LCERPA Working paper No. 2021-5

September 2021

Predictability of Aggregated Time Series

Reinhard Ellwanger, Bank of Canada
and Stephen Snudden, Wilfrid Laurier University

Predictability of Aggregated Time Series*

Reinhard Ellwanger† Stephen Snudden‡§

September 3, 2021

Abstract:

Macroeconomic series are often aggregated from higher-frequency data. We show that this seemingly innocent feature has far-reaching consequences for the predictability of such series. First, the series are predictable by construction. Second, conventional tests of predictability are less informative about the data-generating process than frequently assumed. Third, a simple improvement to the conventional test leads to a sizeable correction, making it necessary to re-evaluate existing forecasting approaches. Fourth, forecasting models should be estimated with end-of-period observations even when the goal is to forecast the aggregated series. We highlight the relevance of these insights for forecasts of several macroeconomic variables.

JEL classification: C1, C53, E47, F37, G17, Q47
Keywords: Forecasting and Prediction Methods, Interest Rates, Exchange Rates, Asset Prices, Oil Prices, Commodity Prices

*The views expressed in this paper are those of the authors; no responsibility for them should be attributed to the Bank of Canada.

†Bank of Canada, International Economic Analysis Department

‡Wilfrid Laurier University; Corresponding author; 64 University Ave W, Waterloo, ON N2L 3C7; Email: ssnudden@wlu.ca

§We thank Lutz Kilian, Ron Alquist, Christiane Baumeister, Edouard Djeutem, Konrad Zmitrowicz, Pierre Siklos, Wing Chan, Willy Chetwin and Luis Uzeda, as well as seminar and conference participants at the Bank of Canada, the European Economic Association Annual Congress, the Canadian Economics Association Annual Conference, and the International Association for Energy Economics International Conference for their detailed discussions, comments and suggestions. This paper supersedes any parallel arguments in Benmoussa et al. (2020).

1 Introduction

Macroeconomic variables are some of the most-common and well-studied time series in economics. Forecasts of these variables are critical for forming expectations, for policy making, and for validating economic models and theories.[1] However, many macroeconomic variables have a distinct structure in that they are temporally aggregated from higher-frequency data, but this factor is frequently underappreciated in existing studies.[2] While seemingly innocent, this form of aggregation has far-reaching consequences for the predictability of a series. For example, aggregation can lead to a loss of information and this can reduce forecasting efficiency (Kohn, 1982). Moreover, aggregation converts a random walk–an entirely unpredictable process–into a cumulative moving average process that is predictable based on past observations (Working, 1960). This result immediately invalidates the common practice of using the predictability of an aggregated series as evidence against the random walk hypothesis of the disaggregated data.

The goal of this paper is to study the consequences of aggregation for the predictability of macroeconomic variables. We explore the theoretical implications of aggregation for predictability, derive an improved test of predictability, and propose a simple forecasting method to correct for the information loss induced by aggregation. We then apply these insights to several macroeconomic series that are frequently expressed as monthly or quarterly averages of daily data in forecasting applications, which is often the case for real variables. This includes the real prices of crude oil,[3] retail gasoline (see, e.g. Anderson et al., 2013; Baumeister et al., 2017) and non-energy commodities (see, e.g., Issler et al., 2014; Chen et al., 2014; Alquist et al., 2020) as well as interest rates and real exchange rates (see, e.g. Carabenciov et al., 2013; Ca'Zorzi et al., 2017; Johannsen and Mertens, 2021; Carriero et al., 2021). Our main findings can be summarized in four points.

First, aggregated macroeconomic series are typically predictable by construction. In an extension to Working's (1960) classical result, it can be shown that aggregated series are predictable for an arbitrary autoregressive integrated moving average (ARIMA) structure of the underlying data. For a large class of plausible data-generating processes, including a random walk, aggregated data follow predictable processes with autocorrelated innovations.

[1] Prominent economic theories that are tied to the predictability of economic variables include the efficient market hypothesis (Fama, 1970), real business cycle models (Nelson and Plosser, 1982), and the permanent income hypothesis (Deaton, 1987).

[2] As in Rossana and Seater (1995), we use the term *temporal aggregation* or simply *aggregation* to refer to constructing a lower-frequency series by averaging or summing non-overlapping observations.

[3] This is the case for the forecasts of the monthly real price of crude oil in Baumeister and Kilian (2012); Alquist et al. (2013); Baumeister et al. (2014); Baumeister and Kilian (2015); Snudden (2018); Funk (2018); Garratt et al. (2019) as well as for forecasts of the quarterly real price of crude oil in Baumeister and Kilian (2014).

Second, the conventional practice of comparing the performance of a forecast against that of a simple no-change forecast is uninformative about the data-generating process. Improvements over this benchmark are commonly used to reject the random walk hypothesis or to advocate for the practical usefulness of a particular model. They have also been used to argue that the prices of certain physical commodities are fundamentally different from those of financial assets (see, e.g. Baumeister and Kilian, 2012). However, we show that improvements over the conventional no-change benchmark are expected under the random walk model and do not necessarily provide evidence against this critical null hypothesis.

Third, a simple improvement of the conventional test leads to a sizeable correction, making it necessary to re-evaluate existing forecasting approaches. An immediate consequence of the random walk model is that a different benchmark should be used to test for predictability: the no-change forecast based on the end-of-period observation. This forecast is the optimal forecast under the random walk hypothesis and leads to large forecast improvements under the null. We show theoretically that for monthly or quarterly data the one-step-ahead improvements to the mean squared prediction error (MSPE) from using the optimal benchmark are around 45 percent. The theoretical improvements decrease with the forecasting horizon but are still measurable for up to two-year horizon forecasts. Empirically, this pattern is borne out remarkably closely in the data of all of our series.

The optimal benchmark hence raises the bar for forecast evaluations. We show that forecasts from simple time-series models often significantly outperform the conventional no-change benchmark, especially at shorter horizons. However, with rare exceptions, the same forecasts do not outperform the optimal no-change benchmark. This suggests that existing models that have derived their legitimacy from improvements over the conventional no-change forecast should be re-evaluated.

Fourth, forecasting models should be estimated with end-of-period observations when the goal is to forecast the aggregated series. We demonstrate that substituting aggregated data with end-of-period observations within lower-frequency models helps to recover the information loss that is introduced by aggregation (Tiao, 1972; Wei, 1978). Our method substantially improves upon time-series models that are estimated with average prices and yield highly significant and robust improvements of 35 to 59 percent in the one-step-ahead MSPE, which is unprecedented in existing forecast applications for many of these series. While a flexible autocorrelation structure should in principle help to deal with the information loss even when the models are estimated using averaged data, in practice, these results show that aggregation does indeed greatly diminish the out-of-sample forecasting performance of model-based forecasts. Contrary to other approaches that deal with aggregation, our approach has the advantage that it allows forecasters to maintain the same frequency as the variable of interest.[4]

These results demonstrate that temporal aggregation is of first-order importance for macroeconomic series that are forecasted in levels.[5] They complement a related body of existing work on the effect of aggregation on forecasts of returns and on tests of predictability in return regressions (see, e.g. Working, 1960; Cowles, 1960; Schwert, 1990; Conlon et al., 2021) as well as on the dynamics of macroeconomic series (see, e.g. Rossana and Seater, 1995; Marcellino, 1999). Given that all of these studies point to potentially serious pitfalls that arise from aggregation, it is surprising that this feature is often ignored in macroeconomic forecasting applications.

