nccc-134

APPLIED COMMODITY PRICE ANALYSIS, FORECASTING, AND MARKET RISK MANAGEMENT

Salmon Futures as Forecasts

Daumantas Bloznelis

Suggested citation format:

Bloznelis, D. 2025. “Salmon Futures as Forecasts.” Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, Chicago, Illinois, April 14-15, 2025. [http://www.farmdoc.illinois.edu/nccc134/paperarchive]

Salmon futures as forecasts

Daumantas Bloznelis[1]

School of Economics and Business

Norwegian University of Life Sciences

First version: March 5, 2020

This version: June 12, 2025

(minor corrections to the version of October 4, 2024)

Work in progress – comments welcome

[1] Mailing address: School of Economics and Business, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Aas, Norway. E-mail: daumantas.bloznelis@nmbu.no.

Abstract

Futures markets serve two main purposes, risk transfer and price forecasting. Both are relevant in the case of Norwegian farmed Atlantic salmon, as the spot price is highly volatile and hard to predict. However, the salmon futures market suffers from low liquidity, thus limiting the effectiveness of risk transfer. What about price forecasting? We consider futures prices as point forecasts of the future spot price. We evaluate them by statistical optimality criteria and find a downward bias that increases with the forecast horizon at a rate of over 10% a year. Simple benchmark forecasts have considerably lower bias, and some of them tend to beat the futures prices in terms of mean absolute error and mean squared error at horizons longer than nine months. However, most of the improvements cannot be claimed with high statistical confidence. The results should be of interest to decision makers who rely on salmon futures prices as point forecasts of the future spot price.

JEL codes:

G17 Financial Forecasting and Simulation

Q02 Commodity Markets

Q22 Fishery • Aquaculture

Keywords: forecasting, forecast evaluation, futures, spot price, salmon.

1. Introduction

Futures markets have two main functions, risk transfer and price forecasting (French, 1986). Both are of interest in the case of Norwegian farmed Atlantic salmon, as the spot price is highly volatile and hard to predict (Bloznelis, 2016c, 2018a; Øglend, 2013). While the risk transfer by salmon futures is limited due to low liquidity (Bloznelis, 2016a, p. 15, 18; 2018a), the market appears to excel in short-term price forecasting (Bloznelis, 2018b). However, for salmon producers, medium- and long-term forecasts might be of greater importance. Due to the length of the production cycle, some production decisions need to be made up to three years ahead of the sales (Mowi, 2020, pp. 51). In this paper, we examine the long-, medium- and short-term forecast quality of salmon futures prices. We track the location and scale of the forecast error distribution over time and seasons and across forecast horizons. We evaluate the forecasts by statistical criteria of absolute optimality and relative to simple benchmark forecasts. Our findings should be of interest to salmon market participants who rely on the futures prices as spot price forecasts and to current and prospective speculators eyeing an exotic seafood commodity accessible through a cash-settled futures contract.

The remainder of the paper is structured as follows. The rest of section 1 introduces the salmon spot and futures markets and reviews earlier attempts at price forecasting. Section 2 presents statistical criteria of forecast optimality and discusses how they will be applied to salmon futures prices viewed as point forecasts. Section 3 contains the empirical findings, and section 4 concludes.

1.1 The physical market

The market for Norwegian farmed Atlantic salmon emerged in 1980s when the salmon farming technology became commercially viable (Asche & Bjørndal, 2011, p. ix). The yearly production volume grew almost continuously until 2012, then plateaued between 2012 and 2018, and then started increasing again (Statistisk sentralbyrå, 2020a; Directorate of Fisheries, 2024). Between the launch of the Fish Pool salmon futures exchange in 2006 and the year 2022, salmon sales soared from 0.630 million metric tons worth NOK 15.6 billion to 1.565 million metric tons worth NOK 102.5 billion (Statistisk sentralbyrå, 2020a; Statistisk sentralbyrå, 2020b; Directorate of Fisheries, 2024); see Figure 1.1.a. Over the same period, the yearly average price fluctuated from NOK 21/kg in 2008 to NOK 65/kg in 2022 (Statistisk sentralbyrå, 2020a; Directorate of Fisheries, 2024) with large variations within each year; see Figure 1.1.b. The salmon price is known to be seasonal (Asche & Guttormsen, 2001; Mowi, 2020, p. 46), highly volatile and difficult to predict. The latter two features create substantial risk for the salmon market participants. Before 2006, forward contracting was perhaps the only straightforward risk management solution. Cross hedging with futures for related commodities does not appear to offer noticeable risk reduction, while hedging with shares of salmon farming companies traded on the Oslo Stock Exchange may be somewhat more effective (Bloznelis, 2016a, p. 29-30, 138-142). Since the opening of the Fish Pool exchange in 2006, hedging with futures contracts has become possible. The futures prices quoted by Fish Pool also offer price forecasts that may be expected to reduce the uncertainty about the future spot price of salmon.

Figure 1.1.a Yearly sales volume (million tonnes, light grey bars) of Norwegian farmed Atlantic salmon, Fish Pool trading volume (million tonnes, dark grey bars) and Fish Pool trading volume as fraction of salmon sales (black line), 2006-2022

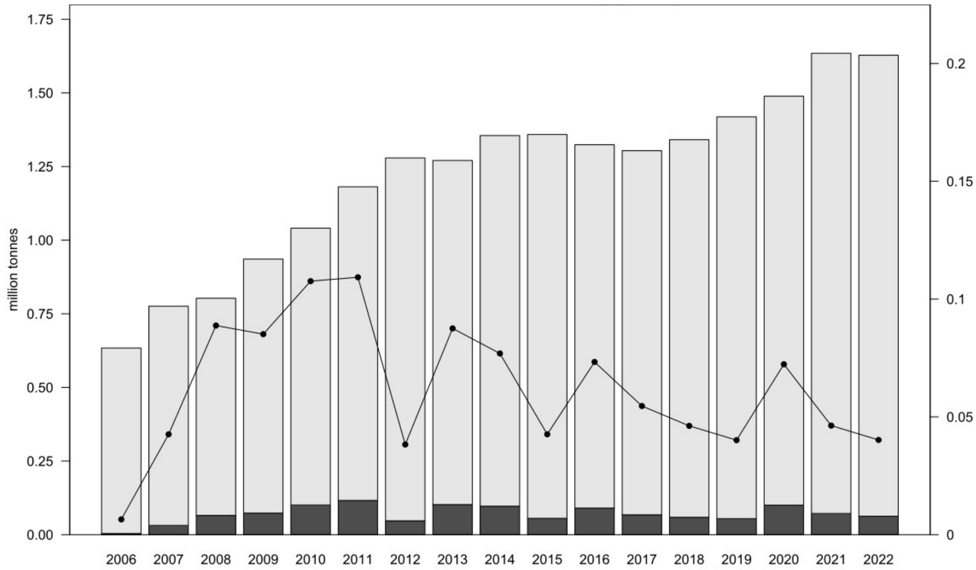
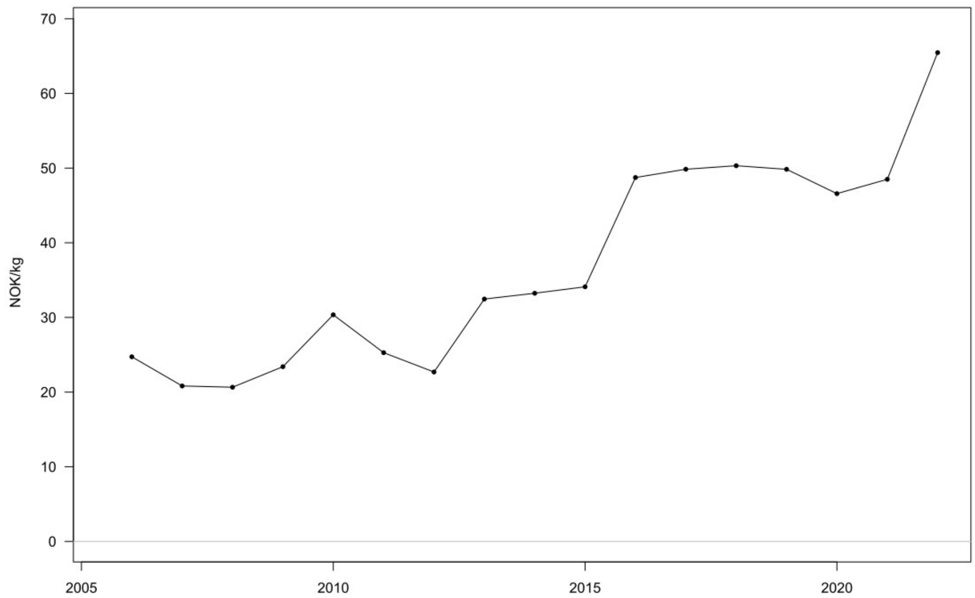


Figure 1.1.b Yearly salmon price (NOK/kg), 2006-2022



1.2 The futures market

There have been several attempts at launching a salmon futures exchange, namely, by European Fish Exchange, FishEx and Fish Pool (Bergfjord, 2007). Only Fish Pool has survived to this day. It has been offering cleared, cash-settled salmon futures contracts since mid-2006. Available for trading are 12 monthly contracts per year up to 60 (initially 30) months ahead. The underlying of a contract is the value of 1 metric tonne of Norwegian farmed Atlantic salmon distributed equally over a period of 4 or 5 weeks (28 or 35 days). The period roughly corresponds to a calendar month. The spot price used for calculating the value of the underlying is proxied by the Fish Pool Index (FPI). Its composition has changed multiple times since 2006, but the main component has remained the NASDAQ Salmon

Index (previously NOS price index) for 3-6 kg size fish. The index is obtained by a weekly survey of salmon exporters representing around one third of the Norwegian production volume (Bloznelis, 2016c). Not all export prices are determined on the spot, and neither the NASDAQ Salmon Index nor the FPI are genuine spot prices; see Bloznelis (2016a) p. 81-83 for a detailed discussion. However, it is common practice to refer to them as such, and we will do that, too. We are not aware of any other publicly available price data that would reflect the spot price of Norwegian farmed Atlantic salmon more accurately.

The trade volume of salmon futures fluctuated between 0.032 and 0.116 million metric tons from 2007 to 2022 (Fish Pool, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023), corresponding to between 4% and 11% of the physical market volume; see Figure 1.1.a. (This excludes the year 2006 in which Fish Pool was open for just over a half of the year.) Fish Pool does not publicly reveal value data for the futures contracts. Compared to some major commodity markets where the futures turnover exceeds the physical turnover by orders of magnitude, the trade volume on Fish Pool can be considered low relative to the size of the physical market. Moreover, the trades are infrequent. According to transaction-by-transaction data for 2006-2012, the year with the highest trading volume was 2011. In 2011, none of the contracts attracted more than one trade every two days; all maturities beyond 12 months received less than 1 trade a week (Bloznelis, 2016b). In all the other years between 2006-2012, trade volume was lower, often substantially so, and the number of trades was likely lower as well. In the years since 2012, the trade volume of the salmon futures contracts has not matched the peak of 2011, indicating the problem with low liquidity has persisted. This makes the risk transfer function of this market questionable at best.[2]

Two peculiarities make Fish Pool stand out among commodity futures exchanges. First, due to the low liquidity, defining a futures price for every contract every day is a challenge. The daily futures prices quoted by Fish Pool and used for determining the required margins in the traders' margin accounts are not set according to a publicly known deterministic formula based on transactions and/or bid/ask quotes for the futures contract of interest. Instead, they are set by a Fish Pool expert based on their best judgement. Thus, the day-to-day futures prices cannot be deterministically manipulated by futures trades and bid/ask offers, though we cannot rule out the possibility that the Fish Pool expert's judgement could be affected by them.

Second, trading in a futures contract is permitted – and the futures price is quoted – for as long as a few days after the end of the underlying target month (or the underlying month for short). For example, the price for the January 2024 contract with its underlying spanning from 2024-01-01 to 2024-02-05 was quoted until and including 2024-02-16 – eleven days after the end of the underlying

[2] The risk transfer function of the salmon futures markets has been investigated in a number of papers such as Ewald (2013), Misund & Asche (2016), Bloznelis (2018a) and Schütz & Westgaard (2018). Unfortunately, the low liquidity has been overlooked in most of these studies, even though the feasibility of the standard hedging strategies hinges on the market being sufficiently liquid for quick entry into and exit from the futures positions without affecting the futures price considerably. An exception is Bloznelis (2018a) who raises the issue explicitly and addresses it by proposing hedging strategies that are tailored to an illiquid market.

month – even though the settlement price was officially known starting from 2024-02-08, and the quoted price did not move after 2024-02-06. We do not see much reason to trade a contract once its settlement price is already known, and we are not aware of any such trades taking place despite the price quotes being published. Both peculiarities may be relevant when interpreting the quoted futures prices as forecasts of the salmon spot price.

1.3 Salmon price forecasting

Price forecasts are the fundament of business and financial planning across the salmon industry, including farming, processing, export and sales. Commercial forecasts are available for purchase from companies like Kontali Analyse and Capia as well as Norwegian brokerage houses' financial analysis divisions. With the advantage of insider knowledge and, in some instances, possibly market power, the largest industry players often maintain their own forecasts. Among them, publicly available are forecasts by salmon farming companies listed on the Oslo Stock Exchange; the forecasts are published in the companies' quarterly financial reports.

Meanwhile, academic literature of salmon price forecasting is scarce. Aside from four articles in the late 1980s and 1990s (Lin et al., 1989; Vukina & Anderson, 1994; Gu & Anderson, 1995; Guttormsen, 1999), the only study from this millennium we are aware of is Bloznelis (2018b). The first three studies used monthly data and focused on markets other than Norway: the European Community, the U.S.A. and Japan. Their findings are briefly summarized in Bloznelis (2018b) but are of limited relevance here, as neither the market nor the time period matches. Guttormsen (1999) produced weekly forecasts for the Norwegian market, but again the time period is well before the start of our sample. Only Bloznelis (2018b) considered the Norwegian market and a time period that overlaps with the one used in this study, making the results at least somewhat comparable. However, he only examined short-term forecasting of 1-5 weeks ahead. The target was the NASDAQ Salmon Index (previously NOS price index) that is not identical but very similar to the Fish Pool Index used in this study. Under a variety of accuracy metrics, Bloznelis (2018b) found salmon futures prices to yield the most accurate forecasts 4-5 weeks ahead, beating 15 alternative forecasting methods, though by a thin margin. Unfortunately, the forecast evaluation period spans only under two years (from 2013 week 1 to 2014 week 39). On the other hand, we will examine 18 years' worth of data and a broad range of forecast horizons stretching from three years ahead to the end of the underlying month or even a few days beyond.

2. Forecast optimality

2.1 Definitions of variables

Denote the futures price on day t for an underlying target month T by F_t^T, and denote the average Fish Pool Index for that underlying month by S_T. Denote the first day of the target month as T_{start} and define the forecast horizon h as the number of days between the current day and the start of the underlying month, $h \coloneqq T_{start} - t$. The horizon can be positive, zero or negative. A positive horizon means the forecast is made before the start of the underlying month. A horizon of zero means it is made on the first day of the underlying month, specifically after working hours. A negative horizon means the forecast is made during some other day of the underlying month or even a few days after its end. (See subsection 1.2 regarding futures price quotes on days after the underlying month has passed.) Define the forecast error for horizon $h = T_{start} - t$ committed on day t as $e_{t+h|t} \coloneqq F_t^T - S_T$. Define its relative counterpart as $re_{t+h|t} \coloneqq \frac{F_t^T}{S_T} - 1$ and the logarithmic counterpart (log-error) as $\ell e_{t+h|t} \coloneqq \log\left(\frac{F_t^T}{S_T}\right)$. Here and further, $\log(\cdot)$ stands for the natural logarithm.

2.2 Absolute optimality

Diebold (2024, p. 334) outlines the core characteristic of errors arising from an optimal forecast; they must not be predictable using the information underlying the forecasts. This is known as the *unforecastability principle*. From it he also deduces four properties that errors from optimal forecasts possess: (1) unbiasedness; (2) 1-step-ahead errors being white noise; (3) h-step-ahead errors being at most moving average of order $h-1$ processes, or MA($h-1$); and (4) forecast error variances being monotonically nondecreasing in horizon h. A fifth property reflecting the unforecastability principle is that a linear regression of the realized values on their optimal forecasts should yield an intercept of zero and a slope of unity.

2.2.1 Unbiasedness

Unbiasedness can be assessed by building and evaluating a model for the (conditional) expected value of the forecast errors. A model without explanatory variables would fit the unconditional expectation:

$$\mathrm{E}(u_t) = \mu \tag{1}$$

where u_t is a placeholder for either e_t, re_t or ℓe_t. (We skip notational reference to the specific forecast horizon for simplicity.) Meanwhile, the conditional expectation could be modelled as a function of season (with respect to either the time t of making the forecast or the underlying target month T), length of the underlying target month (4 vs. 5 weeks) and calendar time.

Seasonality may manifest itself in that the futures price would systematically overpredict the spot price in some periods of the year and underpredict in other periods. Or it may be that the futures price would systematically overpredict some target months and underpredict other ones. The latter alternative appears to be the more plausible one. Since supply and demand of physical salmon are seasonal and so is the spot price (most obviously in its level, but possibly also in other distributional characteristics), net hedging demand may also be seasonal with respect to the target month, yielding different levels of futures' risk premia for different target months. We will henceforth only consider seasonality with respect to the target month.

Length of the underlying refers to two possibilities: a contract with an underlying of 4 vs. 5 weeks. 5-week contracts are more risky than 4-week ones, as one has to forecast an additional week of spot prices. In presence of nonzero net hedging demand, this would imply a higher risk premium.

Calendar time tracks the age of the futures exchange. Participant composition, their (risk) preferences and knowledge evolve over time, making the distributions of the forecast errors change accordingly. The simplest form of such evolution would be a linear trend in the level of the forecast errors. More advanced forms could also be considered, but given the limited length of the time series they may be hard to estimate with sufficient precision. The model for the conditional expectation would thus be

$$\mathrm{E}(u_t|I_{t-1}) = \beta_0 + \beta_1 t + \beta_2 \ell_t + \gamma_2 Feb_t + \cdots + \gamma_{12} Dec_t \tag{2}$$

where $\mathrm{E}(u_t|I_{t-1})$ is the conditional expectation of u_t conditional on an information set I_{t-1} that consists of the explanatory variables in the model, t is time in days since the start of trading at Fish Pool on 12 June 2006, ℓ_t is an indicator variable with a value of 0 for the 4-week underlying months and 1 for the 5-week underlying months, and Feb_t to Dec_t are indicator variables for the underlying calendar months. The model could also be expressed equivalently as

$$\mathrm{E}(u_t|I_{t-1}) = \beta_1 t + \beta_2 \ell_t + \gamma_1 Jan_t + \gamma_2 Feb_t + \cdots + \gamma_{12} Dec_t \tag{3}$$

with an indicator variable for January in (3) replacing the intercept of (2).

