Bank of Israel **Research Department**

The Weighting of the Bank of Israel CPI Forecast—a Unified Model[1]

Dana Orfaig (Flikier)*

Discussion Paper 2013.08
July 2013

Bank of Israel, http://www.boi.org.il

[1] The weighted inflation forecast used in the unified model presented in this study has been presented to the Monetary Committee on a monthly basis since January 2013.

* Research Department, Dana Flikier: Dana.Flikier@boi.org.il, Tel. (9722) 6552634.

I would like to thank Edward Offenbacher, Sigal Ribon, Alon Binyamini, Itamar Caspi, Yossi Yakhin, Tanya Suhoy, and Nadav Steinberg for their helpful suggestions.

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

The Weighting of the Bank of Israel CPI Forecast—a Unified Model

Dana Orfaig (Flikier)

The Bank of Israel Research Department's inflation forecast is an important input in the formulation of monetary policy. The one month ahead forecast is currently based on a simple average of five models that project the change in the Consumer Price Index for the upcoming month. This paper proposes to weight the projections of the models differentially, and to assign one of them a negative weight. The new method enhances the precision of the forecast—an improvement of 50 percent in mean squared errors of out of sample tests.

The results also indicate that including the model which was assigned a negative weight is preferable to removing it from the weighting. The paper presents a theoretical basis that illustrates the inherent benefit of using negative weighting, which derives from utilizing the positive correlation among model errors. Furthermore, professional literature and empirical results which support the theoretical basis are presented.

שקלול תחזית בנק ישראל למדד המחירים לצרכן - מודל מאחד

דנה פליקר

תקציר

תחזית האינפלציה שחטיבת המחקר עורכת הינה רכיב חשוב בעיצוב המדיניות המוניטרית. התחזית לחודש אחד קדימה מתבססת כיום על ממוצע פשוט של חמישה מודלים החוזים מהו השינוי שיחול במדד המחירים לצרכן בחודש הקרוב. מחקר זה מציע לשקלל את תחזיות המודלים באופן שונה ולהעניק לאחד מהם משקל שלילי. התוצאות מראות כי כאשר עוברים משיטת השקלול הקיימת לשיטה החדשה, מוצאים שדיוק התחזית משתפר באופן משמעותי (שיפור של כ-50% בממוצע ריבועי הסטיות במבחני out of sample). עוד עולה מהתוצאות שהכללת המודל שקיבל משקל שלילי עדיפה על השמטתו מהשקלול. המחקר מציג בסיס תיאורטי הממחיש את התועלת הגלומה בשימוש במשקל שלילי, הנובעת מניצול המתאם החיובי בין סטיות המודלים. בהמשך הוא מציג ספרות מקצועית ותוצאות אמפיריות התומכות בבסיס התיאורטי.

1. Introduction

1.1 Theory

The Bank of Israel Research Department's one-month ahead inflation forecast is currently based on a simple mean of five models that predict the change in the CPI in the coming month.[1] This study seeks the optimal method of weighting the forecasts of the models—the weighting method that minimizes the errors, measured in terms of mean squared error (MSE), from the actual CPI.

To accomplish this, a methodology based on Modern Portfolio Theory (MPT) is used. According to MPT, an "efficient frontier" may be constructed given information about the distribution of asset returns and by exploiting the correlation among them. As a simple example, if asset A and asset B are negatively correlated, one would expect, in the event of an exogenous shock, the value of asset B to fall if that of asset A rises (and vice versa). In other words, due to the offset between the returns on those assets, a portfolio that includes both assets will be less risky (have lower variance) than one composed of uncorrelated assets or one asset only. If the assets are positively correlated, the level of risk may be mitigated by assigning a negative weight to one of the assets. (In practice, this is done by opening a short position in that asset.) In the event of an exogenous shock, one expects both assets to decline (or rise) in value concurrently. If asset values decline, an investor loses on asset A and gains on asset B, which s/he has sold short and which has fallen in value. Therefore, when a positive correlation exists, a less risky portfolio (i.e., one with smaller variance) can be constructed by using a negative weight.

Similar reasoning applies to weighting the inflation forecasts. We are interested in maximizing the quality of the unified forecast—that is, minimizing the deviations of the unified forecast from the actual CPI. This can be accomplished by assigning weights utilizing the information implicit in the correlation among the errors of the forecasts, much as risk is mitigated in an asset portfolio.

Specifically, if there is a negative correlation among the forecast errors, then using positive weights will minimize the errors of the weighted (unified) model. For example, when calculating a mean (or any combination of positive weights) composed of an above-forecast error and a below-forecast error, it will yield a more accurate result (more closely approximating the actual CPI). Similarly, if there is a positive correlation between the forecast errors, the use of weights with opposite signs will deliver a result that better approximates the actual CPI.[2]

These conclusions can be expressed in mathematical form—the aim is to minimize the MSE of the unified model. For simplicity, a model for two forecasts is discussed first; further on, the general case is presented as well:

(1) $$\hat{e}_c = w_1\hat{e}_1 + w_2\hat{e}_2$$

[1] The models are shown in Appendix 1.