A potential explanation is that researchers may be unaware that their data is aggregated as lower-frequency data from standard data providers are often averaged by default and without explicit labelling.[6] Moreover, historical price deflators are only available at lower frequencies, which complicates the introduction of higher-frequency data for real variables. This could help to explain why the use of aggregated data is particularly widespread in forecasts of real variables. Finally, averaging data could appear preferable when high-frequency variables are thought to be plagued by measurement error or other forms of random noise. Alleviating any practical concerns, we show that end-of-period observations can help to improve forecasts and inferences for aggregated data, even within the lower-frequency environment that many forecasters prefer.

2 The Optimal Benchmark Forecast for Aggregated Data

This section discusses the effect of aggregation on the predictability of a series that is forecasted in levels. A common test for predictability is to compare the accuracy of forecasts against that of a simple no-change forecast. Improvements over the no-change benchmark are then used to argue that the series is predictable in general and that the specific forecasting approach is more

[4]For example, Lütkepohl (1984) proposed estimating time-series models with high-frequency data and then aggregate forecasts. Ghysels et al. (2004) MIDAS approach uses higher-frequency predictors for lower-frequency macroeconomic variables (see, e.g. Andreou et al., 2013; Baumeister et al., 2015).

[5]Forecasts of aggregated variables in levels are crucial to predicting average economic conditions over a certain time period. Similarly, forecasts of average prices can be more useful than forecasts of end-of-period observations to predict total costs or revenues over a certain time period. Consistent with the literature, we therefore maintain the goal of forecasting the aggregated series in levels throughout the paper.

[6]This is the case for several of our series, which we obtained from Bloomberg, FRED, the US Energy Information Administration, and the IMF.

accurate than a naive forecast. We review this practice in a setting in which the series of interest are temporally aggregated.

We first consider the null hypothesis that the higher-frequency observations follow a random walk. This setup is useful because it highlights the consequences of averaging in critical cases in which all future observations are unpredictable, which is the implicit null hypothesis in comparisons with no-change forecasts. Consider a generic series of daily observations that are labelled "prices" that follow a random walk, such that $p_{t,i}$, the price on day i in month t, is given by

$$p_{t,i} = p_{t,i-1} + \epsilon_{t,i}, \qquad \text{for } i = 0, 1, 2, ..., n. \tag{1}$$

Here, n is the number of daily price observations within a month and we assume that the innovation, $\epsilon_{t,i}$ is a martingale difference series with unconditional variance σ_ϵ^2.[7]

We also define $p_{t,0} = p_{t-1,n}$ to transition between months. The temporally aggregated series is the average monthly price in month t,

$$\overline{p}_t \equiv \frac{1}{n}\sum_{i=1}^{n} p_{t,i}. \tag{2}$$

Throughout the setting, the forecaster's goal is to predict the average monthly price k periods ahead, $\overline{p}_{t+k}$, given time t information.

Proposition 1. *Under the random walk hypothesis, the MSPE-optimal forecast of $\overline{p}_{t+k}$ given time t information is $p_{t,n}$.*

Proof: The forecast that minimizes the conditional expectation $E_t(\overline{p}_{t+k})$, which in our setting is

$$\begin{aligned} E_t(\overline{p}_{t+k}) &= E_t\left(\frac{1}{n}\sum_{i=1}^{n} p_{t+k,i}\right) && (3)\\ &= \frac{1}{n}E_t\left(n \cdot p_{t,n} + n \cdot \sum_{j=1}^{k-1}\sum_{i=1}^{n}\epsilon_{t+j,i} + \sum_{s=1}^{n}(n+1-s)\cdot\epsilon_{t+k,s}\right) && (4)\\ &= p_{t,n}, && (5) \end{aligned}$$

where the last step follows from the assumption that ϵ is a martingale difference series. □

Thus, the price on the last day in period t is the best predictor of the average monthly price in

[7]This assumption is more general than the *iid* assumption. For example, it allows $\epsilon_{t,i}$ to follow a wide range of unpredictable processes with conditional heteroskedasticity.

period $t+k$. We now show that this forecast is strictly better than the no-change forecast that is based on the aggregated data $\overline{p}_{t+k}$.

Proposition 2. *Under the random walk hypothesis, the MSPE of the conventional no-change benchmark is strictly larger than the MSPE of the optimal benchmark.*

Proof: The expected squared forecast error for the k-months-ahead average price under the end-of-period no-change forecast is given by

$$E\left[\left(\overline{p}_{t+k} - p_{t,n}\right)^2\right] = E\left[\left(\sum_{j=1}^{k-1}\sum_{i=1}^{n}\epsilon_{t+j,i} + \frac{1}{n}\sum_{s=1}^{n}(n+1-s)\cdot\epsilon_{t+k,s}\right)^2\right] \tag{6}$$

$$= \left((k-1)\cdot n + \frac{(n+1)\cdot(2n+1)}{6n}\right)\sigma_\epsilon^2, \tag{7}$$

The expected squared forecast error for the k-months-ahead average price under the average no-change forecast is given by

$$E\left[\left(\overline{p}_{t+k} - \overline{p}_{t}\right)^2\right] = E\left[\left(\sum_{j=1}^{k-1}\sum_{i=1}^{n}\epsilon_{t+j,i} + \frac{1}{n}\sum_{s=1}^{n}(n+1-s)\cdot\epsilon_{t+k,s} - \frac{1}{n}\sum_{l=1}^{n-1} l\cdot\epsilon_{t,l}\right)^2\right] \tag{8}$$

$$= \left((k-1)\cdot n + \frac{(n+1)\cdot(2n+1)}{6n} + \frac{(n-1)\cdot(2n-1)}{6n}\right)\sigma_\epsilon^2, \tag{9}$$

which is unambiguously larger since $\frac{(n-1)\cdot(2n-1)}{6n} > 0 \; \forall \; n > 1$. $\square$

In this setting, the forecast improvements from using the last closing price no-change forecast take on a distinct pattern. From equations 7 and 9, it follows that the percent improvement in the MSPE ratio from using the closing-price forecast is

$$\frac{E\left[\left(\overline{p}_{t+k} - \overline{p}_{t}\right)^2\right] - E\left[\left(\overline{p}_{t+k} - p_{t,n}\right)^2\right]}{E\left[\left(\overline{p}_{t+k} - \overline{p}_{t}\right)^2\right]} = \frac{(n-1)\cdot(2n-1)}{6(k-1)\cdot n^2 + (n+1)\cdot(2n+1) + (n-1)\cdot(2n-1)}. \tag{10}$$

The theoretical MSPE improvements only depend on the forecast horizon, k, and on the number of high-frequency observations, n, that are averaged over in a given low-frequency time interval. Because the forecasting error increases with the forecasting horizon, but the forecasting gains remain constant, the relative gains are largest for the one-step-ahead prediction and decrease to zero as the horizon widens.

The closed-form expression in equation 10 can be used to derive theoretical forecast improvements for a typical macroeconomic series that is based on monthly average data with $n \approx 21$. The theoretical MSPE ratios of the end-of-month no-change forecasts relative to the monthly average no-change forecasts are 0.54, 0.88, 0.95, 0.97, 0.99 at horizons of 1, 3, 6, 12 and 24 months, respectively. The relative MSPE ratios are lowest for the one-month-ahead forecast and decrease with the forecasting horizon. Larger improvements are obtained for quarterly data (n=63) with values of 0.51, 0.80, 0.91, and 0.96 at horizons of 1, 2, 4 and 8 quarters, respectively. In our empirical section, we evaluate how closely this pattern is reflected in the actual data of the macroeconomic variables.

