Research Department Bank of Israel

Prediction of the Monthly Change in the Fruit and Vegetables component of the Israeli CPI

Gilad Gaibel*

Occasional Paper 2021.02
November 2021

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

* Gilad Gaibel - Information and Statistics Department (formerly), Bank of Israel. Email: gaibelg@gmail.com
I would like to thank Shir Kamenetsky, Tim Ginker, Yonatan Rosen and Ariel Mantzura from the Statistical Methods and Data Science unit at the Bank of Israel, Itamar Caspi, Sigal Ribon, Dor Goldenberg and the participants of the Bank of Israel Research Department seminar for their helpful comments and suggestions. I thank Boaz Ben Aharon, Merav Oren-Yiftach and Yoel Finkel from the Central Bureau of Statistics for the useful discussions.

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

תחזית לשינוי החודשי בסעיף הפירות והירקות במדד המחירים לצרכן בישראל

גלעד גיבל

תקציר

עבודה זו מתארת מודל נאוקסט חדש לחיזוי שיעור השינוי החודשי בסעיף הפירות והירקות במדד המחירים לצרכן. מחירי סעיפי הפירות והירקות הינם תנודתיים כך שלעיתים יש להם השפעה גדולה על מדד המחירים הכללי. לפיכך, חיזוי נפרד לסעיף זה הוא בעל ערך. מודל החיזוי החדש עושה שימוש במספר מקורות מידע וביניהם מאגרי מחירים פרטניים ומאגרי מחירים סיטונאיים כדי להקדים בשבועיים את הפרסום של סעיף זה במדד. האתגר העיקרי במשימת חיזוי זו הוא מיעוט התצפיות הקיימות לצורך אימון המודל ולבחינת תוצאותיו. בחינת תוצאות המודל מחוץ למדגם מצביעות על שיפור של 33% במונחים של שורש הטעות הריבועית הממוצעת בתקופה שנבחנה, בחודשים שבין מאי 2020 לאפריל 2021.

Prediction of the Monthly Change in the Fruit and Vegetables component of the Israeli CPI

Gilad Gaibel

Abstract

This paper describes a new nowcast model for the monthly percent change in the fruit and vegetables price index. Due to the high volatility of this index, the prices of fruit and vegetables have a large effect on the dynamics of the headline Consumer Price Index. Therefore, a standalone prediction for the fruit and vegetables price index may be beneficial. The new model uses multiple data sources, including granular country-wide retail price and wholesale price databases, to precede the official index by two weeks. The main challenge in this prediction task is the small amount of data points available for training and evaluation. Out-of-sample evaluation shows that the new model improves upon the previous one by 33% RMSE using the months May 2020– April 2021 as test periods.

Contents

1. Introduction

When making monetary policy decisions, central banks rely on information about the dynamics of aggregate prices. This information includes periodical changes in the Consumer Price Index (CPI). The Israeli CPI is published on a monthly basis by the Israel Central Bureau of Statistics (CBS), where the index value of a given month is published on the 15$^{\text{th}}$ of the next month. Hence, the Bank of Israel (BOI) receives aggregate price information with some delay. In order to make better informed decisions, it is useful to incorporate inflation assessments and forecasts before the official versions are published by CBS. Therefore, the BOI generates, for internal uses, a forecast of the changes in the CPI and its various components.

The CPI is composed hierarchically of lower-level price indices. For example, total consumption includes housing, food, fruit and vegetables and more, while the fruit and vegetables price index includes fresh fruit, fresh vegetables, canned fruit, etc. The fruit and vegetables (F&V) price index is highly volatile compared to other indices of the same level. Therefore, although it has a relatively low weight in the CPI ($\approx 3\%$), the monthly changes in the F&V price index may significantly influence the monthly changes in the CPI. In fact, in the years 2016-2020 the monthly changes in F&V prices were responsible for $\approx 16\%$ of the variance of the monthly changes in the CPI.[1] This highlights the need for a stand-alone prediction of the monthly changes in F&V price index.

