performs as well as some of the models and significantly better than others. The model we present joins the suite of monthly models currently in use in the Research Department of the Bank of Israel, and as such is expected to contribute to our ability to understand the inflation environment and project its short-term development as part of the framework for conducting monetary policy.

Several further extensions of this model are possible. One extension is to estimate each of the CPI's major components separately, using MIDAS for the particular components, such as energy or food that may have higher dependence on the commodity indices, and complementing that with macro-financial data such as the interest rates and inflation expectations for a measure of "core" inflation. Factor analysis in the model (as suggested by Andreou, E., E. Ghysels and A. Kourtellos, 2010b and by Modugno, 2011) is also an interesting avenue to explore.

## References


[1] Albert, J.H. and S. Chib (1993) "Bayesian Analysis of Binary and Polychotomous Response," *Journal of the American Statistical Association*, Vol. 88, No.422, p.669-679

[2] Almon, S. (1965), "The distributed lag between capital appropriations and expenditures," *Econometrica*, 33, pp. 178-196.

[3] Andreou, E., E. Ghysels and A. Kourtellos (2010a), Forecasting with mixed-frequency data, prepared for a chapter in *Oxford Handbook on Economic Forecasting*, ed. Clements, M. P. and D. Hendry.

[4] Andreou, E., E. Ghysels and A. Kourtellos, (2010b), "Should macroeconomic forecasters use daily financial data and how ?" Working paper, February 2010.

[5] Armesto, M. T., K. M. Engermann and M. T. Owyang (2010), Forecasting with mixed frequencies, *Federal Reserve Bank of St. Louis Review*, Nov./Dec. 2010, pp. 521-536.

[6] Chen, X. and E. Ghysels (2011), News - good or bad - and its impact on predicting future volatility, *Review of Financial Studies*, 24(1), pp. 46-81.

[7] Chib,S. and E. Greenberg (1995) Understanding the Metropolis Hasting Algorithm – *The American Statistician*, November 1995, Vol 49, No 4, pp..327-335.

[8] George, E.I. and McCulloch, R.E. (1993) " Variable Selection Via Gibbs Sampling", *Journal of the American Statistical Association*, Vol.88, No. 423, p.p.881-889.

[9] George, E.I. and McCulloch, R.E. (1997) "Approaches for Bayesian Variable Selection", *Statistica Sinica*, 7, p.p.339-373

[10] Ghysels, E., P. Santa-Clara and R. Valkanov (2002), "The MIDAS Touch: Mixed data sampling regression models," UNC and UCLA Working Paper.

[11] Ghysels, E. and J. H. Wright (2009), Forecasting Professional Forecasters, *Journal of Business and Economic Statistics*, 27(4), pp. 504-516.

[12] Harvey, D., S. Leybourne and P. Newborne (1997), Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, 13, pp. 281-291.

[13] Ilek, A. (2006), "A Monthly Model for Evaluation of Inflation and Monetary Policy in Israel," Bank of Israel Monetary Studies Discussion Paper.

[14] Kotze, K. L. (2005), "Forecasting inflation with high frequency asset price data," working paper, University of Stellenbosch.

[15] Kuzin, V., M. Marcellino and C. Schumacher (2009), "MIDAS vs. Mixed frequency VAR: Nowcasting GDP in the Euro area", CEPR Working Paper, No. 7445, Sep. 2009.

[16] Modugno, M. (2011), "Nowcasting inflation using high frequency data," ECB Working Paper Series, No. 1324, April 2011.

[17] Monteforte L. and G. Moretti (2010), "Real time forecasts of inflation: the role of financial variables," Banca D'Italia Working Papers, No. 767, July 2010.

[18] Norets, A. and X. Tang (2010) MCMC estimation of a finite Beta mixture – URL: http://www.princeton.edu/~anotets/BetaMixture.pdf

[19] Owyang, M. (2009) "Forecasting Recessions Using the Mixed Frequency Probit",Federal Reserve Bank of St. Louis – URL: https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SNDE2010&paper_id=107

[20] Rodriguez, A. and G. Puggioni (2011), Bayesian model selection approaches to MIDAS regression – URL: http://www.soe.ucsc.edu/research/report?ID=1536

[21] Segal, G. (2010), "A BVAR model of the Israeli economy," Bank of Israel, memo.

[22] Sorezcky, A. (2009), "Forecasting inflation with autoregressive models," Bank of Israel, memo.

[23] Suhoy, T. (2010), "Monthly Assessments of Private Consumption," Bank of Israel discussion Papers, No. 2010.09, August 2010.

[24] Suhoy, T. and Y. Rotberger (2006), "Enhancing the Accuracy of the Seasonal Model for Forecasting CPI in the Short Run," Bank of Israel Discussion Papers, No. 2006.06, October 2006. (Hebrew).

[25] Tay, A. (2007), "Financial variables as predictors of real output growth", Singapore Management University Economics and Statistics Working Paper Series, no. 14-2007.