Proposition 2 has immediate consequences for tests of predictability. In practice, most forecasters evaluate the accuracy of their forecasts based on a criterion that involves relative MSPEs, such as the MSPE ratio. When the average price is used as a benchmark, the forecaster is more likely to find reductions in the MSPE from the use of alternative forecasts and to conclude that prices are predictable. This is the case even if prices are entirely unpredictable, as in our example.

The second effect arises when average prices are also used to estimate models, as is the case with many existing econometric forecasts for aggregated series. Adjacent changes in average prices are mechanically correlated even if the underlying data-generating process is unpredictable Working (1960). This is the case in our setting, where the change in the average price is

$$\overline{p}_t - \overline{p}_{t-1} = \frac{1}{n}\left[\sum_{i=1}^{n} i \cdot \epsilon_{t-1,i} + \sum_{j=1}^{n}(n+1-j) \cdot \epsilon_{t,j}\right], \tag{11}$$

such that each innovation, $\epsilon_{s,k}$, enters both $\Delta\overline{p}_s$ and $\Delta\overline{p}_{s-1}$. Hence, an econometric model that is based on average prices and allows for serial correlation might show statistically significant coefficients and exhibit improved forecasts relative to the no-change forecast $\bar{p}_t$, even in the critical case in which prices are unpredictable. On the contrary, adjacent changes in end-of-month closing prices are uncorrelated under the random walk hypothesis. Thus, both effects can be remedied by using end-of-period observations to construct both no-change and model-based forecasts.

Our results show that the optimal no-change benchmark should be used to test the classical version of the random walk hypothesis under which the underlying data is generated by a random walk. By contrast, comparisons with the conventional no-change forecast implicitly test the hypothesis that the *aggregated* data follows a random walk. While this hypothesis might seem perfectly sensible at first glance, our next proposition shows that the random walk assumption for the aggregated data reflects a rather contrived hypothesis.

Proposition 3. *Whenever the underlying data is generated by an ARIMA(p, d, q) model, the aggregated data does not follow a random walk.*

Proof: The proposition follows readily from the result that temporal aggregation converts an underlying ARIMA(p, d, q) process into an ARIMA(p, d, q^*) process, where $q^* = NLI(p + d + 1 - (q - p - d - 1)/n)$ and $NLI(\cdot)$ is the next-lowest integer function (see, e.g., Rossana and Seater, 1995)[8] and the fact that $p > 0 \ \vee \ q^* > 0$ for all non-negative values of p, d and q.$\square$

An immediate consequence of this result is that a temporally aggregated series is predictable for *any* ARIMA representation of the underlying data-generating process, including the random walk as a special case. The conventional no-change benchmark thus reflects an implicit null hypothesis that not only rules out the random walk model but also a much broader class of plausible time-series models. It is therefore unsurprising that empirical applications to aggregated data often find significant forecast improvements over the conventional no-change forecast. However, such improvements are generally consistent with a broad range of data-generating processes and by themselves provide only limited insights into the practical usefulness of forecasting models or many economically relevant hypotheses.

3 Testing Against the Optimal Benchmark

We now show how forecast comparisons against the optimal benchmark can be implemented in practice. Consider a typical set-up where the forecaster uses the available information at the end of each month to form their predictions for the following months. The forecaster's objective is to predict the average monthly real series k periods ahead, R_{t+k}, given time t information. The average monthly real series is a deflated simple average, $R_t = \frac{1}{n}\sum_{i=1}^{n} p_{t,i}/CPI_t$, where $p_{t,i}$ is the nominal daily price on day i of month t.

Under the random walk null hypothesis, the conditional expectation of R_{t+k} is the last available disaggregated observation, $E_t(R_{t+k}) = R_{t,n}$. The series of real end-of-period observations, $R_{t,n}$, is constructed using $R_{t,n} = p_{t,n}/CPI_t$. This allows us to define the forecast criteria in relation to the optimal forecast under the null of no predictability.

We propose two common forecast criteria in relation to the end-of-period no-change forecast: the MSPE ratio and the success ratio for directional accuracy. The MSPE ratio for the k-steps-ahead forecast, $MSPE_k^{ratio}$, is calculated as the ratio of the MSPE of the model-based forecast to

[8]The next-lowest integer function, $NLI(\cdot)$, means that the expression in brackets is rounded down to the next-lowest integer if it is not already an integer.

the MSPE of the end-of-period no-change forecast:

$$MSPE_k^{ratio} = \frac{\sum_{q=1}^{Q}(R_{q+k} - \hat{R}_{q+k|q})^2}{\sum_{q=1}^{Q}(R_{q+k} - R_{q,n})^2}, \tag{12}$$

where $\hat{R}_{q+k|q}$ is the conditional model forecast for the k-steps-ahead aggregated observation, R_{q+k}, and $R_{q,n}$ is the end-of-period observation for all periods of the evaluation sample, $q \in Q$. The null hypothesis of an equal MSPE for the model-based forecast relative to the end-of-period no-change forecast is tested following Diebold and Mariano (1995) and compared against standard normal critical values as recommended by Carriero et al. (2015).

Directional accuracy is assessed using the mean directional accuracy, also referred to as the success ratio. It describes the fraction of times the forecasting model can correctly predict the change in direction of the series of interest. For comparisons with the optimal no-change benchmark, we compute the success ratio for forecast horizon k, SR_k, as

$$SR_k = \frac{1}{Q}\sum_{q=1}^{n} \mathbb{1}_{sgn(R_{q+k}-R_{q,n})=sgn(\hat{R}_{q+k|q}-R_{q,n})}, \tag{13}$$

where sgn is a sign function and $\mathbb{1}$ is an indicator function. The version of the test in equation 13 differs from the usual application, as the direction of the change is compared against the end-of-period observation and not the monthly average observation. The test statistic is calculated following Pesaran and Timmermann (2009).

4 Application to Forecasts of Macroeconomic Variables

To highlight the generality of the findings, we analyze forecasts of a range of key macroeconomic variables that have attracted considerable attention among forecasters. We focus on series that are expressed as aggregated data in previous forecasting studies, which is typically the case with real variables. Our data comprises the real prices of Brent Crude, US retail gasoline, and copper. Non-commodity variables include short- and long-term real interest rates (the yields of the U.S. 3-Month Treasury Bill and the U.S. 10-Year Treasury Note) as well as the USD/CAD exchange rates. Finally, we also include a measure of real stock prices based on Standard and Poor's 500 index for comparison.[9] Consistent with the literature, our empirical application concentrates on

[9] Forecasts of aggregated real stock prices are less frequent in the literature, which typically focuses on disaggregated data and stock returns. Nonetheless, since the dynamics of stock prices are relatively well-understood, the forecasting

real-time forecasts of monthly average observations.

To study both the effect of introducing the optimal no-change benchmark and the effect of estimating models with end-of-period observations, we provide four distinct forecasting exercises. The first examines the relative forecast performance of the different no-change forecasts. This is done by comparing the accuracy of the end-of-month no-change forecasts against that of the average monthly price no-change forecasts. The second evaluates the forecast performance of ARIMA models that are estimated with aggregated data, where the performance is evaluated against the conventional benchmark. This represents the conventional setting in the existing literature.[10] The third and the fourth exercises evaluate the performance of ARIMA models that are estimated with end-of-period data against the conventional and optimal benchmarks, respectively. Jointly, these exercises allow us to isolate the effects of introducing the optimal benchmark and of estimating forecast models using end-of-period observations.