[2] A simple example will demonstrate this. Say that the two models' errors are positively correlated, that both are biased upward a given month—one by 0.1 and the other by 0.3—and that they predict a CPI of 0.6 and 0.8, respectively, while the actual CPI is 0.5. If one weights the models at 1.5 and (-0.5), respectively, one obtains a unified forecast of 0.5 (=1.5*0.6+(-0.5)*0.8), namely, the actual CPI. Utilizing the positive correlation between the errors of the models and using a negative weight offset the errors. Note that it is impossible to yield such an outcome using positive (or zero) weights.

(2) $E((w_1\hat{e}_1 + w_2\hat{e}_2)^2) = \text{var}(w_1\hat{e}_1 + w_2\hat{e}_2) + (E(w_1\hat{e}_1 + w_2\hat{e}_2))^2$

where:

$\hat{e}_c$	= errors of the unified model (random variable)
$\hat{e}_1$	= errors of Model/Forecast 1 (random variable)
$\hat{e}_2$	= errors of Model/Forecast 2 (random variable)
w_1	= weight for Model 1 errors
w_2	= weight for Model 2 errors
σ_c	= S.D. of $\hat{e}_c$
σ_1	= S.D. of $\hat{e}_1$
σ_2	= S.D. of $\hat{e}_2$
ρ_{12}	= correlation between $\hat{e}_1$ and $\hat{e}_2$.

Assuming that the unified model has no systematic error— $E(w_1\hat{e}_1 + w_2\hat{e}_2) = 0$ —[3] we get:

(3) $E((w_1\hat{e}_1 + w_2\hat{e}_2)^2) = \text{var}(w_1\hat{e}_1 + w_2\hat{e}_2)$

That is, given a two-model forecast, composed of Model 1 and Model 2, error variance needs to be minimized. This is shown below:[4]

(4) $$\text{var}(w_1\hat{e}_1 + w_2\hat{e}_2) = w_1^2 \text{var}(\hat{e}_1) + w_2^2 \text{var}(\hat{e}_2) + 2w_1w_2 \text{cov}(\hat{e}_1, \hat{e}_2) = w_1^2 \text{var}(\hat{e}_1) + w_2^2 \text{var}(\hat{e}_2) + 2w_1w_2\sigma_1\sigma_2\rho_{12}$$

One may see that when the correlation is negative and the weights are positive, the last expression becomes negative and the variance of the weighted forecast's errors decreases relative to a situation of correlation of zero.

Similarly, when the correlation is positive, the last expression becomes negative only when the weights have opposite signs. In this manner, the error variance in the weighted forecast declines relative to a situation in which the correlation equals zero (and certainly relative to a situation where positive weights are used).

The optimal weights obtained for the solution of a first-order condition (assuming that the weights sum to 1) are:

(5) $$w_1^* = \frac{\sigma_2^2 - \rho_{12}\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2} = \frac{\sigma_2(\sigma_2 - \rho_{12}\sigma_1)}{(\sigma_1 - \sigma_2)^2 + 2\sigma_1\sigma_2(1 - \rho_{12})}$$

(6) $$w_2^* = \frac{\sigma_1^2 - \rho_{12}\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2} = \frac{\sigma_1(\sigma_1 - \rho_{12}\sigma_2)}{(\sigma_1 - \sigma_2)^2 + 2\sigma_1\sigma_2(1 - \rho_{12})}$$

It is fairly straightforward to see that this yields the following conditions for the assignment of a negative weight:

[3] Stemming from the assumption that neither of the models has a systematic error.

[4] The weights cannot be resolved to zero because, assuming that the unified model is not systematically biased, the weights must sum to 1. This assumption is tested further on in this paper.

(7) $w_1 < 0$, when $\rho_{12} > \sigma_2 / \sigma_1$

(8) $w_2 < 0$, when $\rho_{12} > \sigma_1 / \sigma_2$

Wherever no correlation exists between the model errors, i.e., ρ_{12} = 0, the last expression representing the variance of the unified model zeros out and the following expression is obtained:

(9) $$\text{var}(w_1\hat{e}_1 + w_2\hat{e}_2) = w_1^2 \text{ var}(\hat{e}_1) + w_2^2 \text{ var}(\hat{e}_2)$$

The optimal weights obtained, respectively, are:

(10) $$w_2^* = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \; w_1^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

In such a case, the weights that minimize the variance of the weighted forecast's errors are determined mainly by the forecasting quality of each constituent model. That is, the smaller the error variance in a given model, the larger the weight that it will receive (relative to the other model). For the optimal solution to be a simple mean, two conditions must be satisfied: identical error variance in both models and no correlation between the errors.

It can also be seen that when the optimal weights are inserted into the variance function of the unified model's errors (the function that was minimized), the expression obtained is by necessity smaller than the error variance of each forecast separately. Consequently, it is better to weight multiple forecasts than to use an individual forecast even if the latter is extremely accurate.