Models (1) and (3) could be estimated as regressions by ordinary least squares. (Model (1) would be a regressor-free, intercept-only model.) However, due to overlapping observations they will be estimated as regressions with autoregressive moving-average (ARMA) errors instead, with the autoregressive and moving-average lag orders selected using the auto.arima algorithm due to Hyndman & Khandakar (2008); see Appendix for details.

2.2.2 White noise

White noise, denoted $WN(0, \sigma^2)$, is a time series process characterized by zero expected value, constant variance and zero autocorrelation at all lags (Diebold, 2024, p. 150). Optimal 1-step-ahead errors are white noise, where in our context 1-step-ahead should be understood as any nonpositive horizon.[3] It is impossible to prove an observed time series is white noise, but we can search for departures from the process's properties. Violations of such properties would imply the series is not white noise.

We will estimate autocorrelations and their p-values (relative to the null hypothesis of the autocorrelation being equal to zero) for lags 1 to 12 and will inspect their distributions across the forecast horizons. A large fraction of sizeable autocorrelations and low p-values will indicate a violation of the white-noise property.

Lag 1 could have a positive autocorrelation if consecutive forecast errors were persistently above or below their average (itself ideally equal to the average of their targets) and a negative autocorrelation if their positions relative to the average were alternating between above and below with every new observation, though not necessarily and not only then. Lag 12 autocorrelation could be positive if consecutive forecast errors were persistently above or below their average for the specific target months of the year (e.g. positive in Septembers or negative in Aprils), though again not necessarily and not only then.

[3] See Appendix for a technical note on autocorrelation and steps vs. horizons of irregularly spaced time series.

2.2.3 MA($h-1$)

A moving average of order $h-1$ process, or MA($h-1$), is given by the equation

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_{h-1}\varepsilon_{t-(h-1)} \quad (4)$$

where $\varepsilon_t \sim WN(0,\sigma^2)$. The process is characterized by its autocorrelation function (ACF) cutting off at lag $h-1$. That is, all autocorrelations beyond lag $h-1$ are equal to zero. The fact that h-step-ahead forecast errors from an optimal point forecast must be at most MA($h-1$) processes derives from two elements. First, the forecast errors must be uncorrelated, unless the periods they cover overlap. Second, for an equally-spaced time series, h-step-ahead forecast errors overlap when the forecast origins are between 1 and $h-1$ periods apart.

The property of h-step-ahead errors being at most MA($h-1$) processes is cumbersome to assess in our case due to overlapping observations and their irregular spacing in time, as detailed in the Appendix. Therefore, we adopt an equivalent alternative property suitable for our data; for a given forecast horizon, any subset of approximately equally spaced nonoverlapping observations should be white noise. Once the relevant subset is obtained, the problem of evaluating to which extent the data are characterized by the property reduces to the one discussed in subsection 2.2.2.

2.2.4 Nondecreasing variances

Nondecreasing variances can be assessed somewhat informally by examining an ordered sequence of empirical variances of the forecast errors, $\hat{\sigma}^2_{H_{\min}}, \dots, \hat{\sigma}^2_{H_{\max}}$, corresponding to an increasing sequence of horizons $H_{\min}, \dots, H_{\max}$. A formal way would be to test the equality (inequality) of variances $\sigma^2_{H_{\min}} \leq \cdots \leq \sigma^2_{H_{\max}}$. We are not aware of any test based on assumptions compatible with the nature of our data, so we will refrain from formal testing.

2.2.5 Mincer-Zarnowitz test

Consider the Mincer-Zarnowitz (MZ) regression (Mincer & Zarnowitz, 1969) that models the realized values of the target variable as a linear function of the forecasts,

$$y_{t+h} = \beta_0 + \beta_1\hat{y}_{t+h|t} + u_{t+h|t}. \quad (5)$$

Here, y_{t+h} is the forecast target, $\hat{y}_{t+h|t}$ is the forecast and $u_{t+h|t}$ is an error term. The forecasts are summaries of the data available to the forecaster at the time of making the forecast. An optimal forecast should not be improved by shifting it by a constant or multiplying it by a scalar. Thus, the coefficients β_0 and β_1 must equal 0 and 1, respectively. Testing whether that is the case constitutes the MZ test.

A complication may arise if y_{t+h} contains a unit root, something that is plausible in the context of commodity prices. (At least it may be difficult to distinguish some commodity prices from unit root processes in small samples, and statistical techniques based on the assumption of stationarity may perform more poorly on such data than ones based on the assumption of a unit root.) Under the null hypothesis of forecast optimality, or $(\beta_0,\beta_1) = (0,1)$, $\hat{y}_{t+h|t}$ will also contain a unit root and be cointegrated with y_{t+h} with a cointegrating vector of $(1,-1)$. The test statistic for $(\beta_0,\beta_1) = (0,1)$ will not have its usual distribution derived under the assumption of stationarity, and its naïve use could yield misleading results. To the best of our knowledge, this issue has not been addressed in the literature, although it has been raised in an unpublished doctoral dissertation by Bloznelis (2016a). We note that given y_t, forecasting y_{t+h} is equivalent to forecasting $y_{t+h} - y_t =: \Delta_h y_{t+h}$. We propose to reformulate the forecasting problem using $\Delta_h y_{t+h}$ as the target. After subtracting y_t from both sides of equation (5), the MZ regression becomes

$$\Delta_h y_{t+h} = \beta_0 + \beta_1 \hat{y}_{t+h|t} - y_t + u_{t+h|t}. \tag{6}$$

Denote the forecast of the h-period change in the target variable of equation (6) by $\widehat{\Delta_h y}_{t+h|t} := \hat{y}_{t+h|t} - y_t$. Under the null hypothesis of $(\beta_0, \beta_1) = (0,1)$, equation (6) becomes

$$\Delta_h y_{t+h} = \widehat{\Delta_h y}_{t+h|t} + u_{t+h|t}, \tag{7}$$

which corresponds to a MZ regression with h-step increments $\Delta_h y_{t+h}$ as the forecast targets and predicted h-period changes $\widehat{\Delta_h y}_{t+h|t}$ as their forecasts. Conveniently, both $\Delta_h y_{t+h}$ and $\widehat{\Delta_h y}_{t+h|t}$ are free of unit roots. Thus, the test statistic retains its usual null distribution as in the stationary case. Anticipating the FPI to be approximately a unit-root process, we will use the reformulated version of the test. Due to overlapping observations, regression (7) will be treated as a regression with ARMA errors instead of estimating it directly by ordinary least squares; see Appendix for details.

2.3 Relative optimality

The four properties from Diebold (2017, p. 334) and the MZ test concern absolute optimality. One can also consider optimality relative to benchmark forecasts. Are there readily available or easily obtainable forecasts that are more accurate than the futures prices? The last observed value of the FPI is a natural candidate. It would be the optimal forecast under square loss[4] if the FPI were a random walk. It would also be an (approximately) optimal forecast under absolute loss if the logarithm of the FPI were a random walk and its increments were (approximately) symmetrically distributed around zero, as elaborated on in subsection 2.5.

Another pair of benchmarks could be obtained by correcting the futures price for the estimated expected value of the forecast error. If the forecast errors were a random walk, the best estimate of their expected value in the future would be the last observed forecast error. If the errors were independently and identically distributed (i.i.d.), the best estimate would be their historical mean.

Correspondingly, one may also consider seasonal benchmarks. The last observed value of the same month of the year could be preferred if the forecast errors were a seasonal random walk. The historical mean of the same month of the year (such as of all Januarys, all Februarys, etc.) could work well if the forecast errors were i.i.d. within the same month across the years (all Januarys, all Februarys, etc.) and uncorrelated with the remaining months (a January forecast error being uncorrelated with errors from Februarys, Marches, etc.). It could be obtained from the linear model in equation (3) by dropping the linear time trend and the long month indicator:

$$\mathrm{E}(u_t | I_{t-1}) = \gamma_1 Jan_t + \gamma_2 Feb_t + \cdots + \gamma_{12} Dec_t. \tag{3'}$$

To save space, we will skip the former seasonal benchmark and only employ the latter one. All of these benchmarks can be considered simple and readily available, as they are based on methods that

[4] In this subsection, we will assume square loss unless specifically indicated otherwise.

are the first ones to be mentioned in an introductory forecasting textbook such as Hyndman & Athanasopoulos (2018).

The futures prices as forecasts can be compared to these benchmarks by visualizing the forecast errors and the corresponding losses such as absolute errors and squared errors. For a formal test, one can use the Diebold-Mariano (DM) test (Diebold & Mariano, 1995) that assesses the null hypothesis of the expected losses from two different forecasts being equal. As was the case with regressions discussed in the previous subsections, also the DM test regression (one with only an intercept and no regressors) will be handled as a regression with ARMA errors.

2.4 Scale of forecast errors

Besides forecast optimality, analysis of the size of forecast errors could be enlightening. The size reflects the difficulty of prediction. The factors that could plausibly affect the error size are the same as these considered under unbiasedness, i.e. seasonality, length of the underlying target month and calendar time. Concretely, we consider the counterpart of equation (3):

$$\mathrm{E}(|u_t||I_{t-1}) = \beta_1 t + \beta_2 \ell_t + \gamma_1 Jan_t + \gamma_2 Feb_t + \cdots + \gamma_{12} Dec_t \tag{8}$$

with variables defined correspondingly and $|\cdot|$ being the absolute value operator. Seeing how the different regressors explain the expected size of the forecast error could enable the forecast users to anticipate the magnitude of the error more precisely based on the month of the year, the length of the underlying month and calendar time. Due to overlapping observations, regression (8) will be treated as a regression with ARMA errors.

The model in equation (8) is not entirely sensible, as we are using regular multiple regression (aside from ARMA errors) with a nonnegative dependent variable. An alternative would be, for example, to take a logarithm of the dependent variable (thus, the logarithm of the error size) and regress it on the same explanatory variables (with ARMA errors). However, the interpretation of the dependent variable would become convoluted, especially in the context of log-errors that we advocate for in section 2.5. For ease of interpretation, we choose to maintain the original specification instead.

2.5 Levels or logarithms?

Most of the statistical measures, models and tests employed in this study assume the observations or the errors are homoscedastic. The level of the salmon futures and spot prices was increasing substantially over the 2006-2024 period. The price fluctuations were increasing with the level; see Figure 3.1.a. Using raw prices, or price levels, in the analysis would most likely result in undesired artefacts caused by the heteroskedasticity. A common modeling approach in such cases is to work with the logarithms of data instead. Figure 3.1.b in subsection 3.1 shows that logarithmic prices are considerably more homoscedastic.

The logarithm being a nonlinear transformation could cause us some trouble due to Jensen's inequality. Let (Ω, F, P) be a probability space, X an integrable real-valued random variable, φ a convex function, E and expectation operator and E_t a conditional expectation operator conditioning on the information available up to and including the time t. Per Jensen's inequality, $\varphi(\mathrm{E}[X]) \leq \mathrm{E}\,[\varphi(X)]$. Applying this to our case, if the futures price were an unbiased forecast of the future spot price, the same need not hold for their logarithms; $F_t^T = \mathrm{E}_t[S_T] \nRightarrow \log(F_t^T) = \mathrm{E}_t[\log(S_T)]$. And if the logarithm of the futures price were an unbiased forecast of the logarithm of the future spot price, the same need not hold for their levels (the raw prices); $\log(F_t^T) = \mathrm{E}_t[\log(S_T)] \nRightarrow F_t^T = \mathrm{E}_t[S_T]$. Also, results from conditional expectation models for logarithms of prices will not translate directly to equivalent results for the corresponding conditional expectation of the price levels.

We believe the unfortunate implications of Jensen's inequality from using logarithms are less harmful than the ones due to heteroskedasticity inherent in the use of data in levels. Analysis of the latter would require heteroskedasticity adjustments that would not only be cumbersome to implement but also contribute measurement noise. On the other hand, logarithmic transformation offers a couple of benefits. First, the interpretation of errors is quite appealing: for small errors, the log-error $\ell e_{t+h|t} = \log\left(\frac{F_t^T}{S_T}\right)$ approximately equals the relative error $re_{t+h|t} = \frac{F_t^T}{S_T} - 1$. E.g. a log-error of 0.05 approximately equals the relative error of 0.05, or 5%. As the monthly FPI values have fluctuated between around NOK 20/kg and NOK 130/kg in our sample period, it could be convenient to consider over- or underprediction relative to the price level instead of in absolute terms. When the price level is close to NOK 20/kg, a forecast error of NOK 2 would be a 10% under- or overprediction and thus not negligible. When the price level is well into the triple digits, it would be less than 2% and thus quite a bit less concerning.

Second, logarithmic transformation preserves quantiles. E.g. the logarithm of the median of a level error distribution remains the median of the log-error distribution, and vice versa; $\log\left(\mathrm{median}\left[e_{t+h|t}\right]\right) = \mathrm{median}\left[\ell e_{t+h|t}\right]$. This is relevant if the log-error distribution is approximately symmetric (which it appears to be in our case), so that its median approximately equals its expectation. Then forecasts from models of the (conditional) expectation of log-errors will also forecast the (conditional) median of these errors. Due to quantile preservation, their exponents will forecast the (conditional) median of errors on the level scale. Why is that relevant? The median on the level scale is the optimal point forecast under absolute loss, i.e. when the forecaster's losses due to missing the target grow linearly in the size (the absolute value) of the error. Meanwhile, the expectation is optimal under square loss. To the extent that actual losses of the forecast users are closer to being linear than quadratic in the forecast errors, the median is the more relevant forecast target. This makes modelling and forecasting on the logarithmic scale attractive. All in all, we find that analysis of log-errors is more promising than that of errors in levels, and we will proceed with the former.

2.6 Overlapping observations and the effective sample size

A common obstacle to precise evaluation of long-term forecasts is short sample span, an issue that also pertains to our case.[5] Between July 2006 and June 2024 there are only 18 years, or 216 months. Considering long horizons, this corresponds to six three-year periods or nine two-year periods.

[5] Meanwhile, samples spanning long periods of time tend to be plagued by another problem. The data generating process tends to evolve over time, and older data in a long sample may poorly reflect the state of the process that generated the newer data.

Allowing for overlapping observations, the numbers increase from six (nine) to 181 (193). However, the information overlap is substantial, and the effective sample size when estimating, say, the expectation of a log-error is much closer to six (nine) than to 181 (193). With a sample of six or nine observations, one may expect only the largest and most pronounced effects to be reliably distinguishable from their values under common null hypotheses. For example, we do not have solid ex ante reasons for $\mathrm{E}_t\left[\ell e_{t+h|t}\right]$ to be particularly distant from zero, so we would not be surprised by null findings for the longer horizons.

3. Data and empirical results

This section presents the data and reports the empirical assessment of forecast quality.

3.1 Data

Our main source of data is Fish Pool. From there, we have obtained daily futures prices for monthly contracts up to five years ahead, for the period of early June 2006 to early July 2024. Also from there, we have obtained weekly values of the Fish Pool Index (FPI) to be used as a proxy for the spot price. They have been aggregated to a monthly frequency according to the definition of the underlying target months at Fish Pool and are depicted in Figures 3.1.a (in levels) and 3.1.b (in logarithms). The forecast errors derived from the futures prices are shown in Figure 3.1.c (with respect to calendar time) and Figure 3.1.e (with respect to time to the first day of the underlying target month), and their logarithmic counterparts are shown in Figures 3.1.d and 3.1.f, respectively. Yearly statistics of Fish Pool trading volume presented in the introduction are from Fish Pool annual reports. Yearly salmon production and prices are from Statistics Norway (data until 2019) and Directorate of Fisheries (data since 2020).

Only minimal cleaning of the data was necessary. Fish Pool futures price data contained prices for two Saturdays: 1 October 2011 and 31 January 2015. As there is no trading on weekends, these observations appear to be erroneous and have been removed before proceeding with the analysis.

The variables used in the study have been defined and their notation introduced in subsection 2.1.