The usual approach to the challenge of predicting the CPI or its components is a forward-looking forecast (for a survey see: Faust & Write 2013). This approach typically uses past dynamics of the target index itself or other exogenous series in order to produce predictions of future values (e.g., Barkan et al. 2021 focus on endogenous dynamics, while Medeiros et al. 2021 use exogenous dynamics). This approach can provide future values of the target index up to some reasonable forecast horizon. Such a method focuses on identifying key patterns either in the dynamics of the index itself or in the dependence of the target index on past values of other series. That is why it cannot capture out-of-pattern behavior. Hence, the quality of this method is limited by the extent to which the target index behaves according

[1] This is calculated as one minus the ratio of the variance of the CPI without F&V to the variance of the general CPI in the years 2016-2020.

to discernible patterns, and this method cannot capture fluctuations due to sudden shocks.

In contrast, an alternative approach would be to focus on streams of observed data, such as observed prices. This way, a model can produce predictions of the target index up to the last available date in the data. If these streams arrive without a significant delay (say within a day), a model could provide a well-informed nowcast. Because this model is not forward-looking, it relies less on inter-temporal patterns (as it does not focus on inter-temporal dependencies), and its performance is typically affected less by shocks. Note, however, that this model cannot produce forecasts for periods for which there is no observed data. This means that in practice, with respect to the F&V price index, this method can precede the official index by two weeks at most. Of course, each of the two types of approaches serves a different purpose.

The BOI previously used a nowcast model which took low-level product-family observed prices as input (e.g., tomatoes, bananas etc.) and produced a prediction for the percent change in the official F&V price index for the last month. The model has two stages. First, an OLS regression is fitted for each product family using monthly percent change in the product's official price index (target) and monthly percent change in available prices data (predictor). The predictions from each OLS model are then weighted using precalculated weights (used by CBS to calculate the original CPI) to generate the output of the model. Recently, an alternative model that uses granular retail data has been proposed. This model takes observed daily prices of specific products from a collection of retail stores and uses the CPI weights to aggregate monthly percent changes to a single index (Goldenberg & Rosen 2021).

This paper describes a new nowcast model for the monthly percent change in the F&V price index. The suggested model expands on the two models mentioned above in several aspects. First, the new model makes use of both retail and wholesale data to enhance the available information. Second, the model incorporates a selection scheme, in which only the best predictors of each lower-level index are taken into account. To avoid over-fitting, the selection scheme is based on a predetermined training set and evaluated on the remaining test examples. Third, the model accounts for in/off season patterns. Fourth, the retail data used for the model covers country-wide retail prices of all major retailers in Israel. Finally, the model is mostly automated and relies on a data-driven hyperparameter tuning scheme.

Essentially, the suggested model follows the lines of CBS's original CPI calculation framework. The model uses four sources of data: country-wide daily retail prices of relevant products, daily wholesale prices of F&V, monthly CBS historical price indices, and the annual Household Expenditure Surveys by CBS. Using these data sources, the model outputs predicted monthly percent changes up to the last available date in the data (usually a 1-day delay). Using the months May 2020 – Apr. 2021 as a test set, the predictions of the suggested model are significantly more accurate than those of the previous one: 33% lower root mean squared error (RMSE) and 34% lower mean absolute error (MAE).

The rest of the paper is organized as follows. Section 2 contains a description of the data. Section 3 contains details about the new F&V nowcast model. Section 4 concludes with results and final remarks.

2. Data

The new nowcast model uses four sources of data:

1. **Annual CBS Expenditure Surveys**. The survey is designed to provide information on both income and household expenditure patterns. It contains detailed data about consumption of products (total expenditure). The products in the survey are encoded in a hierarchical fashion where products are grouped into families in a tree-like structure. Total consumption is composed of ten different sub-categories (housing, food, clothing etc.); each of them is composed of more detailed sub-categories, and so on. There are 4 levels in the hierarchy, encoded by codes of different length: total consumption is a 1-digit level index; its sub-categories are at the 2-digit level; and then there are 3-digit and 6-digit levels. For example: tomatoes are a 6-digit level category that belongs to the 3-digit level category of fresh fruit; fresh fruit belongs to the 2-digit level category of F&V.