Similarly, the general formula (for several models) for the description of error variance in the unified model is[5]:

(11) $$\sigma_c^2 = \sum_i w_i^2 \sigma_i^2 + \sum_i \sum_{i \neq j} 2 w_i w_j \sigma_i \sigma_j \rho_{ij}$$

The mathematical development presented above therefore demonstrates the preference of using several models over using an individual forecast (even if it is the most accurate). It also shows that when there is no correlation between the models' errors, the optimal weight that each model will receive depends on its forecasting quality. In contrast, when a correlation exists, the variance expression of the unified model's errors expands: inserted into it is an expression that may become negative as a function of the correlation between the errors and of the signs of the weights. Thus, the expression may reduce the expected mean squared errors of the unified model.

Given the proposition that the weights should have opposite signs (when the correlation between the errors is positive), the following question arises: which of them should receive negative weights and which should be assigned positive ones? To

[5] This formula is also used to calculate investment portfolio risk according to MPT, which was introduced in the previous subsection.

answer that, note that since the sum of the weights equals 1, the value of the sum of the positive weights will be greater—by 1—that the absolute value of the sum of the negative weights. Therefore, the negatively weighted models should have much greater errors to offset the errors of the positively weighted models, which carry much larger weights.[6] This means that the models that have the larger error variance should be assigned negative weights. Another way to answer this question is that given the existence of two models, the negative weights must be smaller than the positive weights in absolute (and squared) terms (so that the weights sum to 1). Since the variance of the unified model is composed, among other things, of the product of each model's error variance and the squares of its weights, then to minimize variance in the unified model it is better to multiply the model that has the larger variance by the squares of the smaller weights; this must by necessity be the negative weight. Therefore, generally speaking, one may assume that the model that has the larger errors will receive the negative sign.[7]

This conclusion is consistent with the conditions for negative weighting (Equations 7 and 8 above). Observation of these conditions shows that, given a positive correlation (and less than 1, by definition), the forecast that has the larger error variance will receive the negative weight.[8]

1.2 Literature review

There is extensive literature dealing with the combination of forecasts. Bates and Granger (1969) and Newbold and Granger (1974) laid the foundations in this field, showing that this technique is preferable to the use of an individual forecast because even a less-accurate individual forecast contains information that may improve the total forecast. They find optimal weights by minimizing MSE; when the individual forecasts are not biased, this is identical to minimizing the error variance in the total forecast.[9]

Dickinson (1975) offers an explanation for the use of negative weights and links their use to the correlation among the errors. His proof, much like that presented here, shows that when seeking optimal weights for two models that exhibit a (not weak) positive error correlation, one model will receive a positive weight and the other a negative one. Dickinson also sets out conditions that determine which of the models will receive which weight (positive/negative), resembling those in the previous subsection (Equations 7 and 8).

Reinmuth and Geurts (1979) note that the use of weights in contrast to a simple mean improves the results markedly; their explanation relates to the strong correlation between the forecast errors.

Winkler and Clemen (1992) present the sensitivity of the weights to a range of variables and show the probabilities of obtaining different weights. They note that when there is a strong positive correlation between the models' errors, there is a high probability of assigning a negative weighting, even if the variances of the models' errors are almost identical. (If no correlation exists, the optimal weighting would be a

---

[6] This is demonstrated in Footnote 3.
[7] A disclaimer to this is presented in the next footnote.
[8] The opposite—assigning the negative weight to the forecast that has the smaller error variance—is improbable. In certain cases, however, both models may receive positive weights; this happens when a positive correlation exists but is smaller than the ratio of the error variances.
[9] As presented in the previous subsection (Equation 3).

simple mean.) This study, too, tests the aforementioned issues by minimizing error variance of the total forecast.

Bunn (1985), examining several ways of weighting models, offers several rules of thumb that link model weighting methods with sample size. He proposes that at the initial stage, when the number of observations is smaller than five, the models are to be weighted by the simple mean method. When the number of observations exceeds five and the model errors exhibit different variances, the MSE method is to be used for weighting. When the number of observations is smaller than twenty, the method should be used on the assumption that the correlations between the models' errors zero out. Finally, when there are more than twenty observations, these correlations are presumably stable and the assumption can be loosened. In addition, Bunn presents conditions (identical to Equations 7 and 8 in section 1.1 above) for negative weighting.

Another study—Hoeting, Madigan, Raftery and Volinsky (1999)—deals with model selection, the exclusion of less-accurate models from the weighted forecast. The authors contend that model weighting is preferable to model selection because model selection entails the disregard of important sources of information and because a forecast based on several models is more stable.

Handbook of Economic Forecasting (2006) contains a chapter on combination of forecasts. It offers arguments in support of forecast weighting, e.g., it allows the use of all available information, avoids bias, and copes better with structural changes in the market. The chapter also draws a parallel between forecast weighting and asset portfolio diversification to minimize risks.[10] Relating to the negative weight, the writer emphasizes that this is not the same as claiming that the negatively weighted forecast is valueless; instead, under certain conditions,[11] given the correlations between the errors of the forecasts, these are the weights that MSE minimization yields (Timmermann, 2006).

A study published by the Bank of England claims that the use of several forecasts is preferable to the use of just one because individual forecasts may be biased; when all are used, their biases may cancel each other out. The authors also claim that many wrongly think it better to use the most accurate forecast (the one with the smallest variance). This technique, however, yields suboptimal results because it fails to take into account the correlation between the forecast errors (Kapetanios, Labhard and Price, 2008).