Figure 3.1.a Monthly Fish Pool Index (FPI) (NOK/kg), 2006-2024

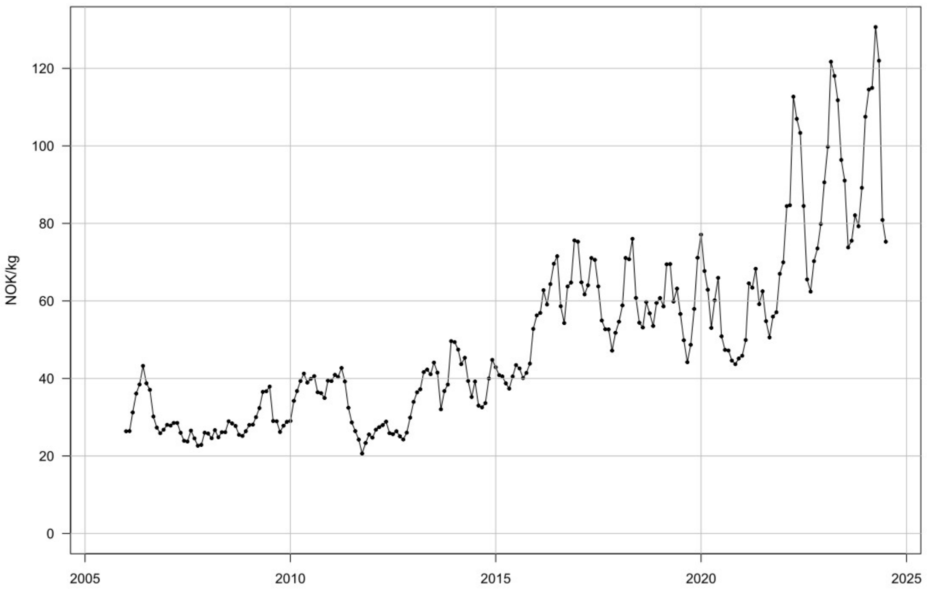


Figure 3.1.b Logarithm of monthly Fish Pool Index (FPI), 2006-2024

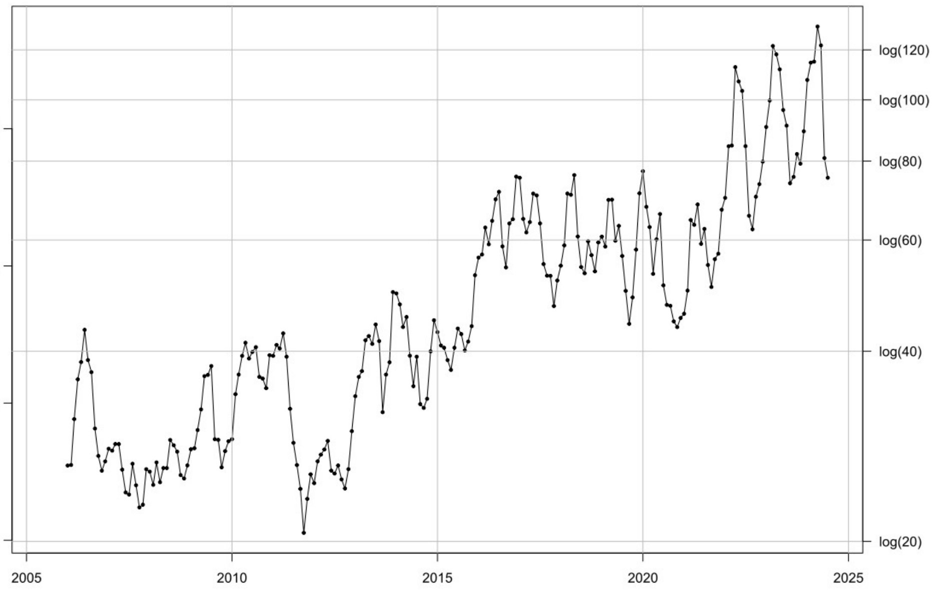


Figure 3.1.c Forecast errors (NOK/kg) from futures prices over forecast horizons of up to five years in calendar time, 2006-2024

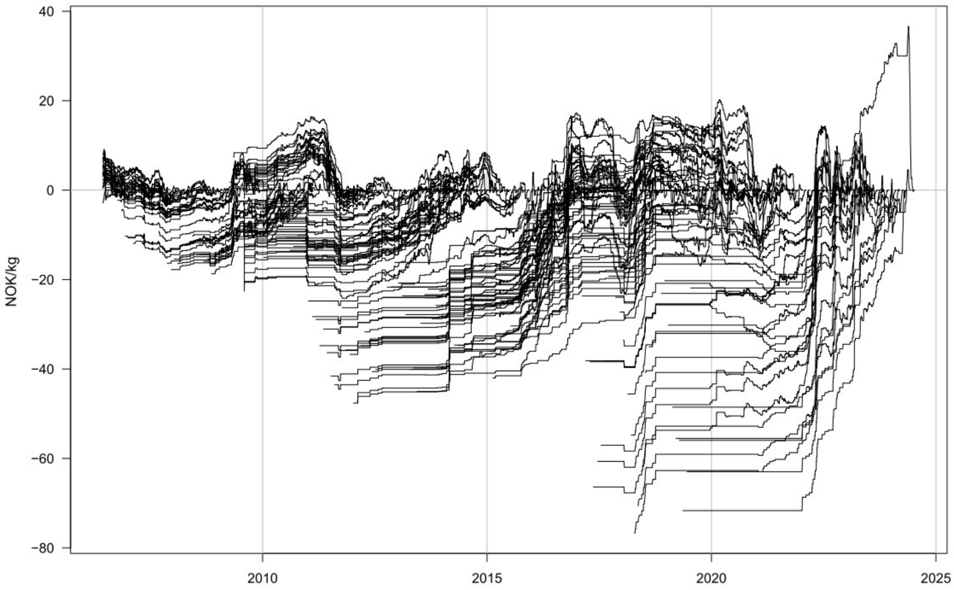


Figure 3.1.d Forecast log-errors from futures prices over forecast horizons of up to five years in calendar time, 2006-2024

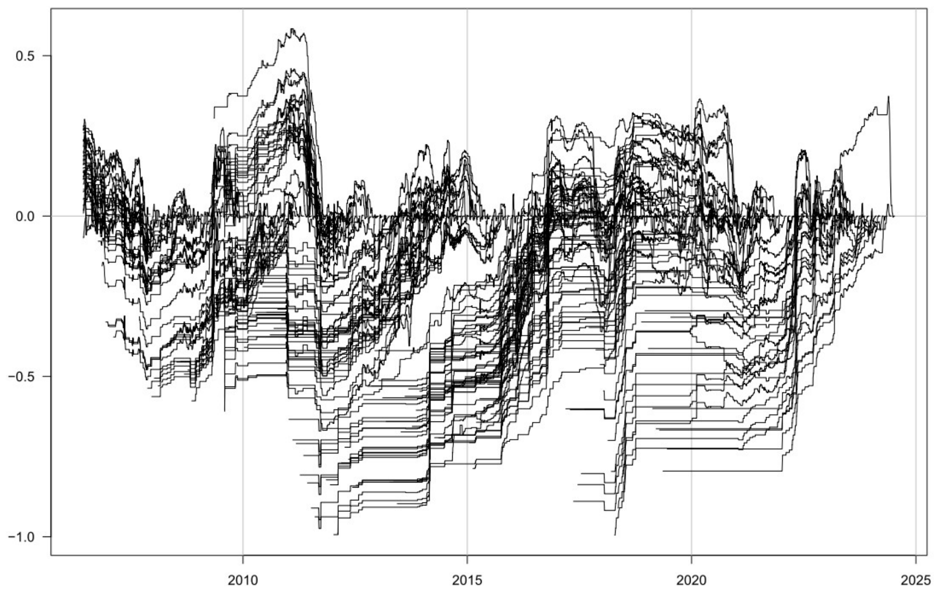


Figure 3.1.e Forecast errors (NOK/kg) from futures prices over forecast horizons of up to five years in event time

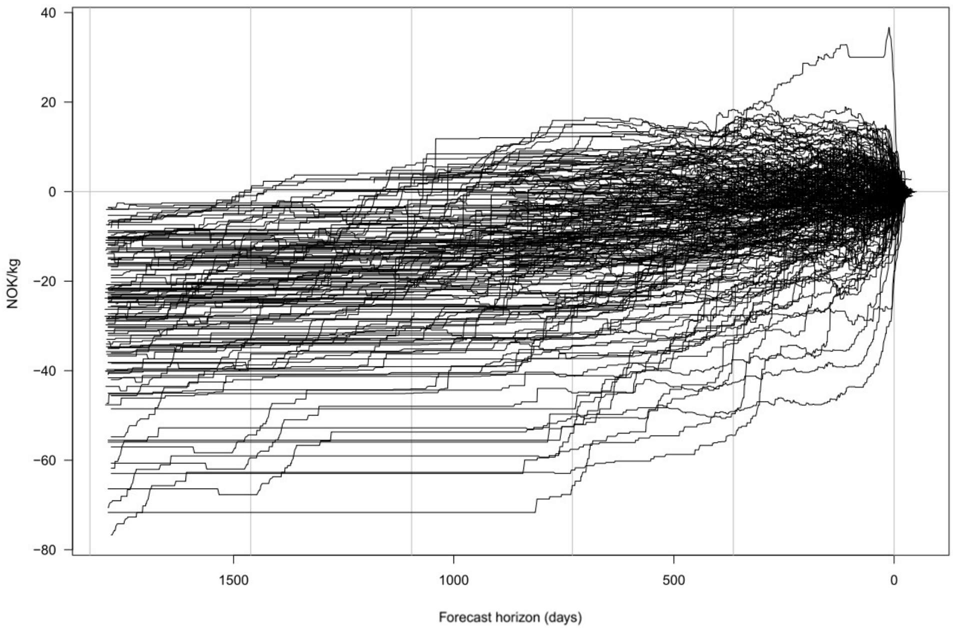
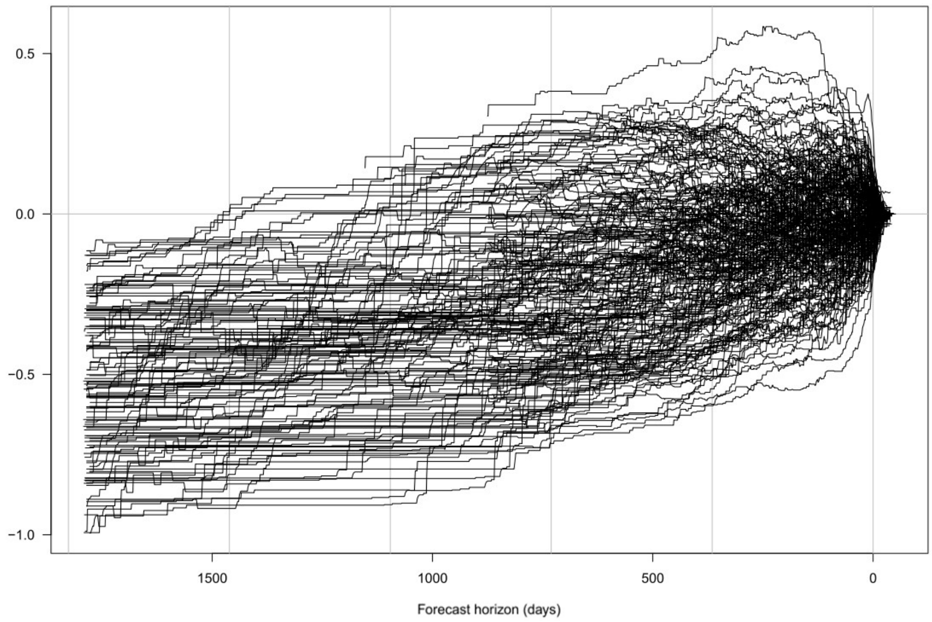


Figure 3.1.f Forecast log-errors from futures prices over forecast horizons of up to five years in event time

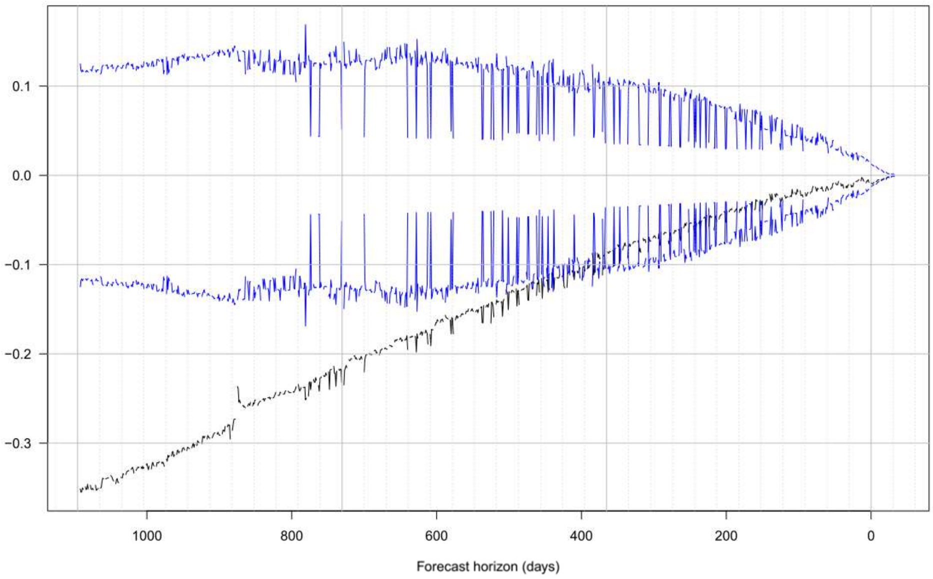


3.2 Absolute optimality

3.2.1 Unbiasedness

Figure 3.2.1.a shows the estimated bias of logarithmic forecasts (the logarithmic bias) across forecast horizons between three years and negative 35 days. The bias is the expected value of the log-error, and our maximum likelihood estimate is approximately equal to the sample mean of log-errors. The 95% critical values stem from ARMA models selected separately for each horizon by the auto.arima algorithm due to Hyndman & Khandakar (2008), as explained in the Appendix.

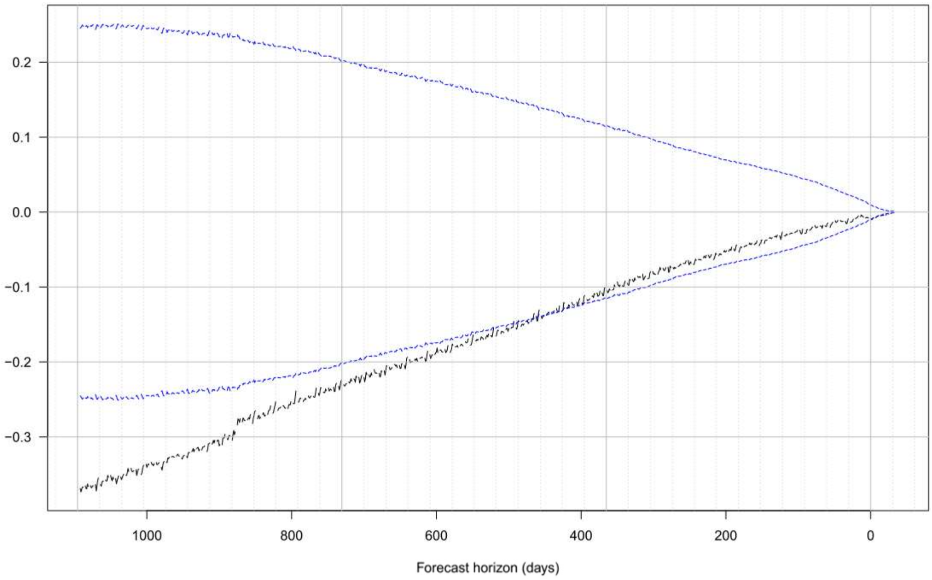
Figure 3.2.1.a Means of forecast log-errors (black) across forecast horizons and 95% critical values (blue) due to regression with ARMA errors



The logarithmic bias is negative and strikingly linear with respect to the horizon, adding about -0.11 on the logarithmic scale (or -12%) a year. For the nonpositive horizons, it tracks the lower critical values very closely, but without exceeding them. For most of the positive horizons between just a few days and about 14-15 months, it is also not statistically distinguishable from zero. At 14-15 months, as it reaches a size of about -0.12 (or -13%), it becomes statistically significant. Thus, we have an indication of forecast suboptimality for the longer horizons.

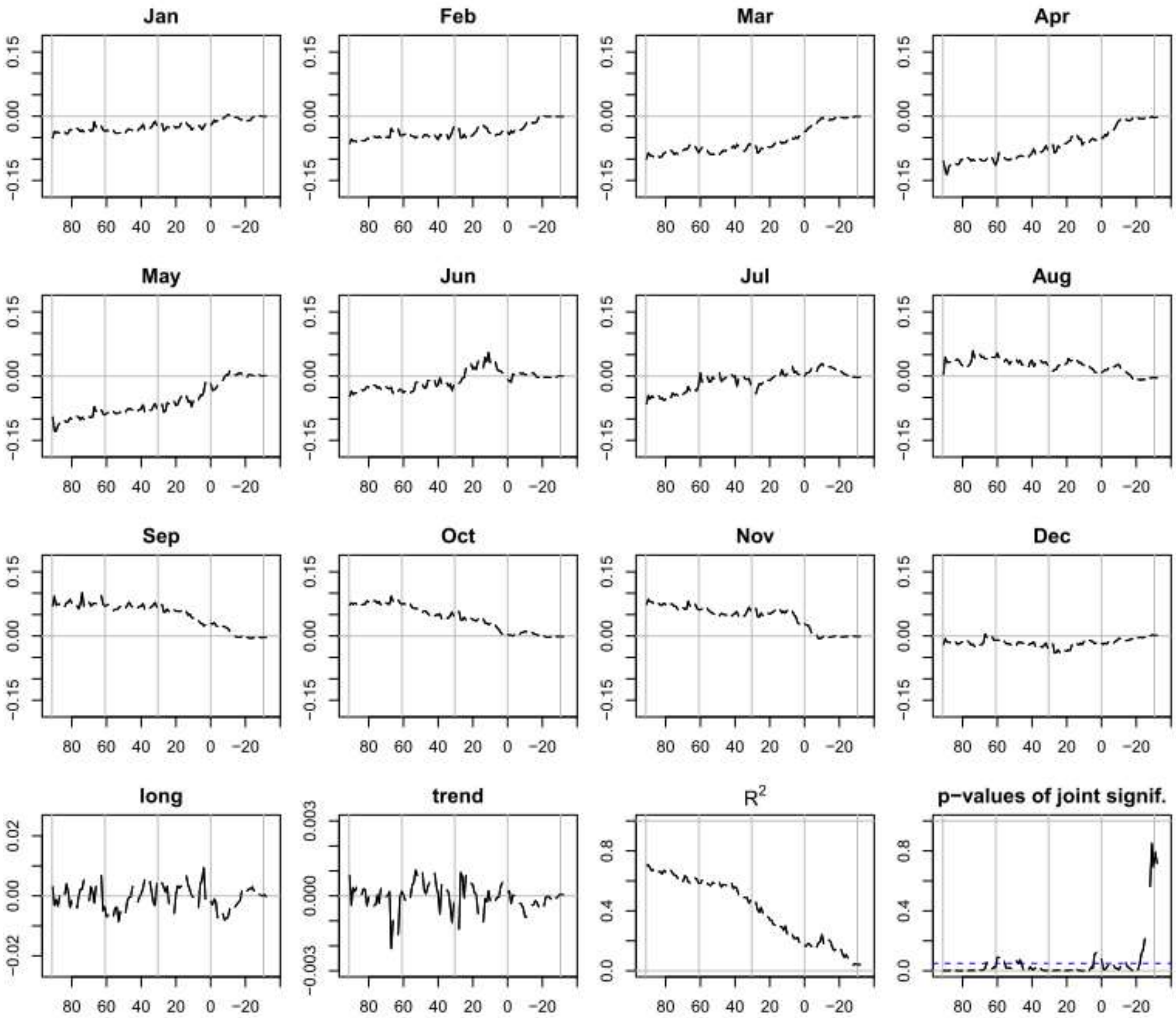
The interval between the critical values conspicuously stops widening after about 21 months. We may suspect that regression with ARMA errors is having trouble approximating the autocorrelation of the log-errors for the longest horizons. As the effective sample size is close to single digits there, we should remain cautious in interpreting the results. For comparison, we produce critical values assuming that (1) there are only as many observations as can fit into the sample's time span without overlapping, and (2) these are i.i.d. As Figure 3.2.1.b shows, such critical values are more conservative (further away from zero). However, the conclusion regarding statistical significance of the bias is unchanged; the expected values of log-errors are statistically borderline nonzero for nonpositive horizons and statistically nonzero for horizons beyond about 14-15 months.

Figure 3.2.1.b Means of forecast log-errors (black) across forecast horizons and 95% critical values (blue) assuming i.i.d. observations and adjusting for effective sample size



Conditional unbiasedness is assessed by an overall F-test for the regression (3) with seasonal (monthly) indicator variables, long-underlying-month indicator and a linear time trend. Given the small effective sample size for the longer horizons, the model with 14 slope parameters is only fit for horizons of under three months. In this range, at least about 5 effective observations per parameter are available. The estimated slope coefficients, p-values of F-statistics for their joint significance (reflecting on the forecasts' conditional biasedness) and the models' coefficients of determination (R^2) are shown in Figure 3.2.1.c.