All models are estimated recursively, and dynamic forecasts are computed out-of-sample. Following the previous literature, and consistent with tests of stationarity, the models for real crude oil prices are estimated in logs, the models for real interest rates are estimated in differences, and the models for all other real series are estimated in growth rates. The forecasts that are estimated in log-levels are converted to real levels using

$$\hat{R}^j_{t+k|t} = exp\left(\hat{r}^j_{t+k|t}\right), \tag{14}$$

where j indicates the series and $\hat{R}^j_{t+k|t}$ is the forecast of the monthly average real price of variable j in levels, $\hat{r}^j_{t+k|t}$ is the equivalent in logs, and exp is the exponential function. For models estimated in percent change, forecasts for the net growth rate between period t and period $t+k$, $\hat{g}^j_{t+k|t}$, are mapped into levels via

$$\hat{R}^j_{t+k|t} = (1 + \hat{g}^j_{t+k|t})R^j_t. \tag{15}$$

Likewise, for models estimated in differences, forecasts for the difference between period t and period $t+k$, $\hat{d}^j_{t+k|t}$, are mapped into levels via

$$\hat{R}^j_{t+k|t} = \hat{d}^j_{t+k|t} + R^j_t. \tag{16}$$

results for this series constitute a useful reference point for the results for the other series.

[10] Using simple ARIMA models as examples allows us to focus the attention on the methodological issues associated with the current practice. A proper evaluation of other commonly applied forecasting approaches for the series are beyond the scope and purpose of this paper.

All forecasts are implemented using real-time data. The price index is the seasonally adjusted U.S. consumer price index obtained from the FRASER database of the Federal Reserve Bank of St. Louis and the real-time database of the Philadelphia Federal Reserve. For the USD/CAD real exchange rate, the seasonally adjusted Canadian consumer price index is obtained from Statistics Canada and the pseudo real-time data is constructed using the latest release and maintaining a one-month release lag. The missing real-time observations for the CPI are nowcasted using the average historical growth rate.

Information on oil and gasoline prices were obtained from the U.S. Energy Information Administration (EIA), information on the nominal USD/CAD real exchange rate and interest rates are from the FRED database of the Federal Reserve Bank of St. Louis, and on Copper prices from Bloomberg. A detailed description of the nominal averaged and high-frequency data is provided in a separate online appendix. End-of-period observations refer to the observation on the last trading day of each month. The only exception is for retail gasoline prices, which are sampled on the Monday of each week. Hence, the end-of-period price of gasoline is the price on the last Monday of each month. All high-frequency observations are assumed to be available in real time and are not subject to historical revisions. This assumption is also implicit in the existing literature, which treats nominal monthly observations for these series as available in real time.

As is standard for oil-price forecasts, oil prices are backcasted to 1973 to estimate the models. For all other series except retail gasoline, where the data starts in 1991M2, we use data from 1990M2 onwards to estimate the models. The forecast evaluation period is 1992M1-2021M1 for the comparisons of the two no-change benchmarks. For the evaluations of the forecast performance of the ARIMA models, which require an estimation period, the evaluation period is 2000M1-2021M1.

5 Forecast Results

Our first set of results in Table 1 compare the forecast performance of the end-of-month no-change forecast against that of the no-change forecast based on averaged data. The directional accuracy of the end-of-month no-change forecast is computed in the conventional way by taking the sign $+$ ($-$) whenever the value of the end-of-month observation is above (below) the monthly average. For the one-step-ahead predictions, the end-of-month no-change forecasts significantly outperform the monthly average no-change forecasts at the one percent significance level for all series. At this horizon, the MSPE improvements of the forecasts range from 37 percent for the USD/CAD real

Table 1. Relative Forecast Performance of the Two No-Change Forecasts

Series	Brent	Copper	Retail Gasoline	3 Month T-Bill	10 Year T-Note	USD/CAD	S&P 500
Horizon				MSPE Ratio			
1	0.58 (0.000)	0.55 (0.000)	0.47 (0.000)	0.56 (0.000)	0.53 (0.000)	0.63 (0.000)	0.61 (0.022)
3	0.92 (0.017)	0.89 (0.007)	0.86 (0.002)	0.84 (0.000)	0.91 (0.002)	0.92 (0.004)	0.91 (0.028)
6	0.98 (0.123)	0.96 (0.027)	0.98 (0.131)	0.92 (0.000)	0.96 (0.039)	0.96 (0.020)	0.97 (0.125)
9	0.99 (0.294)	1.00 (0.404)	1.03 (0.905)	0.93 (0.000)	0.96 (0.031)	0.98 (0.095)	0.99 (0.364)
12	0.98 (0.170)	0.99 (0.321)	1.00 (0.491)	0.95 (0.000)	0.96 (0.006)	0.96 (0.017)	0.97 (0.006)
24	1.00 (0.443)	0.99 (0.314)	1.01 (0.716)	0.99 (0.040)	1.00 (0.366)	0.98 (0.030)	0.98 (0.020)
				Success Ratio			
1	0.72 (0.000)	0.71 (0.000)	0.76 (0.000)	0.75 (0.000)	0.69 (0.000)	0.73 (0.000)	0.76 (0.000)
3	0.57 (0.005)	0.60 (0.000)	0.62 (0.000)	0.65 (0.000)	0.59 (0.000)	0.59 (0.000)	0.62 (0.000)
6	0.56 (0.013)	0.55 (0.020)	0.55 (0.008)	0.66 (0.000)	0.55 (0.058)	0.59 (0.000)	0.59 (0.000)
9	0.55 (0.035)	0.57 (0.001)	0.50 (0.416)	0.65 (0.000)	0.55 (0.117)	0.58 (0.000)	0.58 (0.000)
12	0.59 (0.000)	0.56 (0.004)	0.56 (0.013)	0.64 (0.000)	0.58 (0.008)	0.59 (0.000)	0.59 (0.000)
24	0.56 (0.015)	0.54 (0.061)	0.51 (0.367)	0.54 (0.069)	0.53 (0.248)	0.52 (0.211)	0.53 (0.135)

Note: The table displays the MSPE of the end-of-month no-change forecasts relative to that of the no-change forecast computed from monthly average data, as well as the success ratio of a directional forecast that is based on the value of the end-of-month observation relative to the value of the average monthly observation. Forecasts are computed for real variables in levels. The evaluation period is 1992M1–2021M1. *T*- refers to U.S. Treasury. The brackets show the p-values for serial-dependence-robust tests of Diebold and Mariano (1995) for the null hypothesis of equal MSPEs and of Pesaran and Timmermann (2009) for the null of no directional accuracy.

exchange rate to 53 percent for U.S. 3-Month Treasury Bills. Similarly, improvements in the success ratio range from 38 percent to 52 percent. The forecast gains decrease as the forecasting horizon increases, but in some cases they are still statistically significant at the five percent significance level for the 24-months-ahead forecast. Both the magnitude of the forecast gains and their decay in terms of their relative forecasting performance as the horizon increases is quantitatively similar to the theoretical prediction of the pure random walk model for daily prices in Section 2. This shows that the choice of the no-change benchmark matters in practice.

It can be shown that the forecast improvements are driven by a smaller bias rather than a smaller variance of the no-change forecast computed from the end-of-period observation. For all of our series, the empirical variances of the end-of-period observations are statistically indistinguishable from the variances of the aggregated observations. On the other hand, the empirical biases in the end-of-period observations are smaller than those of the averaged observations, which indicates that reductions in these biases drive our results. Again, this pattern is consistent with a random walk model for daily observations.