A similar study at the Bank of Israel (Blank, 2007) set out to find the optimal weighting of private analysts' inflation forecasts. In the study, the weights were subjected to sign restrictions (only non-negative weights were allowed). The results showed that in the sample considered, the forecast weighted according to the weights obtained by MSE minimization was not significantly different from the simple mean. This may have been due to the sign restrictions.

---

[10] Much like the theoretical analogy presented at the beginning of the Introduction.

[11] The conditions described in the chapter are identical to those presented in Equations 7 and 8 in the previous subsection.

2. Methodology

2.1 Data

When the Bank of Israel's Research Department formulates its one-month ahead CPI forecast, it uses a simple weighting (mean) of five models: econometric, statistical, single-equation, BVAR, and MIDAS. The data examined in the study relate to the period from October 2010 to July 2012 (twenty-two observations). Since the MIDAS model has been run regularly only since July 2011, it was forecast retroactively for the October 2010–June 2011 period with the help of vintage data.[12]

The models include the interventions that are conventionally added to outcomes in cases where external information exists—changes in government-supervised prices (electricity, water, etc.), an increase in the Value Added Tax rate, etc. The inclusion of these interventions in the weighting mirrors the way the models are actually used and allows the models to be compared more accurately. Moreover, the results of the models without the interventions are not systematically recorded, and it is likely that different interventions were made in different models during part of the period. Therefore, in any event, the possibility of testing the models net of the interventions does not exist.

2.2 Examining the weights

To find the optimal weights for the models, the accepted method, of minimizing the mean squared errors of the unified model from the change in the actual Consumer Price Index, was used. This is done by using a numerical solution and also by an OLS estimation, defining the models' forecasts as explanatory variables and changes in the actual CPI as the dependent variable.

The estimated model is:

$$y_t = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + e_t$$

The model is estimated without an intercept, it being assumed that the composite models have no systematic bias that would require the use of an intercept. This assumption will be tested further on in this paper.

The results are also examined in view of restrictions on the sign (non-negative) of the weight and/or their sum (sum to 1). Therefore, the comparison includes four combinations of restrictions:

1. restrictions on both the weights' sign and sum: $\beta_i \geq 0$, $\sum \beta_i = 1$;
2. restrictions on the weights' sign (without restriction on sum): $\beta_i \geq 0$;
3. restriction on the weights' sum (without restrictions on sign): $\sum \beta_i = 1$;
4. no restrictions whatsoever.

[12] The sample could not be enlarged because attempts to forecast the BVAR model retroactively proved unsuccessful: in retroactive forecasting over twenty-one months (January 2009–September 2010), sizable errors were obtained relative to the period examined, in which the model was run regularly. However, according to Bunn (1985), referenced in the literature review, twenty-two observations are a large enough sample for the MSE minimization method, allowing one to assume that the error correlations are stable.

3. Results

Table 1 indicates that the model without restrictions and the model including only a sum restriction yield weights that improve in-sample forecasting quality.[13] Square errors decline by 75 percent relative to the current method and by 65 percent relative to sign-restricted weights (Line 1). To quantify the absolute improvement—the extent of error reduction—root mean squared error (RMSE) data and mean absolute error (MAE) data (Lines 2 and 3, respectively) are presented. It may be seen that in the transition from simple weighting to unrestricted weights or weights including a restriction only on the sum, RMSE decreases by 0.1 (from 0.2 under the current method).

It may also be seen that the coefficient of the weighted forecast will not be biased relative to 1 (with actual CPI change as the dependent variable). In this case, the Bank of Israel's current weighted forecast (simple mean) receives a coefficient of 0.81 as against 1.02 in an unrestricted model and 0.97 in a model including a sum restriction only (Line 4). Note that the coefficient of the current Bank of Israel forecast shows that the forecast was biased due to the weighting method used. Notably, however, according to a Wald test, the coefficient (0.81) is not significantly different from 1.

Furthermore, the quality of fit (R^2) of the current Bank of Israel forecast is 0.53 as against 0.84 in an unrestricted model or in one to which only a sum restriction applies (Line 5).

Table 1. Weighting performance under various restrictions

	Simple mean	Sign and sum restrict-ions (Option 1)	Sign restrictions (Option 2)	Sum restriction (Option 3)	No restrictions (Option 4)
1. $\text{MSE} = \frac{\sum e_t^2}{n}$ *	0.04 (0.09)	0.03 (0.06)	0.03 (0.05)	0.01 (0.02)	0.01 (0.02)
.2 $\text{RMSE} = \sqrt{\text{MSE}}$ **	0.2	0.17	0.17	0.1	0.1
.3 $\text{MAE} = \frac{\sum \lvert e_t \rvert}{n}$ ***	0.15 (0.15)	0.12 (0.13)	0.12 (0.13)	0.10 (0.10)	0.09 (0.08)
4. Coefficient****	0.81 (0.17)	0.96 (0.16)	1.07 (0.18)	0.97 (0.09)	1.02 (0.10)
5. R^2	0.53	0.64	0.64	0.84	0.84

Standard deviations appear in parentheses.