Figure 3.2.1.c Coefficient estimates, coefficients of determination and p-values of the coefficients' joint significance from regressions with ARMA errors due to equation (3) across forecast horizons



The null hypothesis of conditional unbiasedness cannot be rejected for some of the most negative horizons and a small fraction of positive horizons. However, for most horizons, the forecasts are conditionally biased. R^2 values are close to zero for the shortest horizons and approximately monotonically increasing to about 0.7 for the longest three-month horizon. The length of the underlying month and the linear time trend both have estimated effects that vary erratically with respect to the forecast horizon. The estimated slope coefficients on the individual months of the year suggest that for positive horizons, the futures price is biased downward in March to May and upward in September to November. That is, in the spring, the futures prices underpredict the future spot price, while in the fall they overpredict. For the longer horizons, the estimated effect sizes for the spring and fall months are around 0.05 to 0.13 (with differences between months reaching up to around 0.20) and thus considerably larger than the estimated unconditional bias. The winter and summer months show smaller estimated effects or ones with varying signs. While statements about the estimated slope coefficients on the different months are conditional on the values of the length of the underlying month and the linear time trend, unreported models without the latter two variables support essentially the same unconditional statements.

3.2.2 White noise

Figure 3.2.2.a shows autocorrelations for lags between 1 and 12 across the forecast horizons, and Figure 3.2.2.b shows their p-values corresponding to a null hypothesis of the autocorrelation being zero. Results are not available for longer horizons where the approximate effective sample size is less than $\ell + 5$ observations with ℓ being the lag length. Colors code horizons counted in calendar months. For nonpositive horizons ($h \leq 0$) where optimal forecasts would be white noise, autocorrelations are small and hardly economically significant. However, for lags 1, 2 and 8, considerably more than 5% of the p-values are 0.05 or less; see histograms in Figure 3.2.2.c. Thus, formally the errors do not appear to be white noise, even if the deviations are economically negligible.

Figure 3.2.2.a Estimated autocorrelations of forecast log-errors for lags between 1 and 12 across forecast horizons (in days)

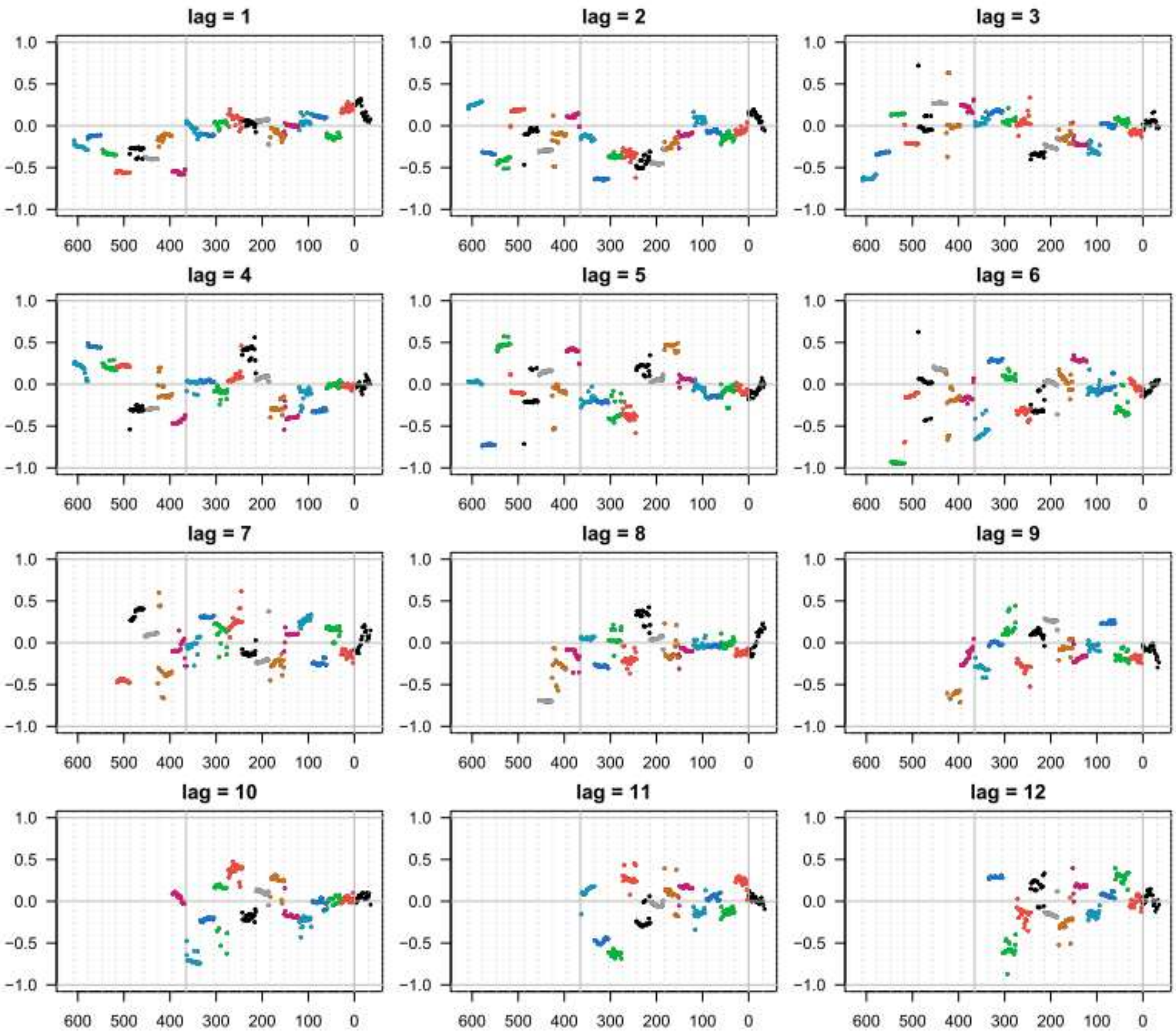


Figure 3.2.2.b P-values of autocorrelations of forecast log-errors for lags between 1 and 12 across forecast horizons (in days) corresponding to a null hypothesis of the true value being zero

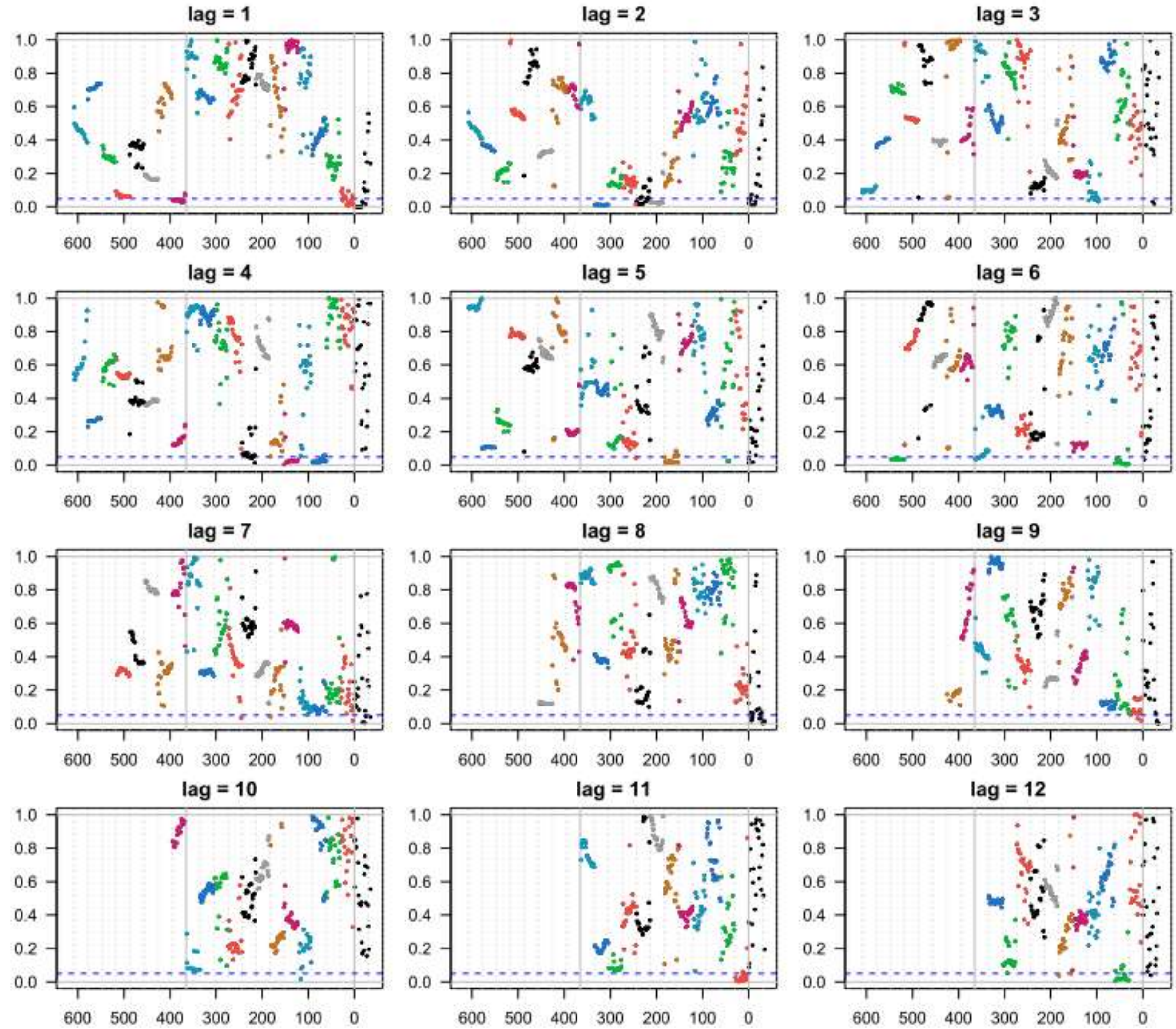
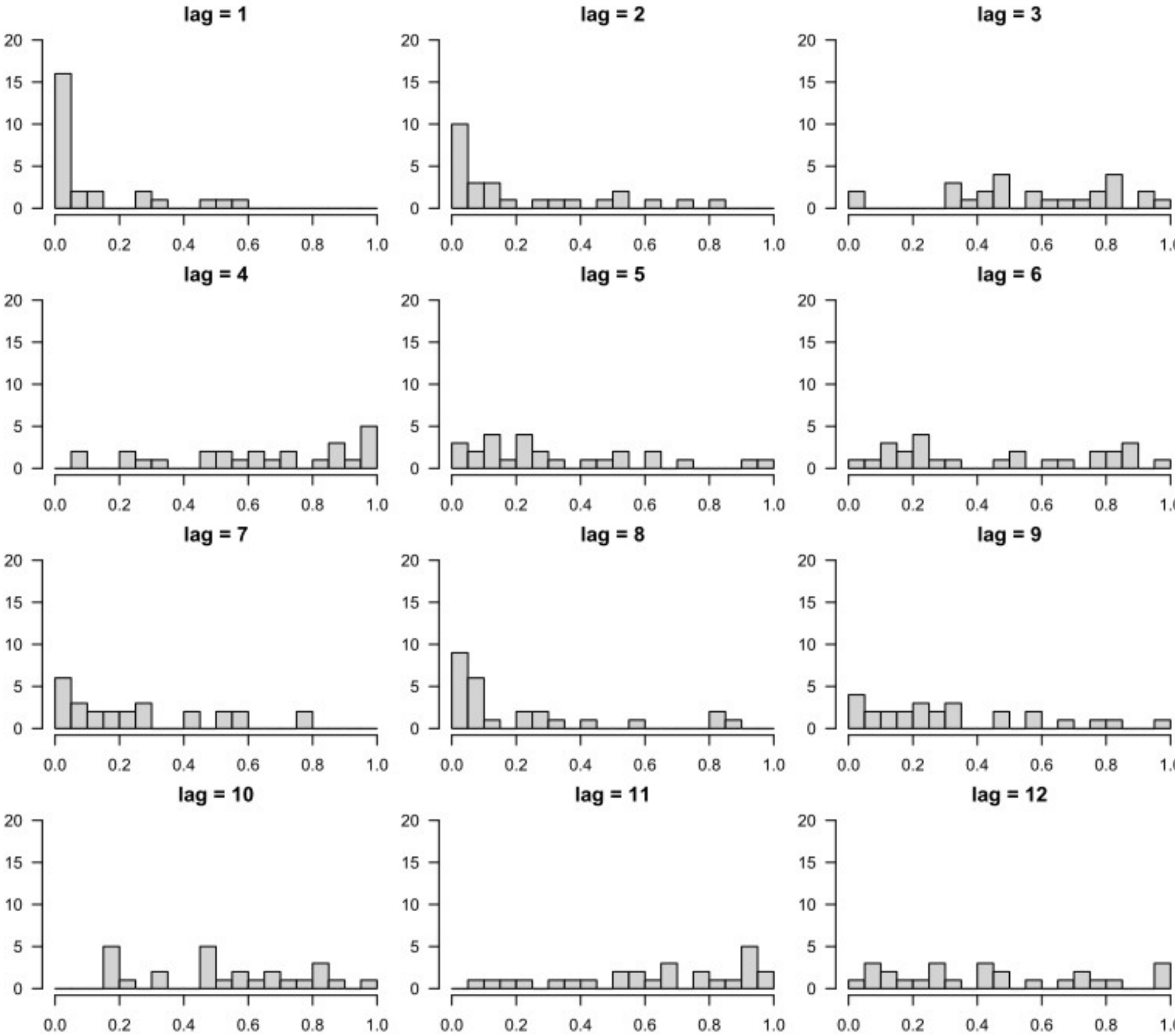


Figure 3.2.2.c Histograms of estimated autocorrelations of forecast log-errors for lags between 1 and 12 aggregated over nonpositive forecast horizons

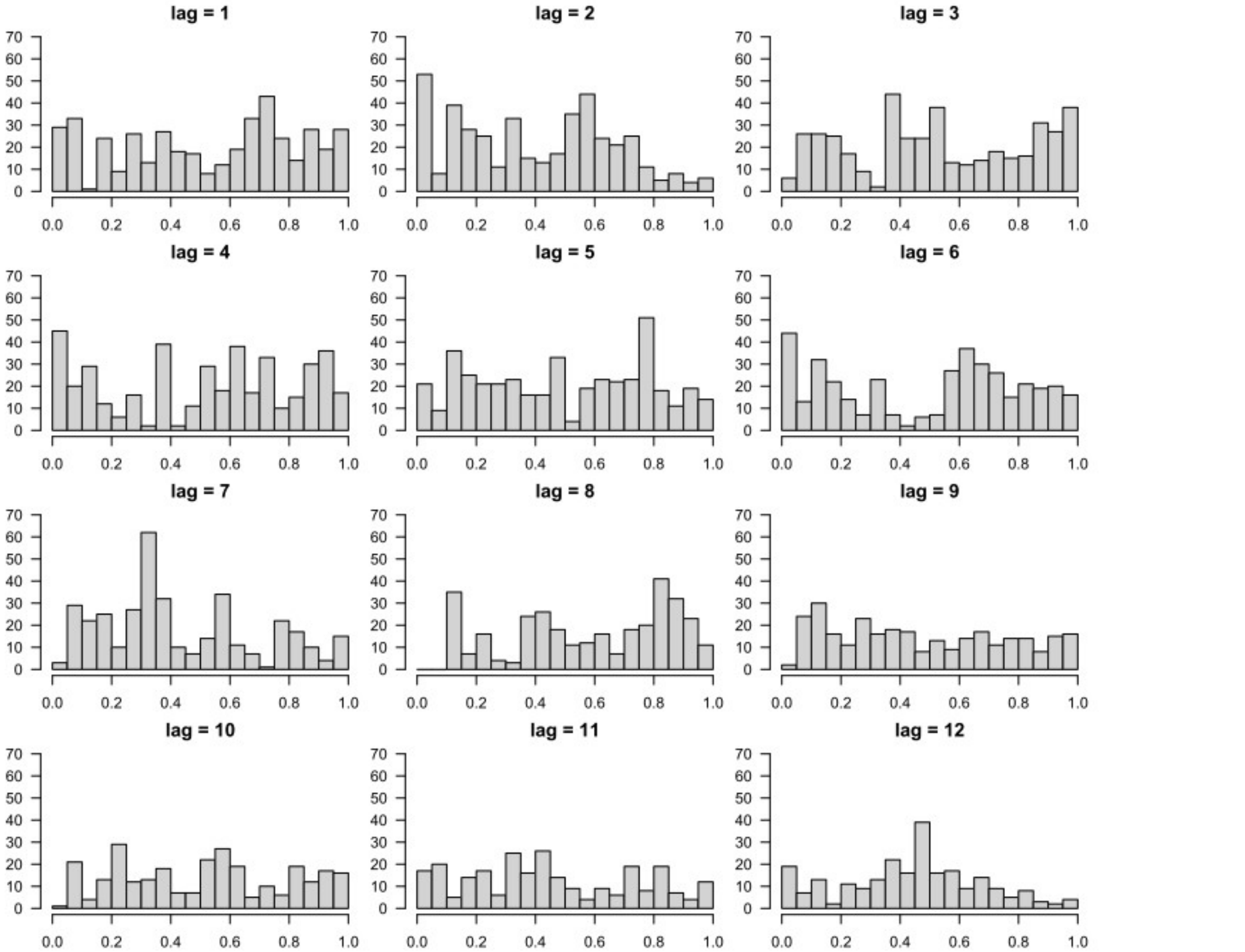


3.2.3 MA($h-1$)

For any positive horizon $h > 0$, instead of analyzing overlapping forecasts and assessing whether they are at most MA($h-1$) processes, we inspect whether a subset of approximately equally spaced nonoverlapping observations is white noise, as proposed in subsection 2.2.3. Thus, we reinspect Figures 3.2.2.a and 3.2.2.b, this time for $h > 0$. We also view the histograms in Figure 3.2.3.a. Most of the autocorrelations are small (between -0.5 and 0.5), and the fraction of p-values at or below 0.05 is generally not substantially greater than 0.05. The largest exception is lag 2 (and smaller ones are lag 4 and lag 6), but these could just as well be due to random chance. All in all, futures prices viewed as forecasts of the spot prices violate the MA($h-1$) property for some horizons, but most of the violations are either economically small or consistent with chance variation, or both.[6]

[6] Lags are measured in terms of time intervals between approximately equally spaced nonoverlapping observations. E.g. for $h = 0$ days, lag 1 (2, 3, …) measures autocorrelation between observations that are one (two, three, …) month(s) apart, while for $h = 10$ days the observations are

Figure 3.2.3.a Histograms of estimated autocorrelations of forecast log-errors for lags between 1 and 12 aggregated over positive forecast horizons



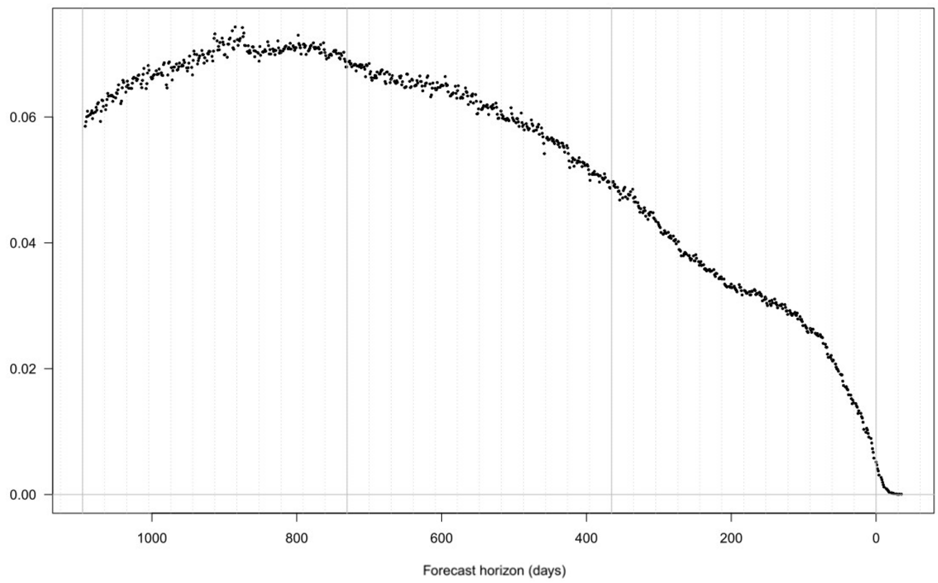
3.2.4 Nondecreasing variances

Figure 3.2.4.a shows that the forecast error variances are monotonically increasing from the most negative horizon to just over two years ahead, but then level off for about six months and turn around and start decreasing. The latter pattern is a violation of forecast optimality. We do not have a hypothesis for why this happens. We should also keep in mind that the effective sample size for the longest forecast horizons is in the single digits, so the variance of the data generating process may not be estimated precisely.