A potential benefit of averaging the data is that averaging reduces the random noise arising from measurement error or market-microstructure dynamics. However, the large forecast improvements we obtain from the use of end-of-period observations instead of averaged observations indicate that this form of random noise in the daily series is largely irrelevant in practice. The result is intuitive given that most of our series are market-based price measures and hence not prone to measurement error. Moreover, it is well-known that for equity returns market-microstructure noise can safely be ignored at the daily frequency (Hansen and Lunde, 2006). Our results suggest that this is also the case for many macro-financial series that are measured in levels.

Table 2. Real-Time Forecasts from Models Estimated with Monthly Average Observations, Evaluated against the Conventional Benchmark

Series	Brent	Copper	Retail Gasoline	3 Month T-Bill	10 Year T-Note	USD/CAD	S&P 500
Horizon				MSPE Ratio			
1	0.93 (0.162)	0.94 (0.210)	0.82 (0.053)	0.78 (0.031)	0.97 (0.214)	0.89 (0.001)	1.00 (0.429)
3	0.97 (0.378)	1.01 (0.568)	0.91 (0.060)	0.79 (0.085)	0.97 (0.226)	1.01 (0.584)	1.04 (0.746)
6	0.96 (0.391)	1.06 (0.918)	0.89 (0.043)	0.77 (0.052)	0.95 (0.200)	1.02 (0.771)	1.07 (0.773)
9	0.96 (0.399)	1.13 (0.970)	1.03 (0.643)	0.76 (0.028)	0.94 (0.224)	1.06 (0.969)	1.14 (0.864)
12	0.96 (0.397)	1.12 (0.965)	1.20 (0.979)	0.82 (0.047)	0.93 (0.217)	1.07 (0.961)	1.17 (0.857)
24	0.92 (0.333)	1.24 (1.000)	1.39 (0.996)	1.06 (0.748)	0.90 (0.248)	1.18 (0.996)	1.32 (0.879)
				Success Ratio			
1	0.48 (0.481)	0.62 (0.000)	0.64 (0.000)	0.68 (0.000)	0.58 (0.011)	0.60 (0.001)	0.57 (0.293)
3	0.50 (0.327)	0.53 (0.234)	0.63 (0.000)	0.68 (0.000)	0.55 (0.344)	0.54 (0.186)	0.59 (0.619)
6	0.56 (0.042)	0.49 (0.683)	0.66 (0.000)	0.68 (0.000)	0.58 (0.266)	0.55 (0.136)	0.63 (0.783)
9	0.55 (0.099)	0.48 (0.712)	0.53 (0.614)	0.66 (0.000)	0.60 (0.591)	0.51 (0.461)	0.65 (0.901)
12	0.59 (0.040)	0.43 (0.927)	0.53 (0.590)	0.63 (0.021)	0.65 (1.000)	0.49 (0.581)	0.69 (1.000)
24	0.68 (0.000)	0.40 (0.953)	0.43 (1.000)	0.49 (0.954)	0.59 (1.000)	0.38 (0.977)	0.73 (1.000)
ARIMA	(2,0,0)	(1,1,0)	(12,1,0)	(3,1,0)	(5,1,0)	(1,1,0)	(1,1,0)

Note: Recursive, dynamic, out-of-sample forecasts of aggregated real variables in levels. Forecast criteria are compared against the no-change forecast based on aggregated monthly data. Bold values indicate significant improvements over the optimal no-change forecast at the five percent significance level. Models are estimated using aggregated monthly real observations beginning in 1990M2 (1973M1 for crude oil series), the evaluation period is 2000M1–2021M1. *T-* refers to U.S. Treasury. The brackets show the p-values for serial-dependence-robust tests of Diebold and Mariano (1995) for the null hypothesis of equal MSPEs and of Pesaran and Timmermann (2009) for the null of no directional accuracy.

Table 2 provides an example of a standard forecast exercise as it would have been implemented in the existing literature. It reports the forecast from the ARIMA(p, d, q) models that are estimated with average monthly prices and evaluated against the monthly average no-change forecast.[11] In some cases, such as short-run horizons for the exchange rate, three-month interest rates and at longer horizons for the crude oil series, the model-based forecasts significantly improve upon the conventional benchmark. Typically, these results would be used to reject the random walk hypothesis and argue for the practical usefulness of the models. However, this conclusion is misleading. None of the model-based forecasts improve upon the optimal no-change benchmark and some are

[11]For all ARIMA models, the lag length is chosen according to the best out-of-sample forecast performance. The results are qualitatively robust to the use of alternative in-sample model selection criteria.

Table 3. Real-Time Forecasts from Models Estimated with End-of-Month Observations, Evaluated against the Conventional Benchmark

Series	Brent	Copper	Retail Gasoline	3 Month T-Bill	10 Year T-Note	USD/CAD	S&P 500
Horizon				MSPE Ratio			
1	0.58 (0.000)	0.57 (0.001)	**0.41 (0.000)**	0.59 (0.001)	0.65 (0.002)	0.64 (0.000)	0.63 (0.089)
3	0.91 (0.199)	0.92 (0.075)	0.85 (0.003)	0.71 (0.016)	0.98 (0.388)	0.95 (0.037)	0.89 (0.129)
6	0.95 (0.364)	1.03 (0.687)	1.04 (0.809)	0.75 (0.032)	0.96 (0.384)	0.99 (0.377)	0.94 (0.318)
9	0.97 (0.419)	1.10 (0.921)	1.17 (0.962)	**0.75 (0.019)**	0.94 (0.358)	1.04 (0.842)	1.00 (0.494)
12	0.97 (0.431)	1.13 (0.933)	1.20 (0.964)	0.81 (0.033)	0.92 (0.327)	1.05 (0.862)	0.98 (0.452)
24	0.97 (0.429)	1.27 (0.999)	1.39 (0.997)	1.05 (0.684)	1.02 (0.528)	1.16 (0.990)	1.01 (0.516)
				Success Ratio			
1	0.71 (0.000)	0.72 (0.000)	0.72 (0.000)	**0.74 (0.000)**	0.70 (0.000)	0.74 (0.000)	0.66 (0.000)
3	0.51 (0.350)	0.62 (0.000)	0.62 (0.000)	**0.69 (0.000)**	0.61 (0.009)	0.58 (0.001)	0.64 (0.119)
6	**0.55 (0.104)**	0.56 (0.108)	0.58 (0.010)	**0.68 (0.000)**	0.60 (0.158)	0.58 (0.018)	0.65 (0.576)
9	**0.54 (0.211)**	0.53 (0.289)	0.52 (0.431)	**0.66 (0.000)**	0.64 (0.000)	0.55 (0.120)	0.68 (1.000)
12	**0.56 (0.123)**	0.49 (0.543)	0.52 (0.169)	0.62 (0.018)	0.69 (0.000)	0.55 (0.126)	0.72 (1.000)
24	**0.61 (0.028)**	0.42 (0.902)	0.42 (0.801)	0.52 (0.849)	0.66 (0.000)	0.42 (0.949)	0.76 (1.000)
ARIMA	(2,0,0)	(1,1,0)	(1,1,1)	(4,1,0)	(5,1,0)	(0,1,1)	(1,1,0)

Note: Recursive, dynamic, out-of-sample forecasts of aggregated real variables in levels. Forecast criteria are compared against the no-change forecast based on aggregated monthly data. Bold values indicate significant improvements over the optimal no-change forecast at the five percent significance level. Models are estimated using end-of-month real observations beginning in 1990M2 (1973M1 for crude oil series), the evaluation period is 2000M1–2021M1. *T*- refers to U.S. Treasury. The brackets show the p-values for serial-dependence-robust tests of Diebold and Mariano (1995) for the null hypothesis of equal MSPEs and of Pesaran and Timmermann (2009) for the null of no directional accuracy.

even considerably worse at short horizons. This shows that the conventional implementation of forecast comparisons could easily result in spurious inference.