* Mean Squared Error; ** Root Mean Squared Error; *** Mean Absolute Error;

**** OLS regression coefficient in a model with an intercept (the results without a intercept are similar). The dependent variable is the actual CPI; the explanatory variable is the forecast weighted in accordance with the optimal weights obtained under the restrictions.

[13] Notably, this is an ex post analysis. Therefore, an unrestricted model will by necessity provide a better explanation that any restricted model. An out-of-sample test is performed below.

Table 2 indicates that when sign restrictions are imposed, the weight of the single-equation model zeros out, causing the loss of information that may be helpful in forecasting. In this case, the weights obtained when the full set of models with sign restrictions are run (middle column) are identical to those obtained after the single-equation model (with no sign restriction) is omitted.[14] By inference, then, it is better to assign the single-equation model a negative weight than to omit it from the weighting because the omission causes MSE to increase considerably (Table 1, Line 1).[15]

Arguably, the estimation may suffer from a multicollinearity problem. Indeed, the correlations among the models' errors fall into a range of 0.56–0.83. However, the insignificance of estimates that is typical of multicollinearity is not evident in the results at all, as all coefficients—apart from the BVAR coefficient—are statistically significant at a high level (Table 2).

Table 2. Weights obtained under various restrictions[16]

	Simple mean	Sign and sum restrictions (Option 1)	Sign restrictions (Option 2)	Sum restriction (Option 3)	No restrictions (Option 4)
Econometric	0.2	0.12	0.11	**0.56** (0.03)	0.55
Statistical	0.2	0.46	0.33	**0.48** (0.00)	0.4
Single-equation	0.2	0	0	**-0.76** (0.00)	-0.74
BVAR	0.2	0.11	0.21	**0.22** (0.20)	0.27
MIDAS	0.2	0.31	0.21	**0.5** (0.00)	0.44
Summing of weights	1	1	0.87	**1**	0.92

The p-values from the estimate appear in parentheses.

3.1 Proposed weights

As Table 1 shows, all indicators show very minor differences between the use of weights without restrictions (Option 4) and the use of weights with a sum restriction only (Option 3).

A Wald test on the hypothesis that the coefficients in the no-restriction model sum to 1 shows that the hypothesis cannot be refuted (p-value: 0.41). Therefore, the weights proposed for use are restricted by sum only (shown in boldface in Table 2).[17]

---

[14] Positive weights were obtained for all four models.

[15] Similarly, the weights obtained under sum and sign restrictions (Column 2) are identical to the weighting obtained after omitting the single-equation model (with a sum restriction but without a sign restriction).

[16] The weights shown are those obtained by MSE minimization in a model with no intercept. Error variance minimization elicits very similar results.

The possibility of adding an intercept to the regression was also tested. Estimating the regression, it was found that the intercept is not significant (p-value: 0.76). Another test was for whether a lagged error variable should be added to the estimation. Here, a Durbin-Watson (D-W) test revealed no serial correlation. Therefore, I decided to use the weights proposed (in a model lacking both an intercept and lags).

The negative weight obtained for the single-equation model is consistent with the discussion in the Introduction, which found that one may assume negative weighting of the model with the highest error variance. Table 3, which presents descriptive statistics of the errors of the various models, shows error variances of 0.11 for the single-equation model and 0.04–0.06 for the other models.

Table 3. Descriptive statistics of model errors

	Econometric	**Statistical**	**Single-equation**	**BVAR**	**MIDAS**
Mean squared error	0.04	0.05	0.11	0.06	0.06
Median squared error	0.03	0.02	0.06	0.04	0.02
Maximum error value	0.59	0.50	0.99	0.69	0.60
Minimum error value	-0.21	-0.37	-0.41	-0.31	-0.56
Error variance	0.04	0.05	0.11	0.06	0.05
Error skewness	0.67	0.31	0.70	0.57	0.04
Error kurtosis	3.28	2.56	3.74	3.27	3.31

3.2 Models' error correlations

From the theory presented in the Introduction, it follows that if the errors of the models are positively correlated, the use of a negative weight will reduce errors of the weighted forecast and thus improve forecast accuracy. It was shown above that the use of a negative weight does, in fact, reduce errors of the weighted forecast. Now let us show that this decrease traces to the positive correlations among the errors of the models. This will support the argument that the positive correlations and the use of the negative weight are what caused the decrease. As Table 4 shows, there are strong positive correlations among the errors of the models, particularly between the single-equation model and the other models.

The especially strong correlations between the errors of the single-equation model and those of the other models are consistent with the fact that the single-equation model received the negative weight. On the basis of the equations that determine which model should receive a negative weight (Equations 7 and 8), the stronger the error correlation among the models is and the wider the spread of the variances are, the greater the probability of being assigned a negative weight.[18]

---

[17] Theoretically, too, one would expect the weights to sum to 1 due to the assumption that the models are not systematically biased.

[18] Even though the equations were developed for a two-model weighted forecast, it is helpful to use them in understanding the general mechanism for the assignment of a negative weight.