---

two (four, six, …) months apart, and for $h = 40$ days the observations are three (six, nine, …) months apart.

Figure 3.2.4.a Estimated variances of forecast log-errors across forecast horizons



3.2.5 Mincer-Zarnowitz test

P-values of the Mincer-Zarnowitz test are shown in Figure 3.2.5.a. Low p-values indicating violations of forecast optimality characterize the most negative forecast horizons, horizons between about 10-11 months and ones above about 15 months. For the latter, this is in line with the results from subsection 3.2.1 pertaining to forecast bias. Since one of the two components that the MZ test accounts for is precisely the bias, this is as expected; see also Figures 3.2.5.b and 3.2.5.d. For the negative horizons where the bias is statistically borderline nonzero but the p-values of the MZ test are considerably below 0.05, the second component of the MZ test – the slope coefficient on the forecasted change in the FPI – is contributing to that by being too low relative to the null hypothesis (i.e. by being under unity); see Figures 3.2.5.c and 3.2.5.e.

Figure 3.2.5.a P-values of Mincer-Zarnowitz test of forecast log-errors across forecast horizons

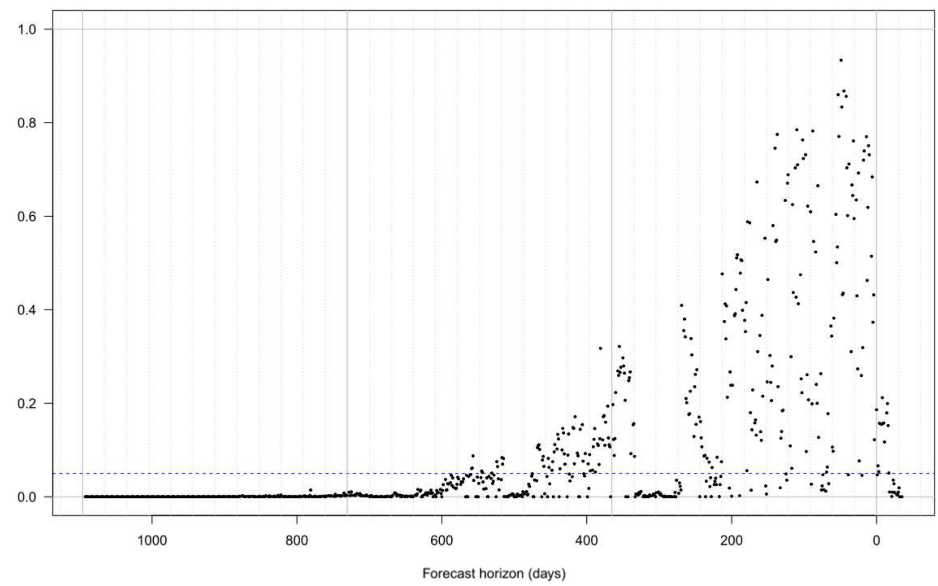


Figure 3.2.5.b P-values of intercepts of Mincer-Zarnowitz regressions of forecast log-errors across forecast horizons

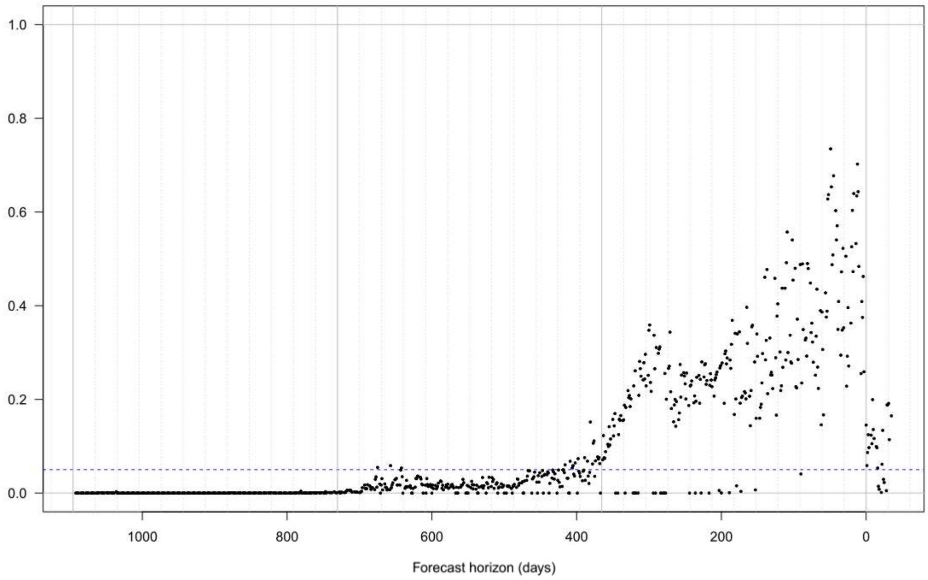


Figure 3.2.5.c P-values of slopes of Mincer-Zarnowitz regressions of forecast log-errors across forecast horizons (red indicates slopes less than unity, black indicates otherwise)

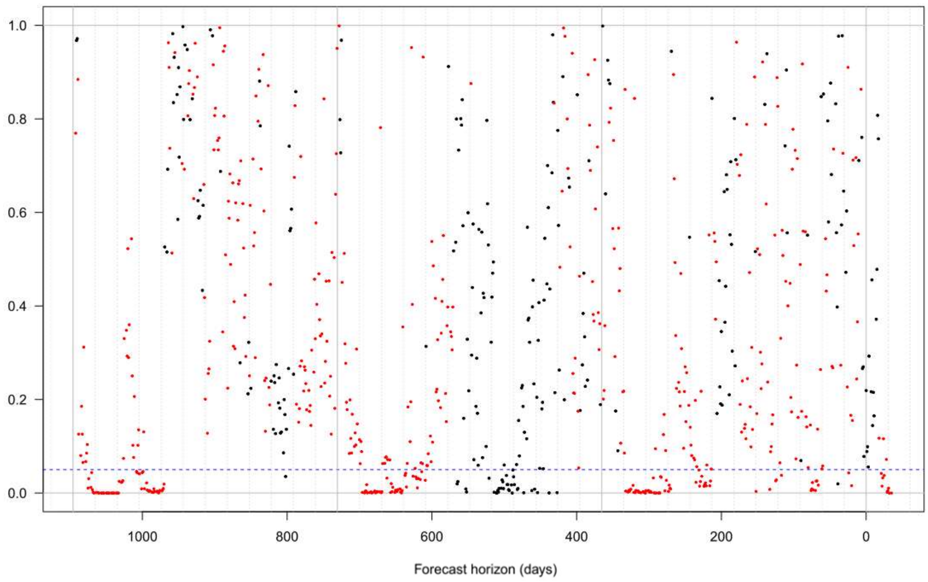


Figure 3.2.5.d Negative of estimated intercepts with corresponding critical values of Mincer-Zarnowitz regressions of forecast log-errors across forecast horizons

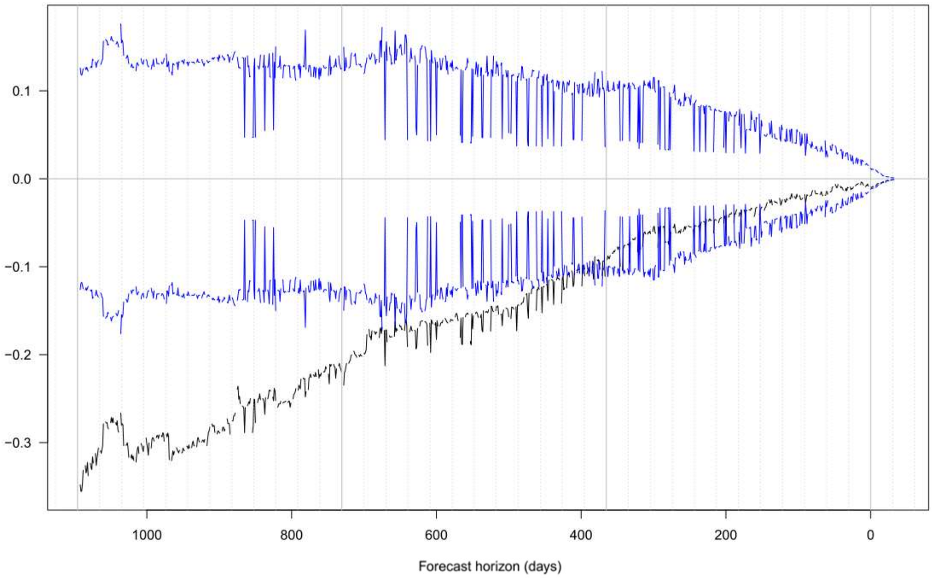
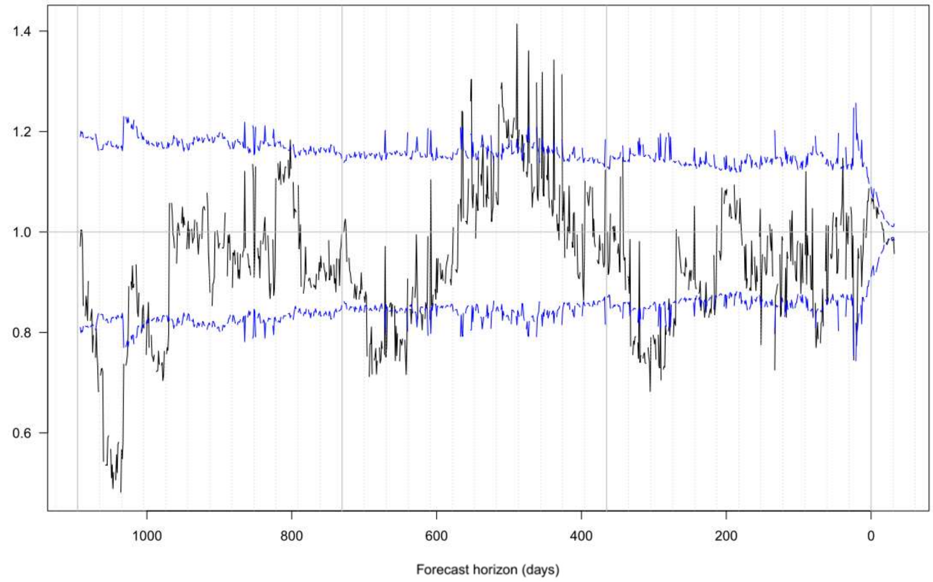


Figure 3.2.5.e Estimated slopes with corresponding critical values of Mincer-Zarnowitz regressions of forecast log-errors across forecast horizons



3.3 Relative optimality

3.3.1 log(FPI) a random walk

If we assumed the logarithm of the FPI were a random walk, we could use its last observed value as a (no-change) forecast of the future values. For horizons of six months and beyond we would thus obtain log-errors that are on average closer to zero than the log-errors of the futures contracts, as can be seen in Figure 3.3.1.a. For shorter horizons, the average log-errors of the two forecasts are about the same.

However, the average of log-errors being small does not imply the log-errors being averaged are small. Therefore, it may be more fruitful to inspect mean absolute errors (MAEs) and mean squared errors (MSEs). Figure 3.3.1.b indicates that MAEs of the futures forecasts are smaller than these of the no-change forecast for horizons of up to about 20 months. For longer horizons, the no-change forecast has lower MAEs, though the difference only becomes noticeable for the longest horizons of 30 months and above. A similar result is observed for the MSEs in Figure 3.3.1.c. The latter two figures also show that the absolute values and squares of the no-change forecast are seasonal. It is easier to forecast 10 (22, 34) months ahead than, say, 6 (18, 30) months ahead, even though the former are the longer horizons.

Figure 3.3.1.a Means of naïve no-change (blue) and futures-based (black) forecast log-errors across forecast horizons

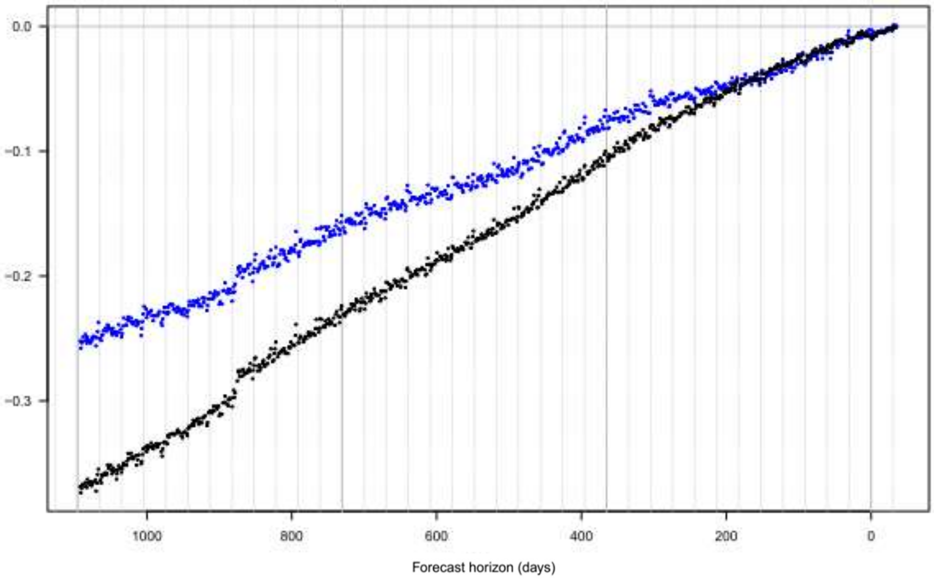


Figure 3.3.1.b Means of absolute values of naïve no-change (blue) and futures-based (black) forecast log-errors across forecast horizons

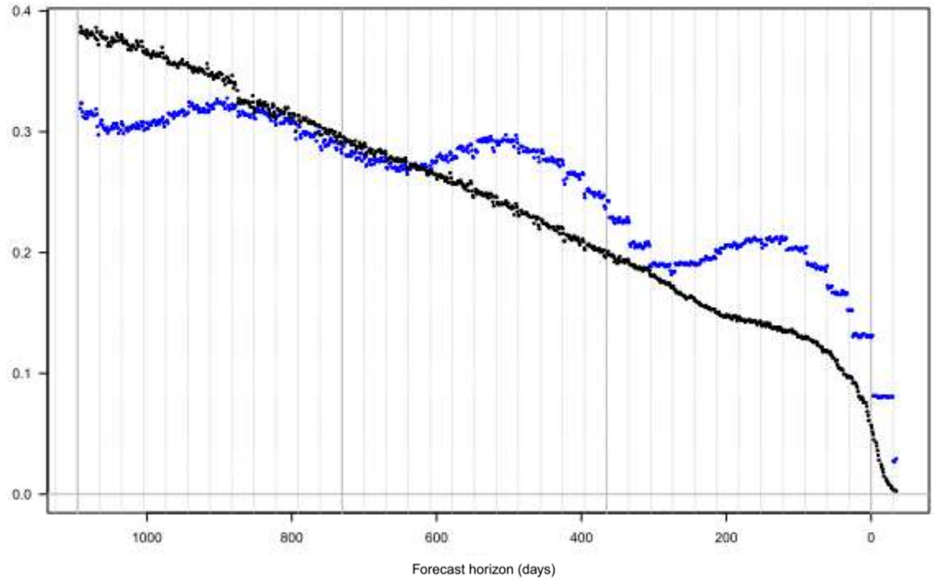
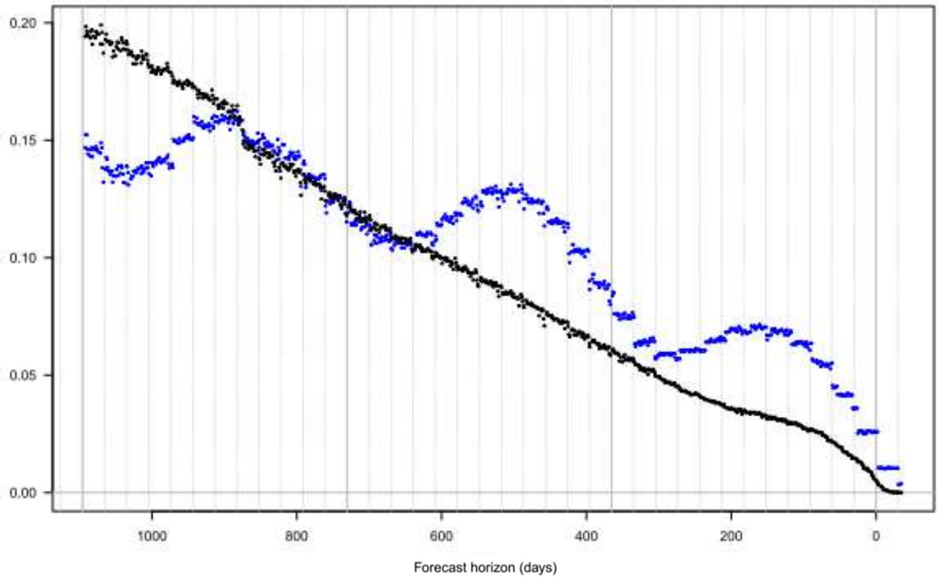


Figure 3.3.1.c Means of squares of naïve no-change (blue) and futures-based (black) forecast log-errors across forecast horizons



We conduct the Diebold-Mariano test for both absolute and squared errors across the forecast horizons to see which of the loss differentials between the benchmark no-change forecast and the futures forecasts are statistically distinguishable from zero. Their p-values are shown in Figures 3.3.1.d and 3.3.1.e. For both absolute and squared errors, the inferiority of the benchmark no-change forecast for the short and medium horizons is statistically significant, but its superiority for the longest horizons is not.