2. **CBS Price Indices Archive (On the CBS's website).**[2] CBS maintains an online database of various price indices. This database includes a selection of 6-digit level indices of F&V, as well as the F&V price index itself, all at a monthly frequency.

3. **Retail Data - Mysupermarket**. The Mysupermarket database contains daily prices of all products that are sold in supermarkets in Israel. The data covers all of the branches of all main retail chains in Israel since Jan. 2016. The dataset contains information such as item name, unified item code, branch, chain, price, price after promotion, and more. For this model, we only use products that are related to F&V. Between Jan. 1st 2016 and June 1st 2021, there are 279,509,642 price records in the data, gathered from 2,108 supermarkets country-wide. Each day, an average of 142,027 additional records are added.

4. **Wholesale Data - Israel Ministry of Agriculture (MOA).**[3] The MOA database contains average daily prices of fresh F&V traded in Jerusalem or Tzrifin wholesale markets. The data tracks prices of each fruit or vegetable type in two quality classes - type A and prime. Between Jan. 1st 2016 and June 1st 2021, there are 114,569 records in the data. The average daily amount of records in trading days is 87.

3. The New F&V Model

In the following section, both technical and theoretical details about the different stages and subroutines in the new F&V model are provided. I begin with the preparations and the calculations of fixed parameters, then the aggregation process, ad-hoc corrections and the hyperparameter-tuning.

As a starting point, it is important to note that the main challenge in predicting the F&V price index is the small amount of available data. The target variable has a monthly frequency. In addition, the available data contains information only since Jan. 2016. This means that, effectively, there are only a little more than 60 data points over all (considering

[2] Available online at https://www.cbs.gov.il/he/Statistics/.
[3] Available online at https://prices.moag.gov.il/.

monthly percent change, up to June 2021 there are 64 data points). This fact has two main implications. First, it is difficult to validate the model, since using a reasonable amount of training examples leaves few data points for out-of-sample evaluation. In addition, calibration of the hyperparameters of the model relies on a rather short period. Nevertheless, to avoid over-fitting the test set, I also limit the search space over the hyperparameters of the model as much as possible. Second, the few training examples relative to the amount of predictors and the granularity of the training data makes it easy to over-fit the training data. Therefore I limit the number of model-parameters and use precalculated parameters as much as possible; also, class of predictive models is limited to simple linear models.

3.1. Precalculated Parameters

CBS's Household Expenditure Survey contains data regarding expenditure patterns of households in Israel. It is possible to compute the share of expenses on a specific product or a set of products out of the total household expenditure. The population mean of such a measure defines the *consumption weight* of a specific set of products (or a single product). The CPI is composed of lower level price indices of different sets of products, in a hierarchical structure. The aggregation scheme from lower level indices to higher levels is by a weighted average, using the proper consumption weights. In addition, in order to account for changes in consumption patterns, CBS uses a two-year average value of each weight. Therefore, an initial step is to calculate the consumption weights of all relevant 6-digit and 3-digit categories using the surveys of the adequate years.

Another set of parameters that is calculated using the Household Expenditure Surveys is the daily probabilities of a household doing grocery-shopping (Figure 1). CBS collects prices of products from a sample of stores each month. The sampling procedure takes into account the mean preferences of households regarding the timing of grocery-shopping and the preferred types of stores. In order to empirically calculate the probabilities of grocery-shopping over the days of the week, one can use the diary table in the survey. This table contains data on purchases of products for each household over a course of two weeks: quantities, prices, store-type, and day of the week. A first step would be to mark days of the week in which a household goes shopping for food. A second step would be to calculate the normalized daily frequency of grocery-shopping of each household. The last step is to calculate the population mean probability of grocery-shopping for each day. A two-year average of the daily probabilities (based on the adequate years) will serve as aggregation weights for daily-to-monthly transformation.