Table 4. Error correlations among the models

	Statistical	MIDAS	BVAR	Econometric
MIDAS	0.13			
BVAR	0.33	0.62		
Econometric	0.58	0.58	0.60	
Single-equation	0.59	**0.73**	**0.71**	**0.86**

Correlations greater than 0.7 are in bold.

A possible theoretical explanation for the correlations among errors of the models is the existence of an unobserved variable and/or a permanent change in the effects of a variable that none of the models captures. Therefore, the models tend to err in a similar direction.

To test the fit between the data and theory, the total proposed weighted forecast may be divided into two parts:

1. A positive component—the contribution of positively weighted models according to the proposed weights;
2. A negative component—the contribution of negatively weighted models according to the proposed weights (in this case, only the single-equation model is negatively weighted).

The closer the correlation between the errors of the two components is to (-1) (after the negative weight is added), the more the errors of the two components will be offset and the more accurate the forecast will be. In this case, given the negative weight, the correlation obtained between the errors of the negative component and those of the positive one is (-0.934).[19] Again, these data reinforce the proposition that the major improvement in reducing MSE is explained by the use of error correlation, as presented in the Theory section. This outcome indicates that the optimal weights obtained sustain a very strong correlation between the errors of the two components of the forecast. By using this correlation and a negative weight, the proposed weighting allows a substantial minimization of errors, as shown.

These matters are described graphically in Figure 1, which illustrates the errors and the result of their counteraction. The negative components (marked in blue) and the positive one (in red) are combined into one model (in green). This reduces the errors (brings them closer to the zero line).[20]

[19] After the deletion of an outlying observation (July 2011), a correlation of (-0.894) is obtained; therefore, it cannot be argued that the positive correlation among the forecast errors is based solely on exceptional observations.

[20] In the figure, the negative component is shown after multiplication by the negative weights.

Figure 1. Component errors and the proposed model

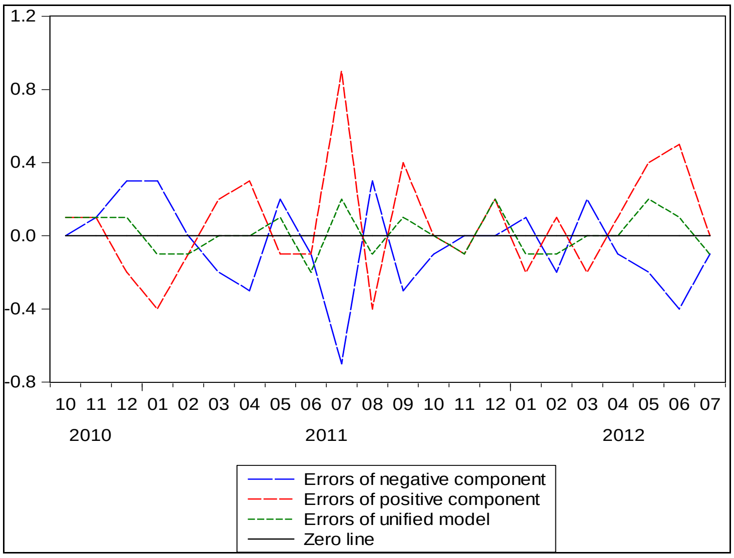
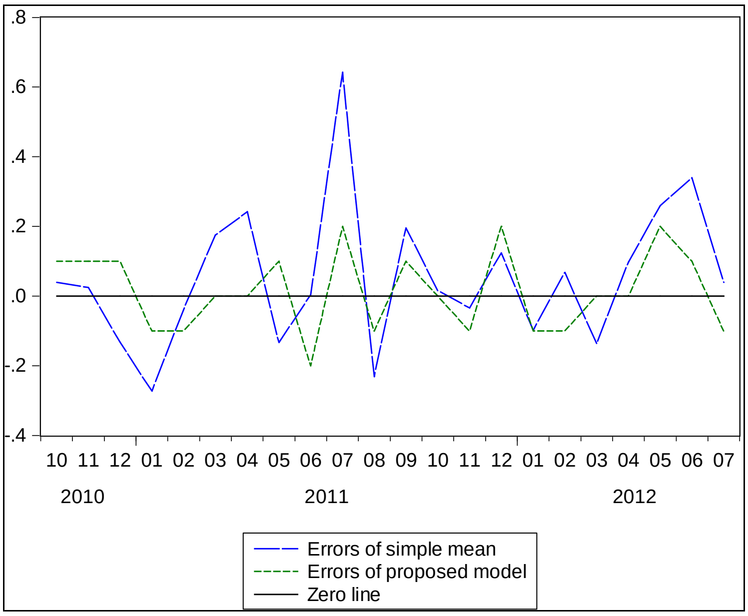


Figure 2 shows the squared errors of the method used today (simple mean—marked in blue) and of the proposed weighting method (in green). The figure graphically demonstrates the major improvement that the proposed method allows (that is, the errors in the proposed method are closer to zero line).[21]

Figure 2. Errors of the simple-mean method and the proposed model



[21] It should be noted that this is an in-sample analysis. An out-of-sample examination follows in the next subsection.