Table 3 investigates whether estimating the models with end-of-period observations can help to recover the information loss that is introduced by aggregation. It shows that the forecast performance does indeed improve considerably. For all series including stock prices, model-based forecasts outperform average no-change forecasts at short forecasting horizons. In the majority of cases, these forecasts are significantly better than those from models that are estimated with averaged prices (Table 3). This result is not obvious ex-ante because the ARIMA models allow for a flexible autocorrelation structure that should in principle help to recover the information loss, even when the models are estimated using averaged data. However, our results show that, in practice, aggregation does indeed greatly diminish the out-of-sample forecasting performance of model-based forecasts.[12] These findings suggest that substantial gains in model-based forecasts can

[12] Consistent with Rossana and Seater (1995), who show that time-averaging changes the dynamics of a series, the optimal lag-length of ARIMA models changes depending on whether they are estimated with end-of-month series or monthly average series.

be obtained from estimating models that have traditionally been estimated with aggregated data.[13] They also suggest that forecasters can maintain the same low-frequency models while incorporating informational content from disaggregated observations.

Table 4. Real-Time Forecasts of Models Estimated with End-of-Month Observations and Evaluated Against the Optimal Benchmark

Series	Brent	Copper	Retail Gasoline	3 Month T-Bill	10 Year T-Note	USD/CAD	S&P 500
Horizon				MSPE Ratio			
1	0.99 (0.432)	1.06 (0.918)	**0.88 (0.019)**	1.04 (0.662)	1.10 (0.973)	1.03 (0.952)	1.02 (0.729)
3	0.99 (0.436)	1.05 (0.901)	1.01 (0.621)	0.85 (0.115)	1.01 (0.623)	1.02 (0.878)	1.00 (0.478)
6	0.96 (0.397)	1.09 (0.926)	1.07 (0.917)	0.82 (0.073)	0.97 (0.289)	1.04 (0.930)	0.98 (0.421)
9	0.96 (0.398)	1.13 (0.952)	1.14 (0.940)	**0.80 (0.037)**	0.96 (0.292)	1.05 (0.948)	1.00 (0.498)
12	0.96 (0.403)	1.18 (0.965)	1.18 (0.937)	0.85 (0.060)	0.95 (0.290)	1.08 (0.968)	1.02 (0.544)
24	0.92 (0.329)	1.36 (0.998)	1.36 (0.980)	1.07 (0.767)	0.93 (0.312)	1.18 (0.996)	1.03 (0.545)
				Success Ratio			
1	0.52 (0.105)	0.47 (0.823)	0.54 (0.098)	**0.61 (0.000)**	0.53 (0.110)	0.53 (0.264)	0.62 (0.110)
3	0.48 (0.453)	0.48 (0.837)	0.54 (0.468)	**0.65 (0.000)**	0.53 (0.397)	0.49 (0.584)	0.66 (1.000)
6	**0.57 (0.022)**	0.51 (0.733)	0.55 (0.891)	**0.67 (0.000)**	0.56 (0.115)	0.47 (0.687)	0.66 (1.000)
9	**0.58 (0.019)**	0.47 (0.906)	0.58 (0.807)	**0.65 (0.000)**	0.60 (1.000)	0.49 (0.556)	0.68 (1.000)
12	**0.59 (0.043)**	0.46 (0.836)	0.60 (0.238)	0.59 (0.088)	0.64 (1.000)	0.50 (0.487)	0.73 (1.000)
24	**0.67 (0.000)**	0.46 (0.846)	0.53 (0.627)	0.52 (0.934)	0.59 (1.000)	0.41 (0.919)	0.76 (1.000)
ARIMA	(2,0,0)	(1,1,0)	(1,1,1)	(4,1,0)	(5,1,0)	(0,1,1)	(1,1,0)

Note: Recursive, dynamic, out-of-sample forecasts of aggregated real variables in levels. Forecast criteria are compared against the optimal no-change forecast. Models are estimated using end-of-month real observations beginning in 1990M2 (1973M1 for crude oil series), the evaluation period is 2000M1–2021M1. *T*- refers to U.S. Treasury. The brackets show the p-values for serial-dependence-robust tests of Diebold and Mariano (1995) for the null hypothesis of equal MSPEs and of Pesaran and Timmermann (2009) for the null of no directional accuracy.

Finally, we put it all together by comparing the model-based forecasts when estimated with end-of-period observations against the optimal no-change forecasts (Table 4). Relative to Table 3, when these forecasts are compared to the conventional benchmark, the evidence in favor of model-based forecasts are much weaker. Only forecasts at selected horizons for retail gasoline, short-term interest rates and crude oil show significant improvements over the optimal benchmark. Moreover, even compared to the traditional exercise of comparing models estimated with aggregated data against no-change forecasts that are based on aggregated data (Table 2), the evidence in favor of model-based forecasts is much weaker. This highlights once again that the choice of the benchmark matters for the evaluation of models and for assessing the predictability of the series more generally.

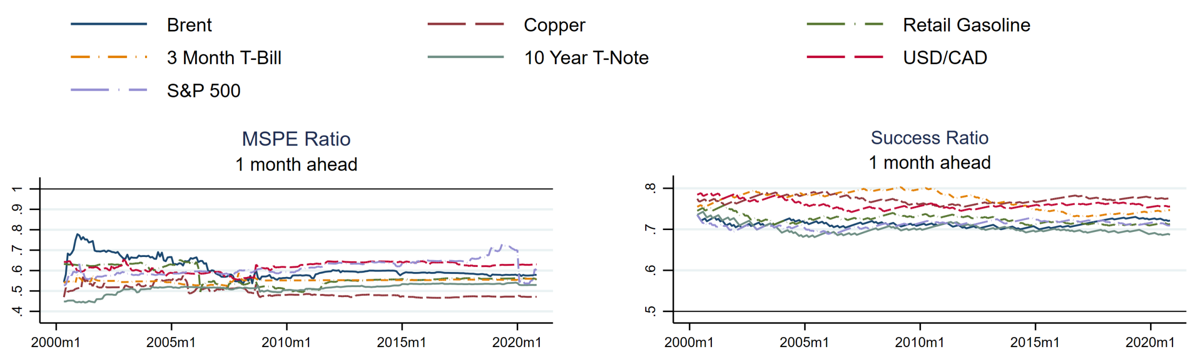
[13] This finding can be shown to extend to wider classes of models, including vector-autoregressions Benmoussa et al. (2020) and estimation-free forecasts based on futures curves Ellwanger and Snudden (2021).

6 Extensions and Robustness

Our results are remarkably robust to alternative modeling choices, including using ex-post revised data instead of real-time data and using nominal series instead of real series. Moreover, our results are robust to alternative methods of deflating the end-of-period observations, including alternative methods of nowcasting the CPI. This result is unsurprising as fluctuations in the CPI deflator are generally small compared to the fluctuations in nominal observations.