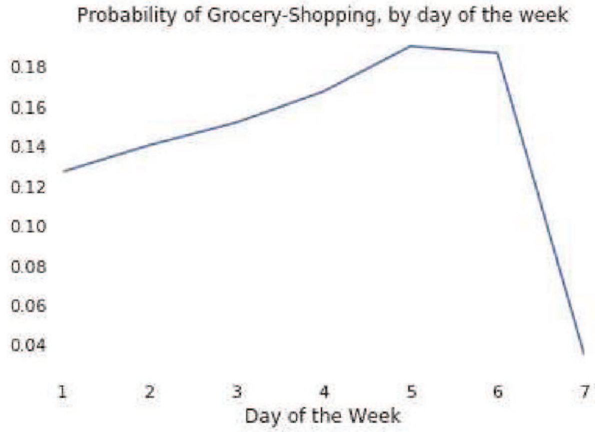

Figure 1: Daily Probabilities of Grocery-Shopping, 2018

In addition, CBS's official website contains data about relevant price indices. First, one can find the target series on the website: F&V 2-digit level index. Second, there are some lower level relevant indices available.[4] These lower level indices provide two layers of information - labeled lower level prices, and in/off season information. In general, if in a given month the price of a product in the labeled series is missing, then it is considered off-season in that month.

Some of the fruit have irregular in/off season patterns, and this classification depends on the monthly sample of stores from which CBS collects prices. It therefore happens that prices for a given fruit at a given month are available in the retail/wholesale data, however CBS disregards them as off-season. Since the target of the model is the official CBS price index, observed dynamics of prices in months when a product is considered off-season would lead to

[4] In particular, it is possible to download price indices of the following: potato, tomato, cucumber, onion, pepper, zucchini, carrot, green beans, lettuce, eggplant, cabbage, cauliflower, apple, banana, orange, mandarin, grapefruit, lemon, grapes, watermelon, melon, peach, plum, pear, avocado, tomato-sauce, pickles, olives, peanuts, nuts, almonds, raisins, fresh fruit, fresh vegetables, conserved vegetables, and dried fruit.

prediction errors. In order to minimize such errors, it is necessary to take the in/off season patterns into account. Hence, using the labeled series, it is useful to calculate the monthly probability of each product being in-season. This is calculated for each month of the year as the proportion of months in which a product was in-season over the last ten years of data.

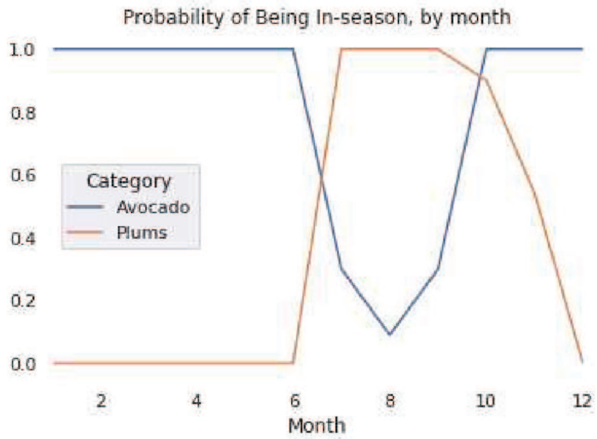

Figure 2: Examples for Probability of Being In-Season

3.2. First Aggregation Scheme

The daily weights and in-season probabilities are used to transform daily granular data into predicted monthly 6-digit level indices. First, the frequency transformation is based on a weighted average, where the weights are the daily weights. This results in monthly measures that are consistent with CBS's store-sampling scheme. Second, each series in the data is classified either as belonging to a known 6-digit level category or not. If the data is small or well-structured, this is a simple task (for example, in wholesale data). However, if the data is very large this may become more complex (for example, in retail data). Third, for each 6-digit level category, the appropriate series are averaged together, to provide an initial prediction. Last, the in-season probabilities are used to refine the predicted 6-digit level indices according to the in/off season patterns observed in the past.

3.3. Retail Prices - Selection Scheme

The retail prices dataset contains many series of daily prices of different products and brands (barcodes) over a country-wide set of stores. Not all barcodes are relevant predictors of a 6-digit level F&V category. For example, shampoo of a specific brand is not related in a natural way to the prices of bananas, or any other F&V. In a small dataset, one can manually assign each series to a corresponding 6-digit level index. However, in a large dataset with varying products this becomes more of a challenge.