Figure 3.3.1.d P-values of Diebold-Mariano test of no-change vs. futures-based absolute forecast log-errors across forecast horizons

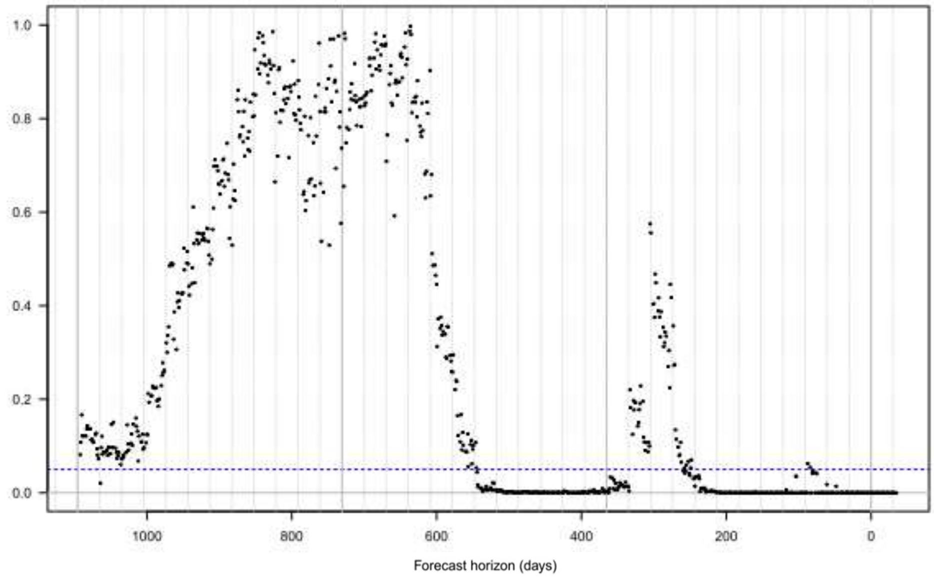
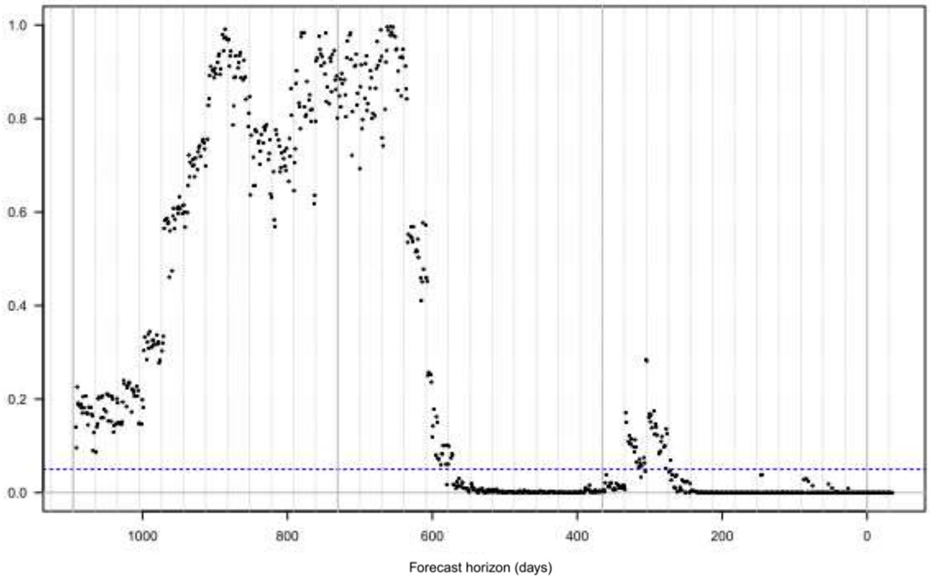


Figure 3.3.1.e P-values of Diebold-Mariano test of no-change vs. futures-based squared forecast log-errors across forecast horizons



3.3.2 Log-error a random walk or i.i.d.

If we assumed the futures forecasts' log-errors were a random walk or an i.i.d. sequence, we could try adjusting them for their last observed value (LOV) or their historical mean (HM), respectively. We would get large improvements in the average log-error; see Figures 3.3.2.a and 3.3.2.b. Looking at the MAEs (Figure 3.3.2.c), however, the LOV-adjusted forecast underperforms the futures prices for all but the longest horizons. According to the Diebold-Mariano test (Figure 3.3.2.g), the underperformance is statistically significant for most of the horizons under 18-19 months but not beyond. The HM-adjusted forecast is about as good as the unadjusted one in terms of MAE for horizons of up to 18 months and better for the longer horizons (Figure 3.3.2.d). However, only a small fraction of the expected loss differentials for the longer horizons are statistically significant as per the Diebold-Mariano test (Figure 3.3.2.h). The results for MSE largely mirror the ones for MAE (Figures 3.3.2.e, 3.3.2.f, 3.3.2.i and 3.3.2.j). Thus, the two benchmark forecasts based on models for the unconditional expectation of the log-error only offer improvements in average forecast losses for horizons beyond one and a half years, and only a small fraction of them are statistically distinguishable from zero. For shorter horizons, the benchmarks yield comparable or larger losses, some of which are statistically significant.

Figure 3.3.2.a Means of last-bias adjusted (blue) and futures-based (black) forecast log-errors across forecast horizons

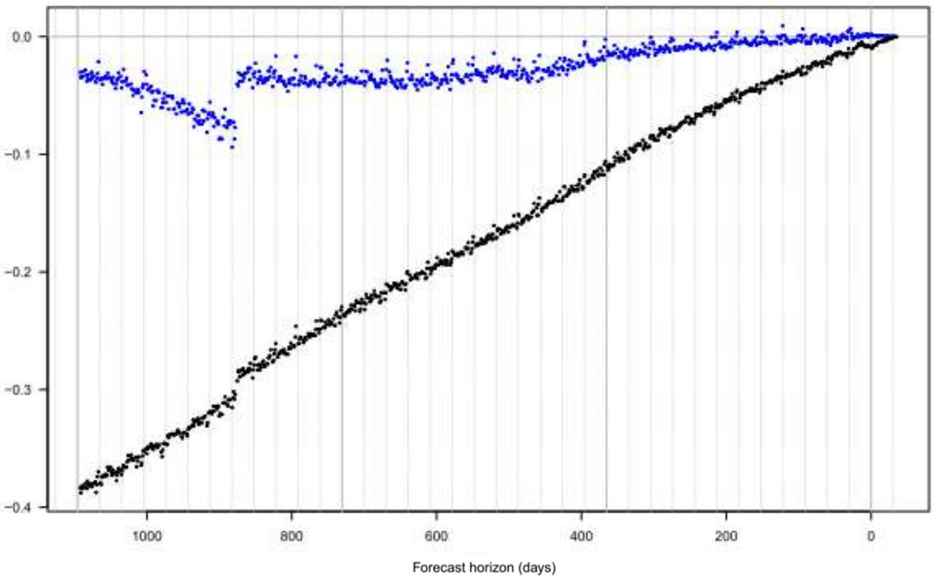


Figure 3.3.2.b Means of average-bias adjusted (blue) and futures-based (black) forecast log-errors across forecast horizons

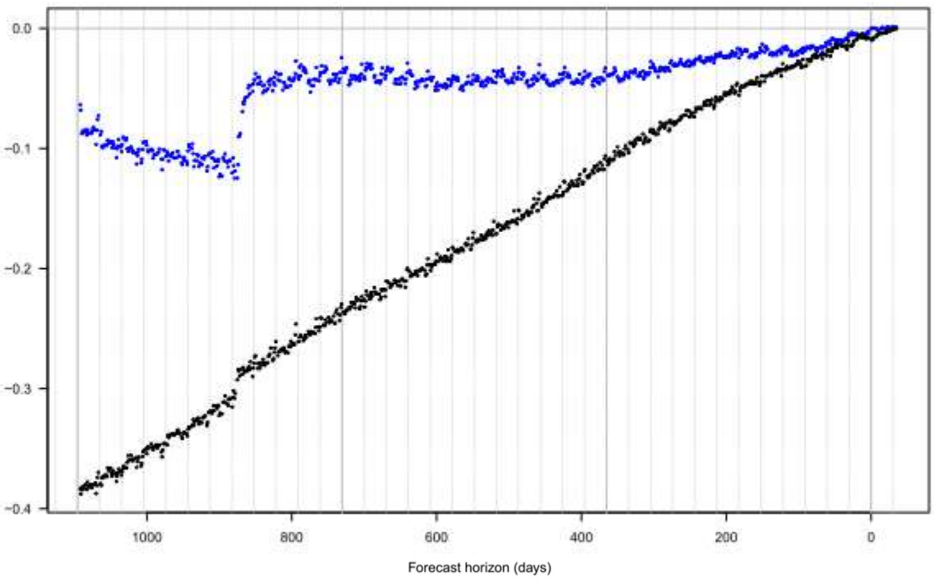


Figure 3.3.2.c Means of absolute values of last-bias adjusted (blue) and futures-based (black) forecast log-errors across forecast horizons

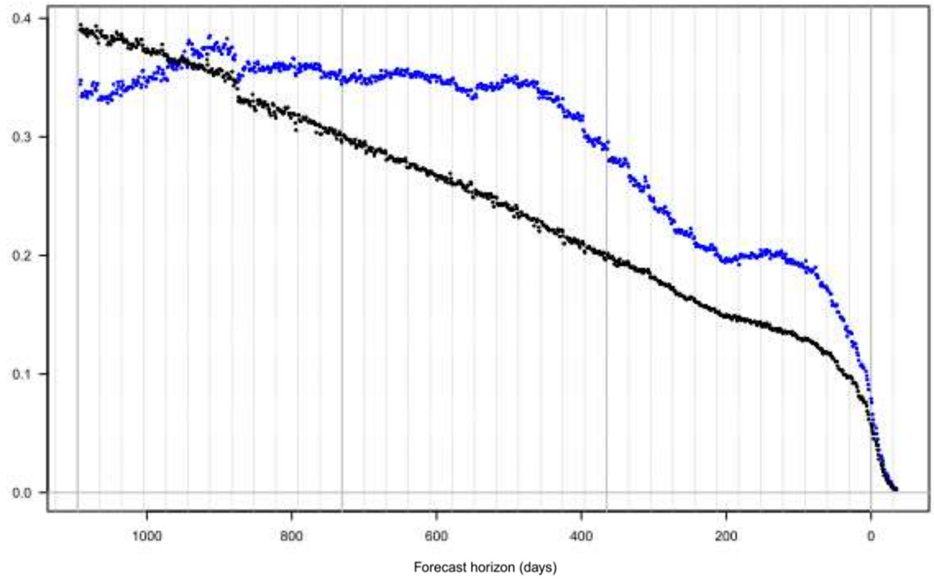


Figure 3.3.2.d Means of absolute values of average-bias adjusted (blue) and futures-based (black) forecast log-errors across forecast horizons

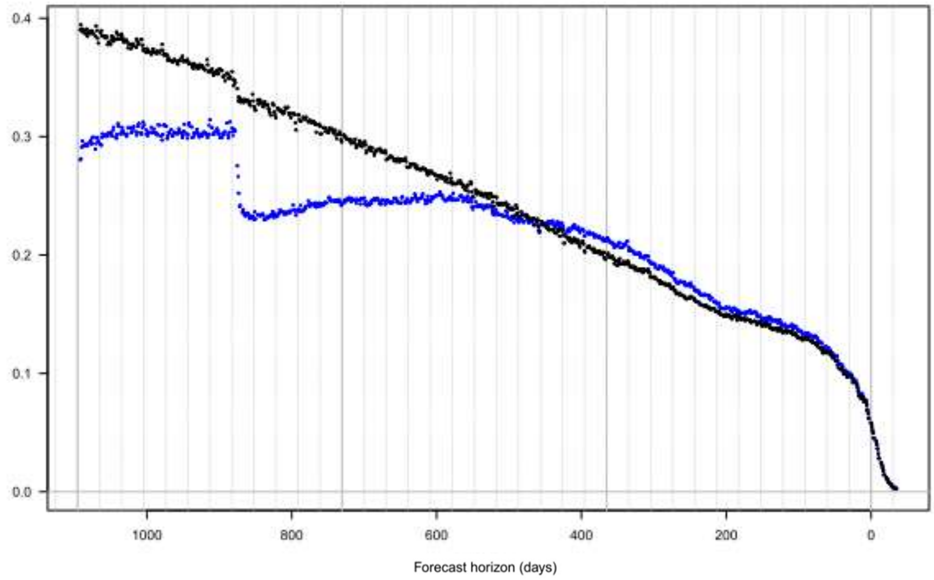


Figure 3.3.2.e Means of squares of last-bias adjusted (blue) and futures-based (black) forecast log-errors across forecast horizons

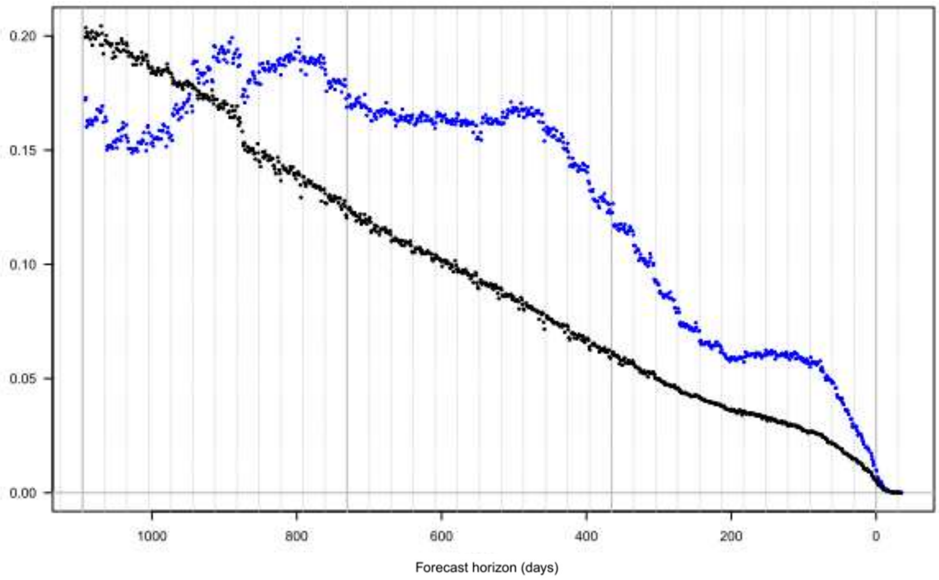


Figure 3.3.2.f Means of squares of average-bias adjusted (blue) and futures-based (black) forecast log-errors across forecast horizons

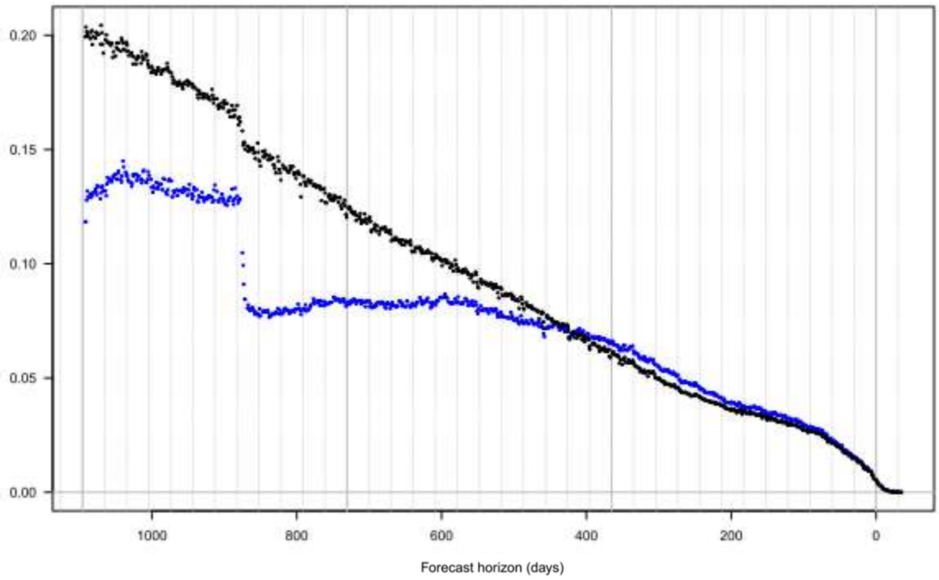


Figure 3.3.2.g P-values of Diebold-Mariano test of last-bias adjusted vs. futures-based absolute forecast log-errors across forecast horizons

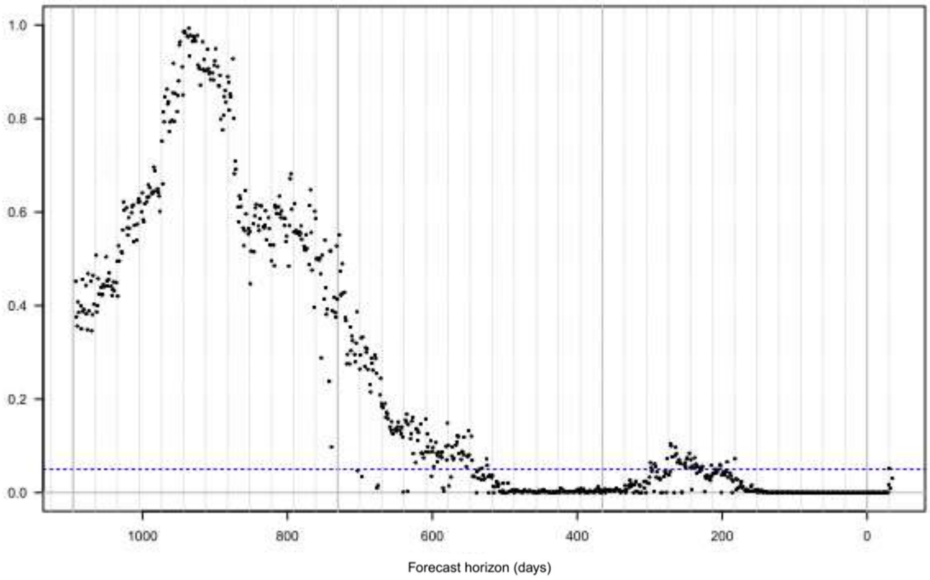


Figure 3.3.2.h P-values of Diebold-Mariano test of average-bias adjusted vs. futures-based absolute forecast log-errors across forecast horizons