Further, it can be shown that the end-of-period no-change forecast is also superior to the aggregate no-change forecast for quarterly and annual data, which are of primary interest to policymakers (Baumeister and Kilian, 2014, 2015). Consistent with the theoretical predictions in Section 2, the improvements in the MSPE ratio at the one-step-ahead prediction are even larger for quarterly and annual aggregations.

Figure 1. Recursively Estimated Relative Forecast Performance of the Two No-Change Forecasts



Note: Recursively updated MSPE ratio of the end-of-month no-change forecasts relative to that of the no-change forecast computed from monthly average data (left-hand panel) and the recursively updated success ratio of a directional forecast that is based on the value of the end-of-month observations relative to the value of the average monthly observations (right panel). Forecasts are computed for real variables in levels. *T*- refers to U.S. Treasury. The estimation period is 1992M1–2021M1. The first 100 months are dropped to reduce starting-point effects.

The performance of forecasts is often sensitive to the sample period. To investigate the robustness of the performance of the optimal benchmark forecast over time, the evolution of the recursively updated MSPE ratio and the directional accuracy relative to the conventional benchmark are reported in Figure 1. The forecasting improvements of the end-of-month no-change forecast relative to the monthly average no-change forecast prices are very stable. For all series, the optimal benchmark outperforms in terms of the MSPE precision and the directional accuracy for 100 percent of the sample, including the Great Recession, the Zero Lower Bound period and the COVID-19 episode. The stability in the forecast gains stands in stark contrast to the evidence from conventional forecasting approaches. It further supports our hypothesis that averaging leads to a loss of information that is mechanical and largely independent of the series' behavior over specific episodes.

7 Conclusion

Our paper demonstrated that temporal aggregation has several far-reaching consequences for the predictability of macroeconomic series. For a large class of data-generating processes, these series are predictable by construction. Moreover, conventional tests of predictability tempt researchers to draw spurious conclusions about predictability that have no basis in the underlying data. This is particularly relevant for the standard practice of evaluating forecasts against the performance of conventional no-change forecasts. We showed that improvements over this benchmark are to be expected even when the data is generated by a random walk. Contrary to existing practice, these improvements should not be cited as evidence against the random walk model or to advocate for the practical appeal of particular forecasting models.

We have also shown how the introduction of end-of-period observations can help to restore meaningful forecast comparisons. Under the random walk null hypothesis, this simple modification of the conventional test leads to a sizeable correction in the MSPE of the benchmark forecast of up to 45 percent. Moreover, similarly large forecasting gains can be obtained by estimating models with end-of-period observations instead of aggregated observations.

We have highlighted the relevance of our findings for a set of real macroeconomic variables. For all series, the gains from using end-of-period observations in the forecasts of these variables are remarkably close to the theoretical predictions from a random walk model for the underlying data. For short-horizon forecasts, these gains are more sizeable than the typical gains obtained from introducing new predictor variables or models for aggregated variables. In contrast to conventional forecasts, they are also astonishingly stable across different sample periods, including the Great Recession, the Zero Lower Bound period and the COVID-19 episode.

Two direct corollaries of our results could help shed new light on the dynamics and economic consequences of expectations. First, since end-of-period observations provide more-accurate real-time forecasts for aggregated variables, they can also provide better measures of the historical expectations for these variables (Baumeister and Kilian, 2016). Second, researchers should be careful when analyzing surveys of expectations that are presented in temporally aggregated form. A prominent example is the Michigan Survey of Consumers, which is typically analyzed at monthly or quarterly frequencies (see, e.g., Anderson et al., 2011, 2013; Malmendier and Nagel, 2016). Our results suggest that the timing of the survey within a given month could have large effects on individual responses. Moreover, the fact that such responses are temporally aggregated could make averaged expectations appear to be predictable, even when each respondent forms expectations according to the random walk.

Our results should concern all forecasters who use aggregated data to construct their forecasts. Incorporating information from end-of-period observations can yield large gains even in the context of the simple models many practitioners prefer. The gains are likely to occur in any setting where forecasters work with aggregated data and the underlying series are persistent, which we have shown to be the case in many macroeconomic applications. Moreover, both forecasts and forecast evaluations are easily implementable in real time and within models that are typically applied to lower-frequency data. Taken together, our findings highlight the need to re-evaluate established forecasting approaches through the lens of aggregation.

References

Alquist, R., Bhattarai, S., and Coibion, O. (2020). Commodity-price comovement and global economic activity. *Journal of Monetary Economics*, 112:41–56.

Alquist, R., Kilian, L., and Vigfusson, R. (2013). Forecasting the price of oil. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, chapter 8, pages 427–507. Elsevier, 1 edition.

Anderson, S. T., Kellogg, R., and Sallee, J. M. (2013). What do consumers believe about future gasoline prices? *Journal of Environmental Economics and Management*, 66(3):383–403.

Anderson, S. T., Kellogg, R., Sallee, J. M., and Curtin, R. T. (2011). Forecasting gasoline prices using consumer surveys. *American Economic Review*, 101(3):110–14.

Andreou, E., Ghysels, E., and Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 31(2):240–251.

Baumeister, C., Guérin, P., and Kilian, L. (2015). Do high-frequency financial data help forecast oil prices? the midas touch at work. *International Journal of Forecasting*, 31(2):238–252.

Baumeister, C. and Kilian, L. (2012). Real-time forecasts of the real price of oil. *Journal of Business & Economic Statistics*, 30(2):326–336.

Baumeister, C. and Kilian, L. (2014). What central bankers need to know about forecasting oil prices. *International Economic Review*, 55(3):869–889.

Baumeister, C. and Kilian, L. (2015). Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business & Economic Statistics*, 33(3):338–351.

Baumeister, C. and Kilian, L. (2016). A general approach to recovering market expectations from futures prices with an application to crude oil. *Bank of Canada Staff Working Paper, 2016-18.*

Baumeister, C., Kilian, L., and Lee, T. K. (2014). Are there gains from pooling real-time oil price forecasts? *Energy Economics*, 46:S33–S43.

Baumeister, C., Kilian, L., and Lee, T. K. (2017). Inside the crystal ball: New approaches to predicting the gasoline price at the pump. *Journal of Applied Econometrics*, 32(2):275–295.

Benmoussa, A. A., Ellwanger, R., and Snudden, S. (2020). The new benchmark for forecasts of the real price of crude oil. *Bank of Canada Staff Working Paper, 2020-39.*

Carabenciov, I., Freedman, C., Garcia-Saltos, R., Laxton, D., Kamenik, O., and Manchev, P. (2013). Gpm6: The global projection model with 6 regions. *IMF Working Papers*, 2013(087).

Carriero, A., Clark, T. E., and Marcellino, M. (2015). Bayesian vars: specification choices and forecast accuracy. *Journal of Applied Econometrics*, 30(1):46–73.

Carriero, A., Clark, T. E., Marcellino, M., and Mertens, E. (2021). Forecasting with shadow-rate vars. *FRB of Cleveland Working Paper.*

Ca'Zorzi, M., Kolasa, M., and Rubaszek, M. (2017). Exchange rate forecasting with dsge models. *Journal of International Economics*, 107:127–146.

Chen, S.-L., Jackson, J. D., Kim, H., and Resiandini, P. (2014). What drives commodity prices? *American Journal of Agricultural Economics*, 96(5):1455–1468.

Conlon, T., Cotter, J., and Eyiah-Donkor, E. (2021). The illusion of oil return predictability: The choice of the data matters! *SSRN Working Paper.*

Cowles, A. (1960). A revision of previous conclusions regarding stock price behavior. *Econometrica*, 28(4):909–915.