The assignment of a barcode to a specific 6-digit level labeled index is therefore performed by an automated selection scheme. Conceptually, the selected barcodes are those with the most similar dynamics to the 6-digit level labeled indices. There are a few initial steps. First, a country-wide average of the daily prices is calculated for each barcode. Second, using the daily weights, the series are transformed from daily to monthly frequency. Third, it is necessary to calculate the monthly percent change in prices of each barcode. The reason for the last step is that the objective is to predict changes in prices, and therefore the similarity between the labeled series to each candidate barcode should be measured using the monthly percent change in each. After these three steps, the Pearson correlation between every label-barcode pair is calculated, and the barcodes with the highest scores per labeled index are considered the best predictors for that index. In addition, in order to avoid spurious correlations and to increase the interpretability of the model, an additional two constraints are forced. First, for each labeled index, only barcodes with names that contain some variation of the name of the labeled index (e.g. for tomatoes - 'tomato') are acceptable. This constraint acts both as a regularization technique and as a sensibility guarantee. Second, only series with high enough correlation (larger than some chosen threshold) with a specific labeled index are kept. Finally, only the top n barcodes per labeled index are used. That is, the correlation threshold and the maximal amount of chosen barcodes per labeled series are hyperparameters of the selection scheme. I limit the search space over the correlation threshold to 0.15 and 0.6 as low and high values, and set the amount of barcodes per series to 5. Of course, the entire selection scheme is performed only using the training periods.

Given the assignment of the most predictive barcodes per 6-digit level index, it is possible to continue the aggregation scheme. For each 6-digit level category, the appropriate series (of the chosen barcodes) are averaged together, to provide an initial prediction for the low-level indices. Then, the in-season probabilities are used to refine the predictions.

3.4. Combining Series

The wholesale data contains prices of two quality classes - prime and type A - for each fruit or vegetable type on a given day. Prices of the two quality classes are not always available, and it could be that at a given period niether, one, or both of them are available. Assuming that produce of both quality classes are supplied to F&V stores and local markets, it is better to incorporate price information from both. In order to reduce the amount of estimated parameters, I choose a simple functional form - simple average - which could potentially capture a representative global price. However, having missing values in some periods requires a more elaborate technique. Therefore, whenever the prices of the two quality classes are available, the average price is used. When only one of the two is available, I scale it by the yearly average ratio of the available quality class prices to the average prices.[5] If instead, none of the prices of the two quality classes is available, I leave a missing value. Note that for the wholesale data, this procedure is done after the first aggregation scheme, as described above, has been performed for each of the two quality classes separately. Hence, this averaging results in wholesale-based predictions for the 6-digit level categories.

Given the predicted 6-digit level categories from both retail and wholesale data, I apply a similar combination technique. This time, instead of using a simple average of retail-based and wholesale-based predictions, it is more reasonable to use a weighted average. The reason is that CBS's store-sampling methodology accounts for different types of stores - supermarkets, local stores, public markets and more. Assuming that wholesale prices data capture some of the price-dynamics in the non-retail store types better than retail prices, incorporating wholesale prices may improve the accuracy of the model. It is straightforward to compute the mean share of purchases made in supermarkets relative to other stores using the Household Expenditure Surveys. However, there is a category of 'local stores' in the survey, and it is not

[5]If the average ratio of a whole year is missing, I use the last known value.

a-priori clear whether these stores are more similar to retail or wholesale in terms of pricing. The proportion of purchases in retail stores, can therefore be ≈ 0.88 or ≈ 0.63, depending on the classification of local stores (these numbers are quite stable over the years). Finally, the actual weight of the retail data (which is one minus the weight of wholesale) is left to be a hyperparameter of the model. Nevertheless, in order to limit the search space over the values of this hyperparameter I consider only the two values - 0.88 and 0.63.