3.3 Testing forecast quality and stability

3.3.1 Forecast quality (out-of sample tests)

To test forecast quality, an out-of sample test in two phases will be used. In Phase 1, optimal weights are calculated on the basis of a partial sample that includes observations from the beginning of the sample to a given month (in-sample). In Phase 2, using a sample (out-of sample) that contains the month or months following the partial sample, the errors of the forecast that is calculated on the basis of the optimal weights found in Phase 1 relative to the actual CPI are examined. The test is performed by two methods:

1. calculation of weights on the basis of part of the sample and forecasting to one step (one month) ahead;
2. calculation of weights on the basis of part of the sample and forecasting for all months up to the end of the sample (cross-validation).

Both tests were run several times on in-sample samples of different sizes (15–21 observations). Table 5 shows the mean squared errors of all the tests and compares it with the weighting method used today (simple mean).

Both tests found approximately a 50 percent improvement in out-of-sample forecasting when the current method is replaced with the proposed one. A Granger-Newbold test (1976)[22] shows that the proposed method improves the out-of-sample results considerably compared with the current method, at <0.1 percent significance.

Table 5. Mean squared errors in out-of-sample test

	Simple mean	**Proposed weights**
* MSE using Method 1 (one month ahead)	0.032 (0.043) [0.18]	0.017 (0.016) [0.13]
* MSE using Method 2 (successive months)	0.040 (0.020) [0.2]	0.020 (0.006) [0.14]

Standard deviations of the square errors appear in parentheses. Root mean squared errors (RMSE) appear in brackets.

* Mean squared errors.

3.3.2 Stability of weights—recursive estimation

To examine the stability of the weights obtained, they were tested in a sample that increased from fifteen observations to twenty-two. This test resembles the updating of weights that will actually be used (at least initially) on the basis of a recursive sample.

[22] This test deals with the evaluation of forecasts; unlike other evaluation tests, it does not require the absence of contemporaneous correlation between the errors of the forecasts.

Table 6. Recursive weights

	Model / Sample size	Econometric	Statistical	Single-equation	BVAR	MIDAS
1	**15**	0.32	0.63	-0.70	0.12	0.62
2	**16**	0.34	0.61	-0.70	0.17	0.58
3	**17**	0.32	0.61	-0.68	0.16	0.59
4	**18**	0.33	0.60	-0.67	0.16	0.58
5	**19**	0.37	0.56	-0.67	0.19	0.55
6	**20**	0.40	0.58	-0.69	0.19	0.52
7	**21**	0.53	0.51	-0.76	0.19	0.54
8	**22**	0.56	0.48	-0.76	0.22	0.50

The results shown in Table 6 indicate that the signs of the weights remain constant but their values change in the transition from the partial sample (15 observations) to the full one (22). The main changes are an increase in the weight of the econometric model (by 0.24) and a decline in that of the statistical one (by 0.15).[23]

We now test to see whether the changes in the weight values, resulting from change in sample size, are meaningful in terms of reducing MSE. To accomplish this, we observe the MSE obtained in a sample of twenty-two observations, in which the weights were determined on the basis of a 15 observation sample (Table 6, Line 1), and compare it with weights determined on the basis of a 22 observation sample (Table 6, Line 8). The test yields an MSE of 0.015 when the Line 1 weights are used (fifteen observations—the smallest sample) as against 0.013 when the Line 8 weights (twenty-two observations—the largest sample) are used. The difference is minimal, especially relative to the MSE obtained via the use of the simple mean method (0.043).

Finally, we test whether the results would be seriously affected if the sample were extended. To do this, a sample of forty-two observations (February 2009–July 2012) for four models (omitting the BVAR model because it wasn't calculated regularly during part of the period) is used. Observing the MSE in the forty-two-observation sample using weights determined on the basis of a twenty-six-observation sample (small sample) [24] and comparing it with the use of weights determined by the forty-two-observation sample (long sample),[25] MSEs of 0.046 and 0.042, respectively, were found. This negligible difference shows that the use of a long sample would not seriously modify the effectiveness of the weights in reducing MSE and that changes in the weights over time are also inconsequential in this respect.[26]

[23] A similar check for a model free of restrictions (in coefficient sum and sign) finds that the sum of the coefficients is not significantly different from 1 in samples of all sizes.

[24] Once the unified model begins to be used on a regular basis, it will have at least twenty-six observations.

[25] Recursive weights determined on the basis of samples of varying lengths are shown in Appendix 2.

[26] The MSEs obtained from the forty-two observation sample (excluding the BVAR model), mentioned in this paragraph, exceeded those obtained from the twenty-two observation sample (including BVAR) mentioned in the previous paragraph. This may be due to the omission of the BVAR model or to the difference in sample lengths. When the MSE was tested in the twenty-two observation sample excluding the BVAR model, a value of 0.015 was found. Therefore, the increase in MSE traces to the change in sample length and not to the omission of the BVAR model.

4. Implementation of the unified model

Due to the sample size and the wish to maximize precision, the weights will be adjusted every month. That is, the optimal weights will be calculated on the basis of a sample that includes the previous month's data.