Deaton, A. (1987). Life-cycle models of consumption: Is the evidence consistent with the theory? In *Advances in Econometrics: Volume 2: Fifth World Congress*, volume 2, page 121. Cambridge University Press.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Ellwanger, R. and Snudden, S. (2021). Futures prices are useful predictors of the spot price of crude oil. *LCERPA Working Paper 2021-4.*

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Funk, C. (2018). Forecasting the real price of oil-time-variation and forecast combination. *Energy Economics*, 76:288–302.

Garratt, A., Vahey, S. P., and Zhang, Y. (2019). Real-time forecast combinations for the oil price. *Journal of Applied Econometrics*, 34(3):456–462.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The midas touch: Mixed data sampling regression models. *Working Paper*.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

Issler, J. V., Rodrigues, C., and Burjack, R. (2014). Using common features to understand the behavior of metal-commodity prices and forecast them at different horizons. *Journal of International Money and Finance*, 42:310–335.

Johannsen, B. K. and Mertens, E. (2021). A time series model of interest rates with the effective lower bound. *Journal of Money, Credit & Banking*.

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past? *Journal of Econometrics*, 18(3):337–349.

Lütkepohl, H. (1984). Forecasting contemporaneously aggregated vector ARMA processes. *Journal of Business & Economic Statistics*, 2(3):201–214.

Malmendier, U. and Nagel, S. (2016). Learning from inflation experiences. *The Quarterly Journal of Economics*, 131(1):53–87.

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics*, 17(1):129–136.

Nelson, C. R. and Plosser, C. R. (1982). Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of Monetary Economics*, 10(2):139–162.

Pesaran, M. H. and Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485):325–337.

Rossana, R. J. and Seater, J. J. (1995). Temporal aggregation and economic time series. *Journal of Business & Economic Statistics*, 13(4):441–451.

Schwert, G. (1990). Indexes of us stock prices from 1802 to 1987. *The Journal of Business*, 63(3):399–426.

Snudden, S. (2018). Targeted growth rates for long-horizon crude oil price forecasts. *International Journal of Forecasting*, 34(1):1–16.

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika*, 59(3):525–531.

Wei, W. W. (1978). Some consequences of temporal aggregation in seasonal time series models. In *Seasonal analysis of economic time series*, pages 433–448. NBER.

Working, H. (1960). Note on the correlation of first differences of averages in a random chain. *Econometrica*, 28(4):916–918.

Online Data Appendix (Not intended for publication)

This appendix describes the construction of the data series used in the empirical exercises. It comprises the prices of Brent Crude, U.S. retail gasoline, and copper, as well as the U.S. refinery acquisition cost of imported crude oil (RAC). Non-primary commodity variables are the yields of the U.S. 3-Month Treasury Bill and the U.S. 10-Year Treasury Note, the USD/CAD exchange rate, and Standard and Poor's 500 stock index. An overview over the series is provided in Table A1.

Table A1. Descriptive Statistics

Real Series	Observations	Mean	Std. Dev.	Data Construction
Monthly Average				
Brent	576	24.13	12.32	Monthly average of daily data
Copper	371	2120.51	932.21	Monthly average of daily data
Retail gasoline	359	1.01	0.30	Monthly average of weekly data
3-month Treasury bill	371	1.66	2.27	Monthly average of daily data
10-year Treasury bond	371	3.36	1.99	Monthly average of daily data
USD/CAD	371	0.68	0.18	Monthly average of daily data
S&P 500	371	640.63	256.56	Monthly average of daily data
End of Month				
Brent	576	24.09	12.33	Last trading day
Copper	371	2131.11	941.05	Last trading day
Retail gasoline	360	1.01	0.30	Last Monday
3-month Treasury bill	371	1.65	2.26	Last trading day
10-year Treasury bond	371	3.34	2.00	Last trading day
USD/CAD	371	0.68	0.18	Last trading day
S&P 500	371	642.06	257.33	Last trading day
Consumer Price Index	575	0.32	0.34	Monthly

Note: Summary statistics for real variables in levels. The 2021M5 vintages are used for the consumer price index.

A1 Nominal series

This section describes the data sources and construction of the nominal series. Monthly observations for all series except U.S. retail gasoline prices are simple monthly averages of daily closing data.

Brent Crude Daily prices for Brent Crude were obtained from the U.S. Energy Information Administration (EIA)'s EIA Petroleum and other liquids. Daily prices are updated by the EIA in "Today in Energy." The data for the previous day is updated between 7:30 a.m. and 8:30 a.m. EST and is not subject to revisions, which justifies treating the data as observed in real time. The series of daily Brent prices provided by the EIA begin in 1987M5. For the estimation of the

ARIMA models, monthly Brent closing prices are backcasted to 1973M1, using the growth rate in the monthly Brent series from the IMF as in Benmoussa et al. (2020).

U.S. retail gasoline We obtained nominal weekly gasoline prices from Table 14 of the EIA's Weekly Petroleum Status Report. We used the series "Weekly U.S. Regular All Formulations Retail Gasoline Prices," for which the longest history is available. The EIA's weekly price corresponds to the price on the Monday of the week and is published on the same day, except on government holidays, for which the data is released on Tuesday.[14] We treat this price as being observable in real time.[15] Our end-of-period observation is the last observed weekly price; i.e., the price on the last Monday of each month. For the monthly prices, we took simple averages over all weekly observations (i.e., all Mondays) within a month. The monthly series obtained this way exactly corresponds to the series for "Regular Motor Gasoline, All Areas, Retail Price" available in the pen-ultimate column in the EIA's April 2021 Monthly Energy Review. This confirms that we are using the same method as the EIA to compute monthly data.

Copper Daily prices for copper are the cash price from the end of LME day Final Evening Evaluations in USD per Megatonne. The series were obtained from Bloomberg (Ticker LMCADY LME Comdty). We verified manually that the monthly averages obtained from this data are identical to the Copper data published by the IMF that is used in previous studies; the exception is for negligible differences in the earlier sample period.

Interest rates Daily data on interest rates were obtained from FRED. The short-term rate is the daily 3-Month Treasury Constant Maturity Rate (Ticker DGS3MO). The long-term rate is the 10-Year Treasury Constant Maturity Rate (Ticker DGS10).

USD/CAD exchange rate The daily Canadian Dollar vs. US Dollar exchange rate was obtained from FRED (Ticker DEXCAUS).

S&P 500 Daily prices for the S&P 500 Index were obtained from Bloomberg (Ticker SPX).

A2 Real series

All forecasts are implemented for real variables. The variables are originally obtained as nominal observations and converted into real terms using the US CPI. We use the same deflator to deflate monthly averages and end-of-month observations. The price index is the seasonally adjusted U.S. consumer price index obtained from the FRASER database of the Federal Reserve Bank of St.

[14] See `https://www.eia.gov/dnav/pet/TblDefs/pet_pri_gnd_tbldef2.asp`.

[15] Since the information set for our forecasts includes all data released in the current month, the only concern would be the rare case in which a public holiday falls on a Monday which is also the last day of the month.

Louis and the real-time database of the Philadelphia Federal Reserve. For the USD/CAD real exchange rate, the seasonally adjusted Canadian consumer price index is obtained from Statistics Canada and the pseudo-real-time data is constructed using the latest release and maintaining a one-month release lag. The missing real-time observations for the CPIs are nowcasted using the average historical growth rate.