3.5. In/Off Season Adjustments

Some of the low-level categories of the CPI exhibit irregular availability patterns over time. For example, plums are usually available in most of the stores in Israel only during the summer. When the CBS collects prices from stores, if a certain product is unavailable in many stores it is considered off-season, and its price is disregarded. Instead, CBS inserts a new value such that it will not affect the index value of the parent-category. This is done by calculating the monthly percent change in the parent index without the off-season category, and assigning this category a price value that corresponds to the result of the calculation. While past in/off season patterns are known, the in/off season realization of the upcoming month is unknown and can add variability to the predictions.

Note that given predictions for the 6-digit level categories from both the wholesale and retail data, there are two valid approaches: a price-based approach or percent-change-based approach. That is, one can use the monthly percent change in each category or maintain the predicted index value (price) itself for each of the wholesale or retail 6-digit level predictions. Then, combine the information from the retail and wholesale data as described above, and proceed. There is a difference between the two approaches is in how in/off season patterns are handled. When using the percent-change-based approach, price information is unavailable, so CBS's imputation technique (for prices of off-season categories) cannot be implemented.

Hence, the in/off season refinement routine depends on the chosen approach. For the percent-change based approach, I choose a classification threshold that is used to transform monthly probability of being in-season (as described above) to binary classification. Then, for each category and month of the year, if it has been classified as being off-season, the index value of that month is disregarded. The classification threshold is then another hyperparameter of the model, and I limit the search space to only two values - 0 or 0.5 - where the value 0 means that no in/off season adjustment is performed.

Alternatively, if the price-based approach is chosen, I use a more elaborate refinement technique. CBS's imputation technique relies on the in/off season *binary* value being known for each category and month. However, I only know the *probability* of a 6-digit level category being in season in a given month of the year. Therefore, CBS's imputation technique cannot be implemented directly. Instead, I calculate the expected index value, taken over all possible joint assignments (states) of in/off season of the 6-digit level predictions for each month. Given some state (a joint assignment of in/off season *binary* values), one can can extract the imputed value of an off-season category using CBS's technique. The probability of such state is calculated as the product of the probabilities corresponding to the in/off season assignment of all the sibling 6-digit level categories.[6] Then, by the law of total expectation, the expected index value of each category equals the sum over the value at each state, weighted by the corresponding joint probability.

3.6. Second Aggregation Scheme

Given the final 6-digit level predictions, after combining the retail and wholesale information and after the in/off season refinement routine, an initial prediction for the 2-digit level target is calculated. This calculation is performed by a hierarchical averaging scheme, using the proper consumption weights. Another hyperparameter of the model being used is the number of 6-digit level categories to drop. That is, the set of predicted low-level indices may contain categories with consistently bad predictions. Such a category can hurt the final prediction of the 2-digit level target, while the prediction when dropping this category may be superior. Therefore, I consider the best subset of categories, leaving out the specified amount of categories to drop, in terms of root mean squared prediction error over the training periods. I limit the search space over this hyperparameter to the values 0 or 1.

[6]Note that using joint probabilities implies a conditional independence assumption: independence of the probabilities of the different categories being in/off-season given a particular assignment.

3.7. Additional Ad-hoc Corrections

Given the initial predictions for the 2-digit level target, I use additional ad-hoc correction routines. First, I fit an OLS regression of the monthly percent change of the target on the monthly percent change in the predictions (using the training periods). The new predictions of this fitted model then satisfy the property of zero mean error (since the regression is fitted with a constant term). In addition, I consider the possibility of annual seasonality in the prediction error, and correct for it using the calculated errors in the train set.

Finally, choosing either the percent-change based approach or the price-based approach yield different predictions. An additional ad-hoc procedure is then to use both predictions as intermediate values, and to output a combination (ensemble) of the two. In order to combine them, I calculate the optimal linear combination of the predictions from both approaches using an additional OLS regression of the target on them. The purpose of that is to stabilize and improve the performance of the F&V model. As before, the OLS model is trained only on the training periods, and evaluated using the test periods.