In addition to update frequency, changes in the models should be addressed. The various CPI forecasting models are occasionally revised to improve their estimation ability, meaning that the models which are weighted may change from time to time. Practically speaking, such changes will be considered on the basis of the following decision-making rule: if the change is minor, the existing weighting will remain in use and the change will be ignored, and if the change is meaningful, the model will be calculated as a new model and the forecasting will be redone retroactively with the help of vintage data. Note that the rule will also be applied to each model in the sample tested in this study.

5. Conclusion

The Research Department's one-month ahead inflation forecast is currently based on a simple mean of five models that predict CPI change in the coming month. This study proposes differential weighting of the models' forecasts and the assignment of a negative weight to one of them. The study began by reviewing the sources of inspiration for this proposal; they relate to exploiting the potential in a positive correlation among errors of the models, namely by weighting some of them negatively. Further on, the study presented empirical results that support its central proposition. The results show that when the existing weighting method is replaced with the new one, forecasting accuracy improves considerably (by around 50 percent in MSE in out-of-sample tests). The results also demonstrate that it is preferable to include the negatively weighted model rather than exclude it from the weighting.

The Research Department's inflation forecast is an important element in the formulation of monetary policy, which aims to maintain price stability. An improvement in this indicator will supply policymakers with more accurate forecasts and improve their decision making.

The study also establishes grounds for the insight that the use of a negative weight in the weighting of forecasts allows us to exploit positive correlations among errors. This insight may be useful in weighting forecasts of other economic variables and improving the Bank of Israel's forecasting ability in different contexts.

6. Sources

Bates, J. M., & C. W. J. Granger (1969). "The combination of forecasts". *OR*, 451-468.

Bunn, D. W. (1985). "Statistical efficiency in the linear combination of forecasts". *International Journal of Forecasting*, *1*(2), 151-163.

Blank, M. (2007). "Estimations of private analysts' inflation forecasts", *Monetary Studies 2007.07*, Bank of Israel (in Hebrew).

Dickinson, J. (1975). "Some comments on the combination of forecasts". *Operational Research Quarterly*, 205-210.

Granger, C. W. J., and P. Newbold (1974). "Spurious regressions in econometrics". *Journal of Econometrics*, *2*(2), 111-120.

Granger, C. W., and P. Newbold (1976). "Forecasting transformed series". *Journal of the Royal Statistical Society. Series B (Methodological)*, 189-203.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C.T. Volinsky (1999). “Bayesian model averaging: A tutorial”. *Statistical Science,* 382-401.
Ilek, A. (2006). “Monthly model for estimation of inflation and monetary policy in Israel”. *Monetary Studies 2006.4,* Bank of Israel (in Hebrew).
Kapetanios, G., V. Labhard, and S. Price (2008). “Forecast combination and the Bank of England's suite of statistical forecasting models”. *Economic Modelling, 25*(4), 772-792.
Reinmuth, J. E., and M. D. Geurts (1979). “A multideterministic approach to forecasting”. *TIMS Studies in the Management Sciences, 12*, 203-211.
Ribon, S., and T. Suhoy (2011). “Forecasting short-run inflation using mixed frequency data (MIDAS)”, *Occasional Paper, 2011.02*, Bank of Israel.
Sorezcky, A. (2009). “Inflation forecasts by auto-regressive models”, Internal Memorandum, Bank of Israel (in Hebrew).
Suhoy, T., and Y. Rotberger (2006). “Enhancing the accuracy of the seasonal model for forecasting CPI in the short run”. *Discussion Paper 2006.06,* Bank of Israel (in Hebrew).
Timmermann, A. (2006). Forecast combinations. *Handbook of Economic Forecasting, 1*, 135-196.
Winkler, R. L., and R.T. Clemen (1992). “Sensitivity of weights in combining forecasts”. *Operations Research, 40*(3), 609-614.

7. Appendices

Appendix 1. CPI forecasting models

1. **Statistical model**—based on separate forecasts for each of the ten main components of the CPI (several components are divided into subcomponents) (Suhoy & Rotberger, 2006).
2. **Econometric model—**based on the Phillips curve approach and generates a forecast for the overall CPI by forecasting a "core" index—the CPI excluding fruit and vegetables, clothing and footwear, energy, and housing (Ilek, 2006).
3. **BVAR model**—a VAR model estimate that applies a Bayesian method to a set of variables: the CPI excluding housing, the housing component, inflation expectations, exchange-rate changes, the Bank of Israel interest rate, exports, and private consumption.
4. **Single equation model**—an autoregressive equation for the price index and a separate equation for the housing component of the CPI (Sorezcky, 2009).
5. **MIDAS model—**based on data of differing frequencies; it allows the inclusion of financial data and global commodity prices with daily frequency by adjusting distributions (Ribon & Suhoy, 2011).

Appendix 2. Recursive weights and an expanded sample omitting the BVAR model

Model / Sample size	**MIDAS**	**Single-equation**	**Econometric**	**Statistical**
26	0.29	0.24	0.09	0.38
28	0.32	0.17	0.10	0.41
30	0.34	-0.02	0.11	0.58
32	0.39	-0.07	0.11	0.58
34	0.40	-0.08	0.11	0.57
36	0.38	-0.06	0.15	0.53
38	0.38	-0.07	0.15	0.54
40	0.36	-0.07	0.17	0.55
42	0.38	-0.12	0.26	0.48