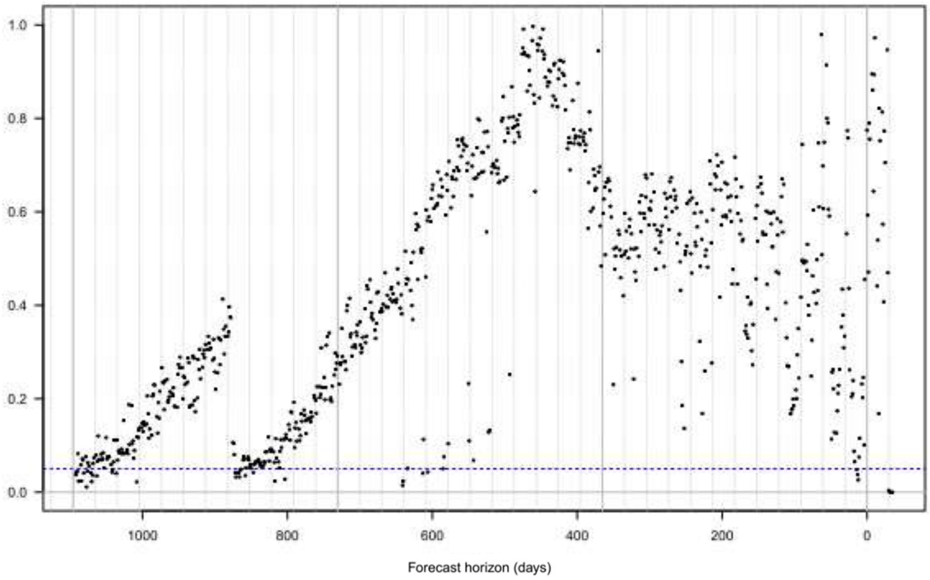


Figure 3.3.2.i P-values of Diebold-Mariano test of last-bias adjusted vs. futures-based squared forecast log-errors across forecast horizons

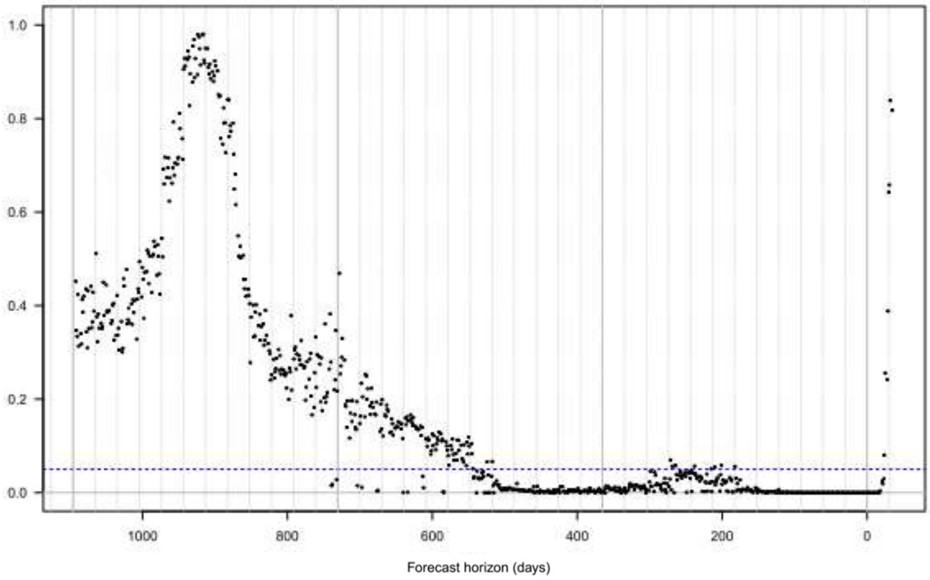
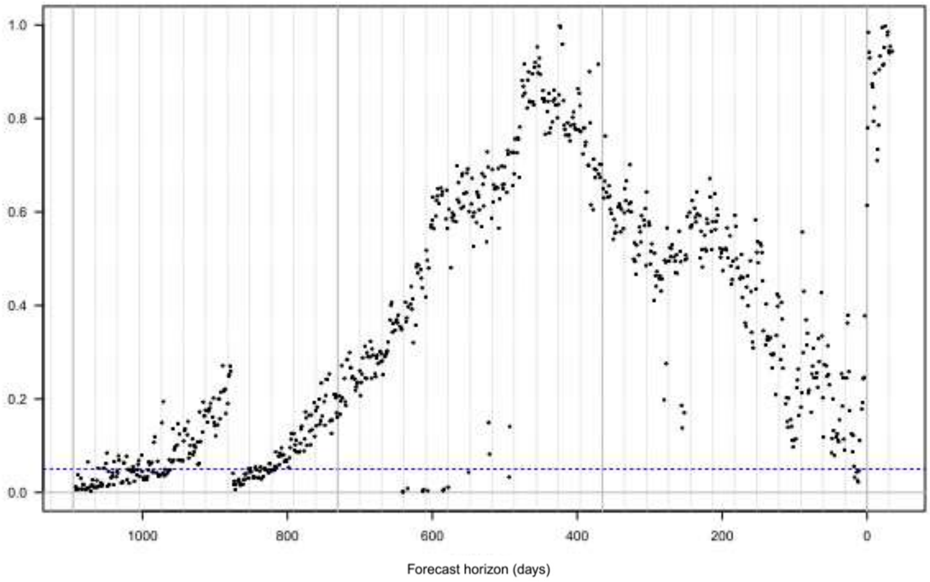


Figure 3.3.2.j P-values of Diebold-Mariano test of average-bias adjusted vs. futures-based squared forecast log-errors across forecast horizons



3.3.3 Log-error i.i.d. aside from seasonal level shifts

If we assumed the futures forecasts' log-errors followed the seasonal linear model in equation (3') with i.i.d. errors, we could adjust them for the historical means of the different underlying target months of the year (the historical seasonal means). Such bias-adjusted forecasts would produce log-errors with means, MAEs and MSEs shown in Figures 3.3.3.a, 3.3.3.b and 3.3.3.c. Substantial improvement over the unadjusted forecast is apparent with respect to all three metrics, the performance gap increasing with the forecast horizon. MAEs and MSEs show consistent superiority of the adjusted forecast for horizons above nine months. However, the estimation precision is insufficient to distinguish the expected loss differentials from zero in the Diebold-Mariano test

(Figures 3.3.3.d and 3.3.3.e), except for MSE at 26–29-month horizons. Thus, the benchmark forecast based on a seasonal model of the log-errors' conditional expectation does not convincingly beat the unadjusted forecast in terms of expected losses.

(A break in the values appearing between 29- and 30-months horizon stands out compared to the plots in the previous subsections. It is an artefact of shortening of the evaluation period; we use the first 60 observations to estimate the 12 slope coefficients and start evaluating the forecasts only from observation 61 and onward. Since we are using the same evaluation period for both the adjusted and the unadjusted forecasts, the break is seen in the values of both series.)

Figure 3.3.3.a Means of conditional bias-adjusted (blue) and futures-based (black) forecast log-errors across forecast horizons

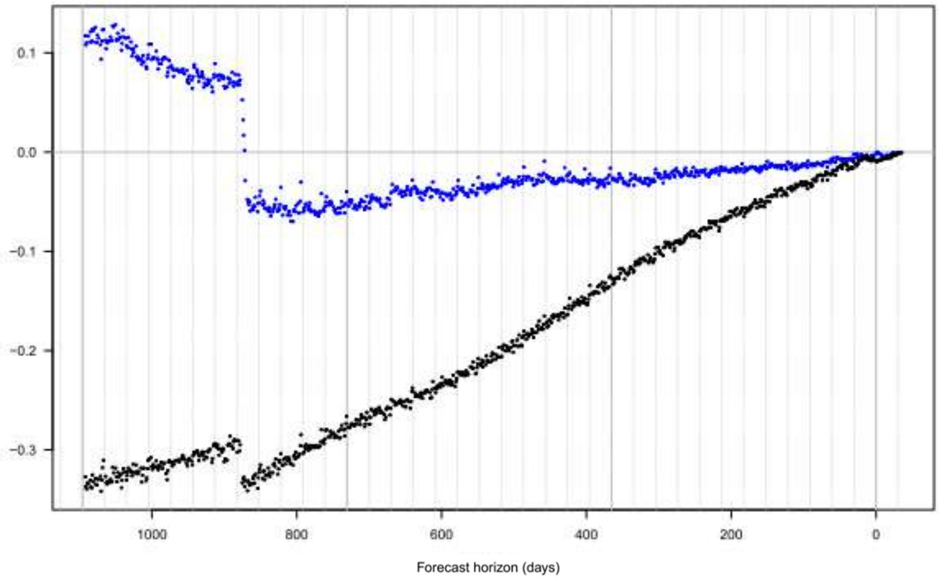


Figure 3.3.3.b Means of absolute values of conditional bias-adjusted (blue) and futures-based (black) forecast log-errors across forecast horizons

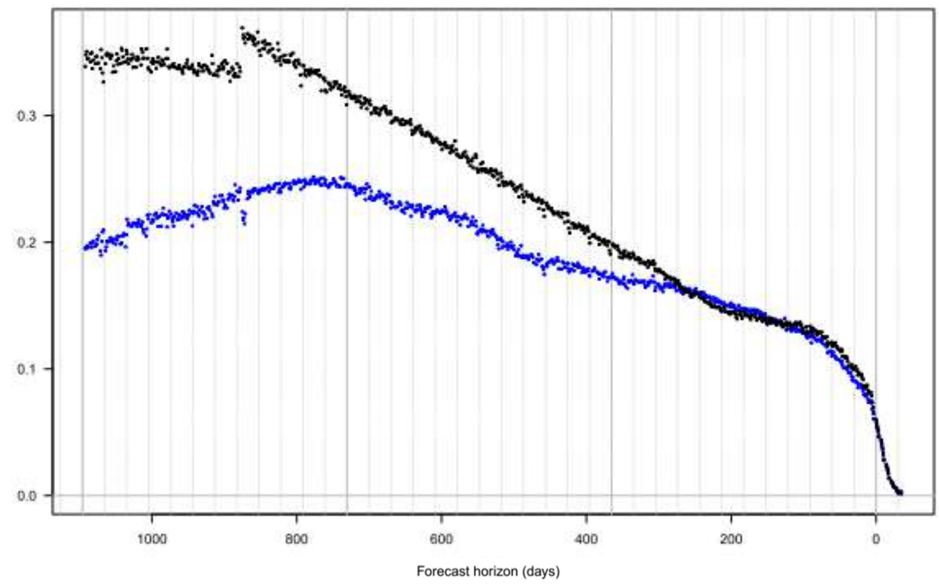


Figure 3.3.3.c Means of squares of conditional bias-adjusted (blue) and futures-based (black) forecast log-errors across forecast horizons

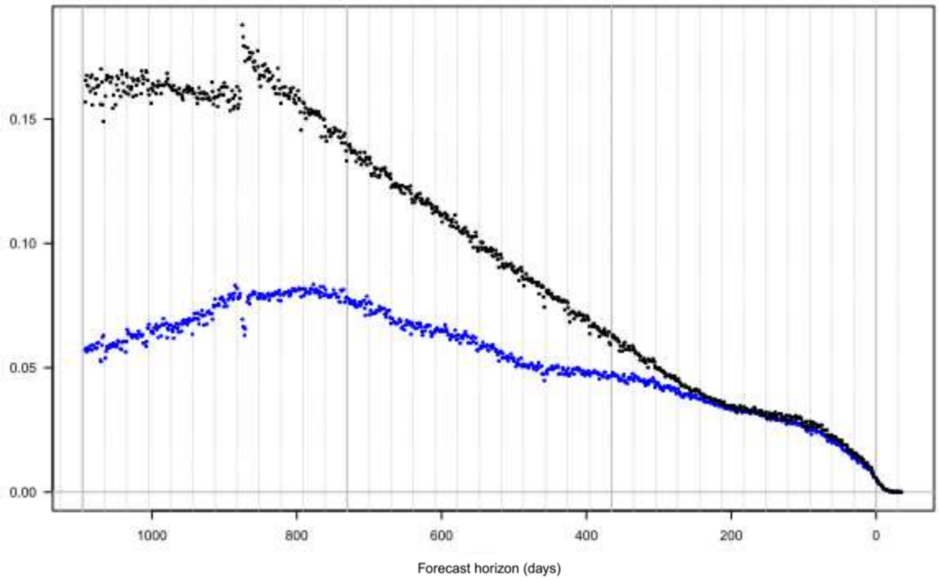


Figure 3.3.3.d P-values of Diebold-Mariano test of conditional bias-adjusted vs. futures-based absolute forecast log-errors across forecast horizons

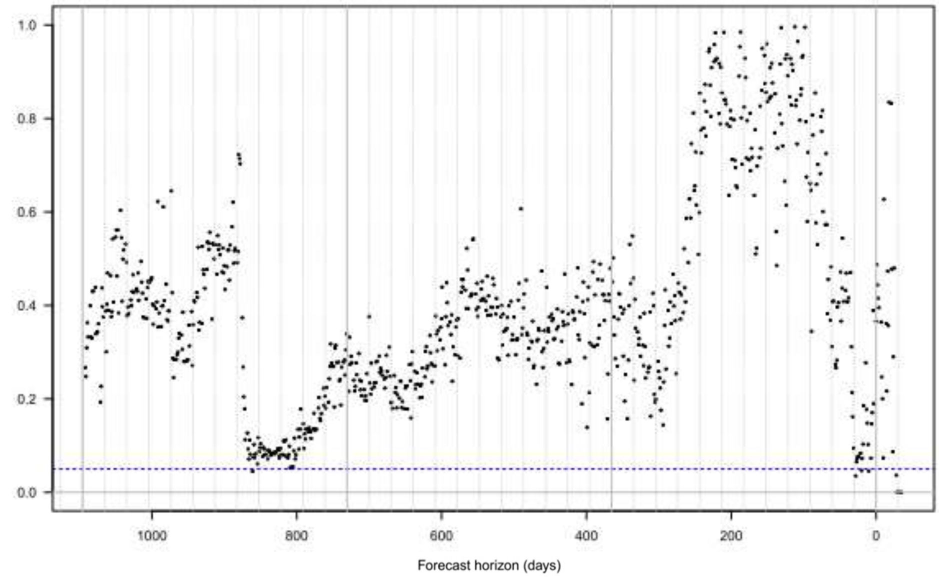
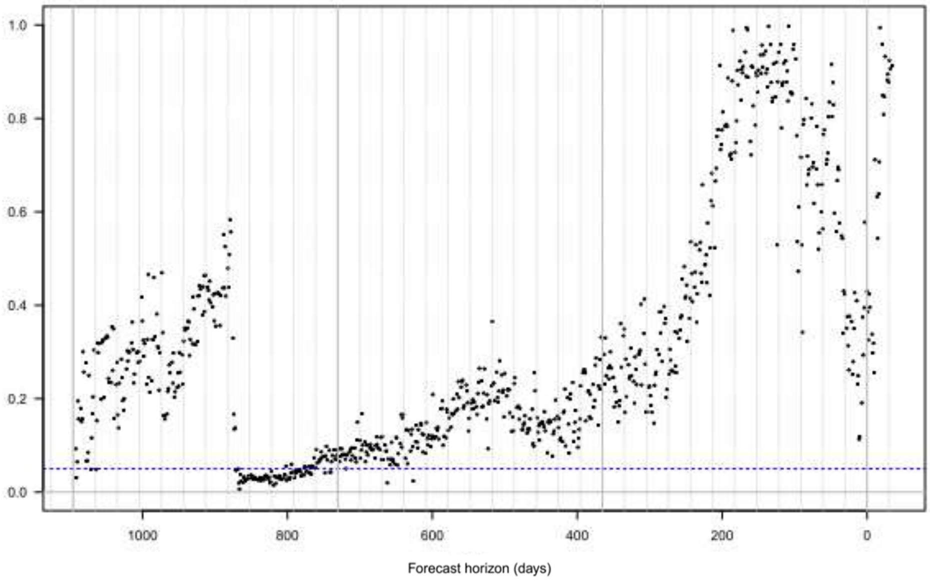


Figure 3.3.3.e P-values of Diebold-Mariano test of conditional bias-adjusted vs. futures-based squared forecast log-errors across forecast horizons

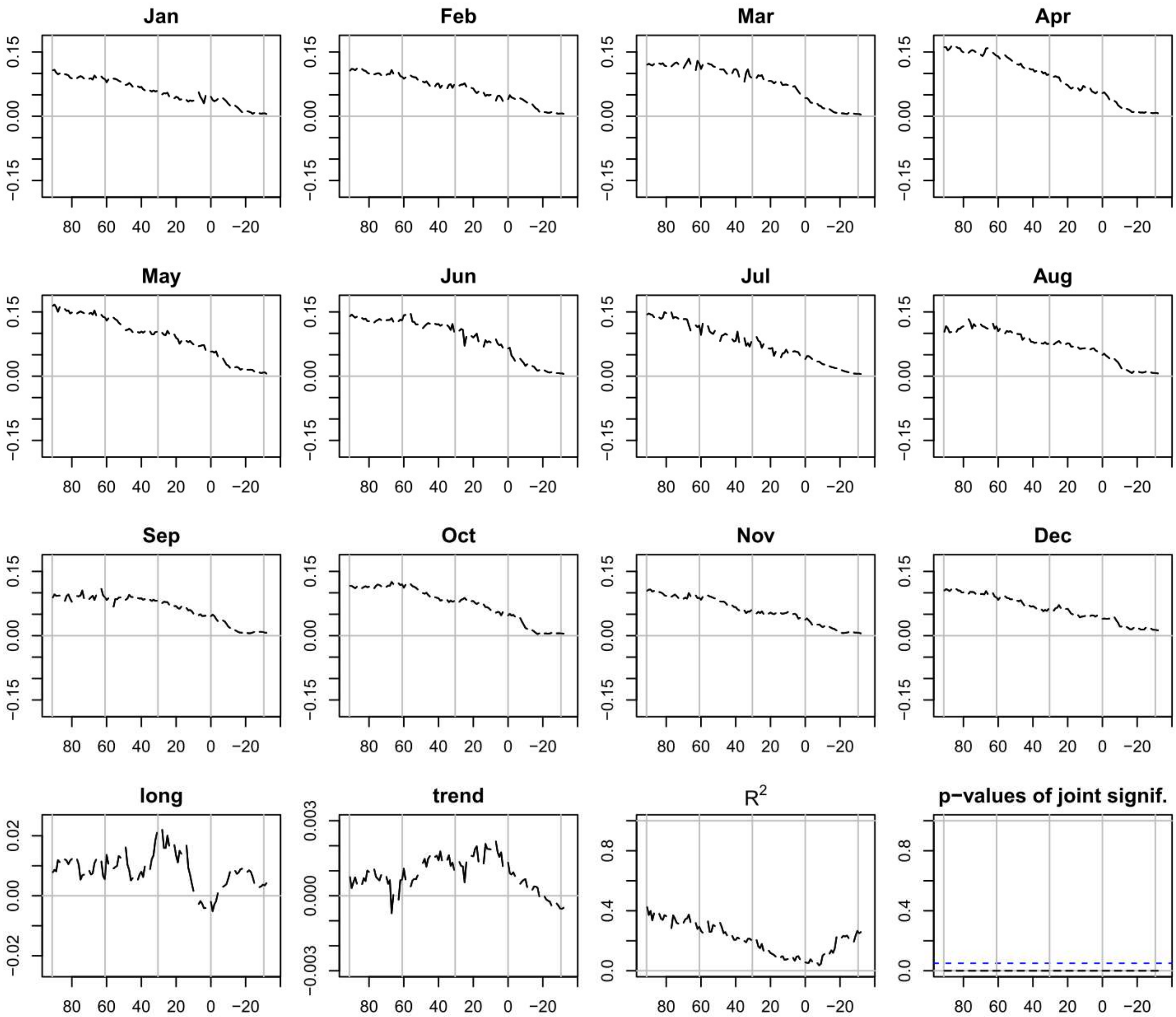


3.4 Scale of forecast errors

If we model the absolute value of the log-error as a linear function of the underlying target month of the year, the length of the underlying target month and calendar time, we get coefficient estimates, p-values of their joint significance and coefficients of determination shown in Figure 3.4.a. The size of the error increases substantially with horizon in every underlying target month of the year. The largest log-errors of about 0.15 are observed in April to July for the longest three-month horizon. For the same horizon, the smallest errors are about 0.10 in September to February. Longer underlying months tend to produce larger errors, but the effect is small. The estimated slope coefficient of the linear time trend varies somewhat erratically across the forecast horizons. The models' explanatory power ranges between close to zero for slightly negative horizons and about 0.40 for the longest three-month horizon. While this is considerably lower than in the case of regressions for the levels of the same log-errors, the coefficients are jointly significant at all horizons (more so than for the levels).[7]

[7] While it may not be quite clear from the graph, all p-values in Figure 3.4.a are virtually equal to zero.