3.8. Hyperparamater-tuning

Table 1: Summary of Hyperparameters

Hyperparameter	Values	Search Space
Number of barcodes per 6-digit level category	$n > 0$	5
Minimal correlation threshold	$q \in [0, 1]$	$0.15, 0.6$
Weight of retail data	$q \in [0, 1]$	$0.63, 0.88$
In-season threshold	$q \in [0, 1]$	$0, 0.5$
Amount of 6-digit level categories to disregard	$n \geq 0$	$0, 1$
Correct for seasonality in the error	True, False	True, False

As described, there are a few hyperparameters that govern the behavior of the model and affect its performance. A summary of the hyperparameters of the model is presented in Table 1. In general, one can either set the values of the hyperparameters according to some relevant domain knowledge. However, since I do not know of *a-priori* good reasons to justify a particular set of hyperparameters for this model, it is best practice to use a data-driven approach. In regular datasets, one can use the standard leave-one-out cross-validation (CV) scheme. However, since regular CV implicitly assumes some sort of independence between observations, it is not useful for time-series data. Instead, I use a rolling one-step time-series CV scheme (Hyndman and Athanasopoulos, 2018). For the desired test month, I calculate an aggregate prediction error for the previous validation months for each of the possible sets of hyperparameters. The prediction error for each of these preceding validation months is calculated based on an expanding one-step prediction. Then, the chosen set of hyperparameters is that which leads to the minimal aggregate prediction error over the validation periods. The error aggregation function I use is the mean squared error.

4. Results and Concluding Remarks

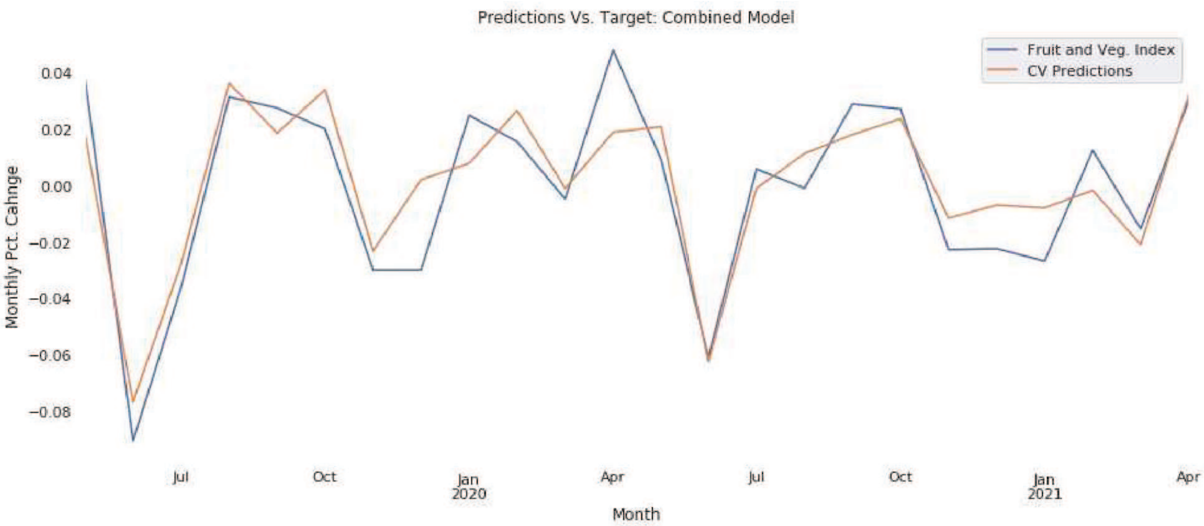

Figure 3: Predictions of the Combined Model

The performance of the model is evaluated mainly by the root mean squared error (RMSE). RMSE provides an estimate for the mean size of the prediction error. Since the RMSE is sensitive to extreme values (as it emphasizes large errors), I also track the mean absolute error (MAE). Typically, these metrics are highly correlated, but still they provide together a richer picture of the prediction error. All of the predictions are compared with two benchmarks: the previous F&V model used in the BOI and the aggregate F&V index that is calculated