# 6 Appendix

Table A.1: Regression coefficients (estimated simultaneously with the Beta parameters), for the CPI seasonally adjusted rate

| | 1 | 2 | 3 |
|---|---|---|---|
| | Coeff. (std. error) | Coeff. (std. error) | Coeff. (std. error) |
| dp(-1) | 0.24 (0.07) | 0.22 (0.07) | 0.24 (0.07) |
| Effective nominal exchange rate | 2.44 (0.47) | 2.44 (0.48) | 2.44 (0.47) |
| Dollar exchange rate * contract ratio | 1.94 (0.53) | 2.09 (0.55) | 1.97 (0.54) |
| Fuel price ($) | 0.23 (0.07) | 0.18 (0.07) | 0.22 (0.08) |
| Inflation expectations | 0.10 (0.02) | 0.10 (0.02) | 0.10 (0.02) |
| Bloomberg Food index | 0.08 (0.22) | | |
| Agricultural commodities index | | 0.72 (0.29) | |
| CRB index | | | 0.11 (0.40) |
| BoI interest rate | -0.01 (0.01) | -0.02 (0.01) | -0.01 (0.01) |
| Seas. dummies and holiday shift | NO | NO | NO |
| Adjusted $R^2$ | 0.50 | 0.53 | 0.49 |
| RMSE in sample | 0.0028 | 0.0028 | 0.0028 |

Table A.2: The posterior proabailities of the Beta distribution parameters

| Lag length | Posterior prob. | s (std) | m (std) | $\theta_1$ | $\theta_2$ |
|---|---|---|---|---|---|
| **NER** | | | | | |
| **Lag length** | **Posterior prob.** | **s (std)** | **m (std)** | $\theta_1$ | $\theta_2$ |
| 60 | 63.6 | 7.2 (1.9) | 0.6 (0.1) | 3.9 | 3.3 |
| 45 | 26.0 | 6.3 (1.7) | 0.5 (0.1) | 2.8 | 3.4 |
| 55 | 8.2 | 9.2 (5.5) | 0.8 (0.0) | 7.3 | 1.9 |
| 50 | 2.2 | 3.6 (3.8) | 0.9 (0.0) | 3.1 | 0.6 |
| **DOL** | | | | | |
| **Lag length** | **Posterior prob.** | **s (std)** | **m (std)** | $\theta_1$ | $\theta_2$ |
| 45 | 60.2 | 6.7 (2.0) | 0.4 (0.2) | 2.8 | 3.9 |
| 60 | 35.4 | 8.8 (2.6) | 0.5 (0.1) | 4.7 | 4.1 |
| 55 | 4.4 | 4.4 (4.5) | 0.9 (0.1) | 3.9 | 0.5 |
| **FUEL** | | | | | |
| **Lag length** | **Posterior prob.** | **s (std)** | **m (std)** | $\theta_1$ | $\theta_2$ |
| 120 | 56 | 9.7 (2.1) | 0.6 (0.1) | 6.0 | 3.8 |
| 50 | 21 | 9.9 (2.1) | 0.7 (0.1) | 6.5 | 3.4 |
| 40 | 13 | 9.4 (2.0) | 0.6 (0.2) | 5.4 | 4.0 |
| 110 | 7 | 3.9 (0.8) | 0.8 (0.0) | 10.2 | 2.7 |
| 90 | 3 | 3.5 (0.5) | 0.5 (0.0) | 6.3 | 6.1 |
| **INF_E** | | | | | |
| **Lag length** | **Posterior prob.** | **s (std)** | **m (std)** | $\theta_1$ | $\theta_2$ |
| 30 | 51.2 | 3.8 (1.5) | 0.4 (0.1) | 1.4 | 2.4 |
| 15 | 26.0 | 11.4 (1.0) | 0.4 (0.1) | 4.4 | 7.1 |
| 25 | 22.6 | 3.4 (0.6) | 0.6 (0.1) | 2.2 | 1.2 |
| 20 | 0.2 | 0.8 (.) | 0.9 (.) | 0.7 | 0.1 |
| **AGR** | | | | | |
| **Lag length** | **Posterior prob.** | **s (std)** | **m (std)** | $\theta_1$ | $\theta_2$ |
| 70 | 73.0 | 9.1 (2.2) | 0.6 (0.1) | 5.8 | 3.2 |
| 130 | 16.6 | 11.1 (2.1) | 0.6 (0.0) | 6.7 | 4.4 |
| 120 | 6.2 | 12.7 (3.7) | 0.7 (0.0) | 8.2 | 4.5 |
| 100 | 2.4 | 7.8 (4.4) | 0.8 (0.0) | 6.2 | 1.6 |
| 80 | 1.0 | 8.2 (0.1) | 0.3 (0.0) | 2.6 | 5.7 |
| 11 | 0.8 | 9.2 (3.9) | 0.7 (0.0) | 6.7 | 2.5 |
| NER=Effective exchange rate; DOL=Dollar/Shekel exchange rate; FUEL=Fuel prices INF_E=Inflation expectations; AGR=Agricultural commodities' price index. | | | | | |