Figure 3.4.a Coefficient estimates, coefficients of determination and p-values of the coefficients' joint significance from regressions with ARMA errors due to equation (8) across forecast horizons



4. Conclusion

We set out to evaluate salmon futures prices as point forecasts of the future spot price. We assessed forecast optimality for horizons between essentially zero and three years ahead both in absolute terms and relative to four simple benchmark forecasts. The futures prices display a major indication of absolute suboptimality. They are downward biased, with the size of the bias increasing exponentially with the forecast horizon.[8] The bias is statistically distinguishable from zero for horizons above 14-15 months, where its size reaches between -13% and -31%. For horizons of up to three months, the bias can be explained by the length of the underlying month, a linear time trend and monthly seasonality, the latter contributing considerable variation in the level of the forecast errors. The sample span is too short to extend this analysis to longer horizons. On the other hand, the futures prices mostly meet two other criteria of absolute optimality. They do not show

[8] As the negative mean of log-errors grows linearly with the forecast horizon.

autocorrelations beyond what is consistent with chance variation, and the error variance is nondecreasing with the forecast horizon except for the longest horizons of over 30 months.

The forecast bias is easy to reduce by adjusting the futures prices for the forecast errors' historical mean, historical seasonal mean (based on the month of the year) or, for the best result, the last observed value. However, the latter adjustment increases the mean absolute error (MAE) and mean squared error (MSE) considerably for all but the longest horizons. This is not the case for historical mean or historical seasonal mean that both offer reductions in MAE and MSE for horizons beyond 18 and nine months, respectively. However, most of the reductions in the expected absolute or square forecast loss are not statistically distinguishable from zero. Thus, for an agent interested in the most accurate forecast of the future spot price of salmon up to nine months ahead, there is no reason to replace the futures price with any of the simple benchmarks. For longer horizons, historical seasonal mean may be a good alternative, even if its superiority cannot be established with a high degree of statistical confidence.

Regarding limitations of this study, Fish Pool announced in April 2024 that its contracts would be denominated in euros instead of Norwegian kroner and traded on Euronext's commodity market from 22 July 2024. The company believes these steps will increase both trading volume and liquidity in the futures market. The quality of spot price forecasts derived from the futures prices may also change, limiting the applicability of our findings to the new state of the markets and requiring a new investigation. However, assessing the forecast properties of the longer-term contracts after these changes will require several years' worth of data past July 2024. The problem is not specific to salmon futures and remains a fundamental challenge in trying to obtain up to date knowledge on long-term forecast quality.

5. Acknowledgements

The author is indebted to Mindaugas Bloznelis and Erik Smith-Meyer for their helpful comments and discussions. There being no co-authors, the author assumes full responsibility for any omissions or commissions.

Bibliography

Asche, F., & Bjørndal, T. (2011). *The Economics of Salmon Aquaculture*. Chichester: John Wiley & Sons.

Asche, F., & Guttormsen, A. G. (2001). Patterns in the relative price for different sizes of farmed fish. *Marine Resource Economics*, *16*(3), 235-247.

Baillie, R. T., Diebold, F. X., Kapetanios, G., Kim, K. H., & Mora, A. (2024). *On robust inference in time series regression* (No. w32554). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w32554/w32554.pdf

Bergfjord, O. J. (2007). Is there a future for salmon futures? An analysis of the prospects of a potential futures market for salmon. *Aquaculture Economics & Management*, *11*(2), 113-132.

Bloznelis, D. (2016a). *Management of short-term price uncertainty in the salmon spot market* [Doctoral dissertation, Norwegian University of Life Sciences]. Brage. https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/2497330

Bloznelis, D. (2016b). *Price and volatility forecasting for risk management in the Fish Pool market* [Lecture]. Ås, Norway. https://www.nmbu.no/forskning/disputaser/1-july-phd-defence-daumantas-bloznelis-hh

Bloznelis, D. (2016c). Salmon price volatility: a weight-class-specific multivariate approach. *Aquaculture Economics & Management, 20*(1), 24-53.

Bloznelis, D. (2018a). Hedging salmon price risk. *Aquaculture Economics & Management, 22*(2), 168-191.

Bloznelis, D. (2018b). Short-term salmon price forecasting. *Journal of Forecasting, 37*(2), 151-169.

Diebold, F. X. (2024). *Forecasting*. Department of Economics, University of Pennsylvania, http://www.ssc.upenn.edu/~fdiebold/Textbooks.html

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

Directorate of Fisheries (2024). *Grow out production: Sale 1994-2022* [Data set]. https://www.fiskeridir.no/English/Aquaculture/Statistics/Atlantic-salmon-and-rainbow-trout/grow-out-production

Ewald, C. O. (2013). Derivatives on nonstorable renewable resources: fish futures and options, not so fishy after all. *Natural Resource Modeling*, *26*(2), 215-236.

Fish Pool. (2015). *Årsrapport 2014*. Oslo/Bergen/Florø: Fish Pool.

Fish Pool. (2016). *Årsrapport 2015*. Oslo: Fish Pool.

Fish Pool. (2017). *Årsrapport 2016*. Oslo: Fish Pool.

Fish Pool. (2018). *Årsrapport 2017*. Bergen: Fish Pool.

Fish Pool. (2019). *Årsrapport 2018*. Oslo: Fish Pool.

Fish Pool. (2020). *Annual accounts 2019*. Oslo: Fish Pool.

Fish Pool. (2021). *Annual report 2020*. Bergen, Florø, Oslo, Amsterdam: Fish Pool.

Fish Pool. (2022). *Annual report 2021*. Oslo: Fish Pool.

Fish Pool. (2023). *Annual report 2022*. Oslo: Fish Pool.

Fish Pool. (2024). *Info-flash: Listing of Salmon Futures on Euronext Paris Commodity Derivatives Market*. https://fishpool.eu/press-release-listing-of-salmon-futures-on-euronext-paris-commodity-derivatives-market/

French, K. R. (1986). Detecting spot price forecasts in futures prices. *Journal of Business,* S39-S54.

Gu, G., & Anderson, J. L. (1995). Deseasonalized state-space time series forecasting with application to the US salmon market. *Marine Resource Economics*, *10*(2), 171-185.

Guttormsen, A. G. (1999). Forecasting weekly salmon prices: risk management in fish farming. *Aquaculture Economics & Management*, *3*(2), 159-166.

Hyndman, R. J. (2010, October 4). The ARIMAX model muddle. *Hyndsight*. https://robjhyndman.com/hyndsight/arimax/

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts. https://otexts.com/fpp2/

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, *27*, 1-22.

Lin, B. H., Herrmann, M., Lin, T. Y., & Mittelhammer, R. C. (1989). Forecasting the price of farmed Atlantic salmon: an integrated econometric and time series approach. *Agribusiness*, *5*(5), 477-488.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance* (pp. 3-46). NBER.

Misund, B., & Asche, F. (2016). Hedging efficiency of Atlantic salmon futures. *Aquaculture Economics & Management*, *20*(4), 368-381.

Mowi. (2020). *Salmon Farming Industry Handbook 2020*. https://corpsite.azureedge.net/corpsite/wp-content/uploads/2020/06/Mowi-Salmon-Farming-Industry-Handbook-2020.pdf

Øglend, A. (2013). Recent trends in salmon price volatility. *Aquaculture Economics & Management*, *17*(3), 281-299.

Schütz, P., & Westgaard, S. (2018). Optimal hedging strategies for salmon producers. *Journal of Commodity Markets*, *12*, 60-70.

Vukina, T., & Anderson, J. L. (1994). Price forecasting with state-space models of nonstationary time series: case of the Japanese salmon market. *Computers & Mathematics with Applications*, *27*(5), 45-62.

Statistisk sentralbyrå. (2020a). *07326: Aquaculture. Sales of slaughtered fish for food, by fish species (C) (closed series) 1976 – 2019* [Data set]. https://www.ssb.no/en/statbank/table/07326/chartViewColumn/

Statistisk sentralbyrå. (2020b). *Aquaculture (terminated in Statistics Norway)* [Data set]. https://www.ssb.no/en/fiskeoppdrett

Appendix

A.1 Time aspects of forecast errors

When modelling forecast errors due to the salmon futures prices, two complications arise: (1) the forecast horizons are overlapping and (2) the observations are irregularly spaced in time. For example, consider using futures prices to forecast the spot price for January 2024, February 2024, March 2024 and April 2024. Let each forecast be made 78 hours ahead of the start of the underlying target month that the settlement price of a Fish Pool futures contract is based on; see Table A.1 for details.

Table A.1. Example of the time frame of Fish Pool salmon futures contracts and forecast errors derived from them

Forecast origin (also date of futures price)	Time difference from the previous row	Forecast horizon[9]	Full forecast horizon[10]	Name of underlying month	Start of underlying month	End of underlying month	Length of underlying month
…	…	…	…	…	…	…	…
2023-12-28 at 18:00	…	3 days	38 days	January 2024	2024-01-01	2024-02-04	5 weeks
2024-02-01 at 18:00	35 days	3 days	31 days	February 2024	2024-02-05	2024-03-03	4 weeks
2024-02-29 at 18:00	28 days	3 days	31 days	March 2024	2024-03-04	2024-03-31	4 weeks
2024-03-28 at 18:00	28 days	3 days	31 days	April 2024	2024-04-01	2024-04-28	4 weeks
…	…	…	…	…	…	…	…

Define the time index of the forecast errors by the forecast origin. (Other options would be to define it by (1) the end of the underlying target month when error values are determined or (2) three days after that when error values become publicly available.) The forecast errors are thus indexed on 2023-12-28, 2024-02-01, 2024-02-29 and 2024-03-28, the same as the futures prices they are derived from. The corresponding spot prices are the FPI values for the underlying target months of January, February, March and April.

Events occurring on 2024-02-02 that affect the spot price can affect the FPI for both January and February. In this sense, the two forecast errors of 2023-12-28 and 2024-02-01 are overlapping. Similarly, events occurring on 2024-03-01 that affect the spot price can affect the FPI for both February and March, so the two forecast errors of 2024-02-01 and 2024-02-29 are overlapping. Finally, events occurring on 2024-03-29 that affect the spot price can affect the FPI for both March and April, making the two errors of 2024-02-29 and 2024-03-28 overlapping.

[9] Time until the start of the underlying month.

[10] Time until the end of the underlying month.

Moreover, the time difference between the first two errors (or the first two forecast origins) is 35 days, while the time differences within the second and third pairs of errors are both 28 days. Thus, the observations are irregularly spaced in time.

Both overlapping observations and irregular spacing present challenges for the empirical analysis. First, overlapping observations tend to be positively autocorrelated, as they are affected by the same events and usually in a similar way. Second, the time series measures (including autocorrelation) and models (such as white noise or MA($h-1$)) discussed in section 2 are typically defined under regular (or equal) spacing in time. Irregular spacing complicates the definitions and requires adjustments of the forecast optimality criteria.

A.2 Irregular spacing

To address the problem of irregular spacing, we may consider resorting to state space models for daily data or to continuous-time models. The daily data would have streaks of 28-1=27 or 35-1=34 unavailable (NA) observations with a single observed value in between each pair of consecutive streaks. The high proportion of missing values would hinder estimation of the model's parameters. Also, the model would be based on questionable assumptions about what happens between the available observations. Otherwise, we could consider continuous-time models such as the Ornstein–Uhlenbeck process. However, they are not flexible enough to accommodate the rich autocorrelation dynamics that may be present in our data. Thus, these solutions do not appear satisfactory.

The other alternative is to approximate irregularly spaced data by a model for equally spaced observations. For simplicity and computational feasibility, we will ignore the difference between 28 and 35 days and proceed as if the spacing were perfectly regular. This should not create major distortions as long as the spacing is approximately regular, as is in our example; an underlying month is always either 28 or 35 days. Occasionally, the time difference between consecutive observations is greater than 35 days, e.g. 28+28=56 days or 28+35=63 days. This is mainly due to holidays when the futures market is closed, and the price quotes are not available. In such cases we will impute NA values, so that the time series become approximately equally spaced. Treating the resulting time series as equally spaced will allow the use of standard measures and models that are easier to estimate and/or more flexible than the alternatives discussed above.

A.3 Overlapping observations

Regarding the overlapping horizons of consecutive forecasts, we will proceed in one of two ways depending on the question we are addressing. First, we may keep and use all observations despite their overlap but adjust the uncertainty estimates to reflect the approximate effective sample size. For example, when estimating the expected value of overlapping errors for a given horizon, we will use all observations. To account for the autocorrelation due to the overlap, we will use regression with autoregressive and moving-average (ARMA) errors similarly to how Bailie et al. (2024) use dynamic regression in a procedure called DURBIN. We prefer regression with ARMA errors to dynamic regression, as the former facilitates a more convenient interpretation of the parameter estimates (Hyndman, 2010) and allows using the auto.arima algorithm (Hyndman & Khandakar, 2008) for computationally efficient automated selection of the lag orders of the autoregressive and the moving-average components. When there are no regressors besides the intercept such as when modelling the unconditional expected value of logarithmic forecast errors (log-errors), there is no difference between using regression with ARMA errors and DURBIN.

Second, when assessing whether h-step-ahead forecast errors are at most MA($h-1$) processes, we will remove a minimal number of observations to avoid the overlap. In the example due to Table A.1 above, it is sufficient to remove the rows corresponding to the February and April contracts. More generally, consider a fixed forecast horizon and a series of forecast errors where NA values have already been imputed where necessary to obtain an approximately equal spacing in time. We will keep the first observation and remove all other observations that overlap with it. We will move on to the next remaining observation and do the same. We will reiterate the procedure until we exhaust our sample. After having removed the overlapping observations, one may wonder if irregular spacing could become an issue again. However, we expect it to be minor. For example, when removing every other forecast error from the set of the 3-day horizon errors as in Table A.1, we will end up with observations that are either 28+28=56 or 28+35=63 days apart. This is a series of approximately equally spaced, nonoverlapping forecast errors. The procedure will be done separately for different forecast horizons, resulting in different, horizon-specific subsamples of the original data.

A.4 Steps vs. horizons

To translate the forecast optimality properties from regularly spaced time series to irregularly spaced ones, one must again consider the overlap. In this context, a nonpositive forecast horizon of $h \leq 0$ days corresponds to a 1-step-ahead forecast. That is, errors from an optimal forecast for $h \leq 0$ days must be white noise. Meanwhile, a positive forecast horizon $h > 0$ approximately corresponds to $\left\lceil \frac{h+3}{365.25/12} \right\rceil + 1$ forecast steps. Here, $\lceil x \rceil$ is the integer that x rounds up to (e.g. $\lceil 1.01 \rceil = \lceil 1.1 \rceil = \lceil 1.5 \rceil = \lceil 1.9 \rceil = 2$), and $365.25/12$ is the average length of an underlying target month. $\left\lceil \frac{h+3}{365.25/12} \right\rceil + 1$ is the number of whole months (rounded up) from the forecast origin (time t) to the end of the underlying T_{end} plus three days, i.e. when the price of the underlying becomes publicly available.

For example, when forecasting 3 days ahead of the start of the underlying ($h = 3$), we could consider ourselves to be forecasting 2 steps ahead, so that the optimal forecast would produce errors that have positive autocorrelation up to lag 2-1=1 and zero autocorrelation at lags 2, 3, …. Similarly, when forecasting 40 days ahead of the start of the underlying ($h = 40$), we could consider ourselves to be forecasting 3 steps ahead, so that the optimal forecast would produce errors that have positive autocorrelation up to lag 3-1=2 and zero autocorrelation at lags 3, 4, …. This is approximate, as the length of the underlying month can be either 4 or 5 weeks, so not exactly $365.25/12$.

Moreover, when forecasting 30 days ahead ($h = 30$), we could be within the underlying month that precedes the underlying target month, or we could be within one month before that. We could thus be generating positive autocorrelation either up to lag 2-1=1 or up to lag 3-1=2, though the autocorrelation would be small in the latter case. We choose to tolerate these minor overlaps here, as otherwise we would have to remove a substantial fraction of the observations, thereby lowering estimation precision and reducing the power of hypothesis tests. We regard the minor overlaps to be a negligible problem in practice and a lesser evil than the removal of a large fraction of observations.