Table 2: Prediction Results

	Test		Average on Train	
	RMSE	MAE	Average RMSE	Average MAE
A. Test size = 12				
CBS Labels	1.09	0.75	0.88	0.74
Previous Model	1.62	1.43	-	-
Prices Based Model	1.69	1.42	2.45	1.92
Pct. Change Based Model	1.38	1.11	1.94	1.49
Combined Model	1.09	0.95	1.64	1.25
B. Test size = 18				
CBS Labels	0.98	0.71	0.89	0.75
Previous Model	1.79	1.58	-	-
Prices Based Model	2.39	1.89	2.74	2.12
Pct. Change Based Model	1.59	1.29	1.65	1.25
Combined Model	1.45	1.19	1.44	1.09
C. Test size = 24				
CBS Labels	0.95	0.70	0.90	0.76
Previous Model	2.07	1.79	-	-
Prices Based Model	3.13	2.42	2.81	2.17
Pct. Change Based Model	1.50	1.23	1.50	1.13
Combined Model	1.40	1.18	1.34	1.00

* This table summarizes the prediction results of the new F&V model. The last month in the evaluation is April 2021. The three panels A, B and C, differ in the length of the test set: 12 months (begin in May 2020), 18 months (begin in November 2019), and 24 months (begin in May 2019). For each month, the new F&V model has been fitted with hyperparameters chosen in cross-validation. The metrics presented here are the root mean squared error (RMSE) and mean absolute error (MAE). The results regarding the test sets are over one month per fitted model, while the results regarding the training sets are a summary of models' performance on multiple training samples per fitted model.

using the CBS 6-digit level labeled categories. Comparing the performance of the new model with the previous model quantifies the benefit of using the new model. In addition, since the aggregation of the labeled low-level indices represent the best predictions one could expect from the new model, comparing the performance of the model with that potentially indicates how close the predictions are to optimal.

The prediction results are summarized in Table 2 and visualized in Figure 3. Table 2 contains the result of the new model by the price-based approach, the model by the percent-change based approach and the final prediction which combines the two. In addition, the table reports metrics from the two benchmarks - the previous model used by the BOI's research department and the aggregation of the 6-digit level labeled categories. A first observation would be that the new model by the combined version outperforms the former model by 33%, 19%, and 32% in terms of RMSE for the test spans of 12, 18, and 24 months accordingly; similar magnitudes are observed in terms of MAE. A second observation is that the combined model is consistently better than either the percent-change-based model or the price-based model by themselves. This is true both in the train and the test periods.

In addition, comparing the performance of the combined model with the aggregation of the labeled low-level indices is encouraging. With the test size of 12 months, the combined model reaches the performance of the the aggregated labeled series in terms of test RMSE and is rather close in terms of test MAE. It is apparent that for larger test spans the performance of the combined model is not as good. It could be that the good performance on the 12-months test horizon is by chance, however it could be also that this is due to the accumulation of training examples over time. That is, as time passes, more training data becomes available and it is reasonable to presume it will further contribute to the performance of the model in the future.

References

- Barken O., Benchimol J., Caspi I., Hammer A. & Koenigstein N. (2021). "Forecasting CPI Inflation Components with Hierarchical Recurrent Neural Networks." Discussion Paper 2021.06, Bank of Israel.

- Faust J. & Write J.H. (2013). "Forecasting Inflation." *Handbook of Economic Forecasting* Vol. 2 (Eds. Elliot G., Granger C. & Timmermann A.). Elsevier. pp. 2-56.

- Goldenberg D. & Rosen, Y. (2021). "Prediction of Changes in the Fruit and Vegetables Price Index Using the Retail Prices Database," *Statistical Bulletin 2020*, Bank of Israel.

- Hyndman R.J., & Athanasopoulos G. (2018) *Forecasting: principles and practice*, 2$^{\text{nd}}$ Ed., OTexts: Melbourne, Australia.

- Medeiros M., Vasconcelos G., Veiga A. & Zilberman E. (2021). "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods." *Journal of Business & Economic Statistics* Vol. 39.

- *The Consumer Price Index, Technical Publication No. 60*, Israel Central Bureau of Statistics, 1992.