

Forecasting for Economics and Business

Lecture 2: Generating and Evaluating Forecasts

David Ubilava
University of Sydney

Pseudo forecasting routine



When we make forecasts, we would want to have some idea of how accurate they are.

We can get an idea about forecast accuracy by comparing the forecast for some future period, $\hat{y}_{t+h|t}$, with the actual future realization, y_{t+h} .

But to assess forecast accuracy we need the data from the future - a seemingly unfeasible task.

So, we can resort to the so-called *pseudo-forecasting* routine.



In-sample and out-of-sample data segments

The pseudo forecasting routine involves splitting the available data into two segments referred to as 'in-sample' and 'out-of-sample'

- We use the in-sample segment of a series to decide on the model and estimate its parameters.
- We use the out-of-sample segment of a series to compare the forecasts, generated based on the data from the in-sample segment, to the realizations of the variable.



Genuine pseudo-forecasts

Thus, we make the so-called "genuine" forecasts using only the information from the past and present, and we assess the accuracy of these forecasts using the information from the future.

Such chronological ordering is usually relevant, because forecasting is often performed in a time series context. In non-dynamic settings, e.g., if we use one subset of data to predict the other subset of data, such chronological ordering may not be needed.



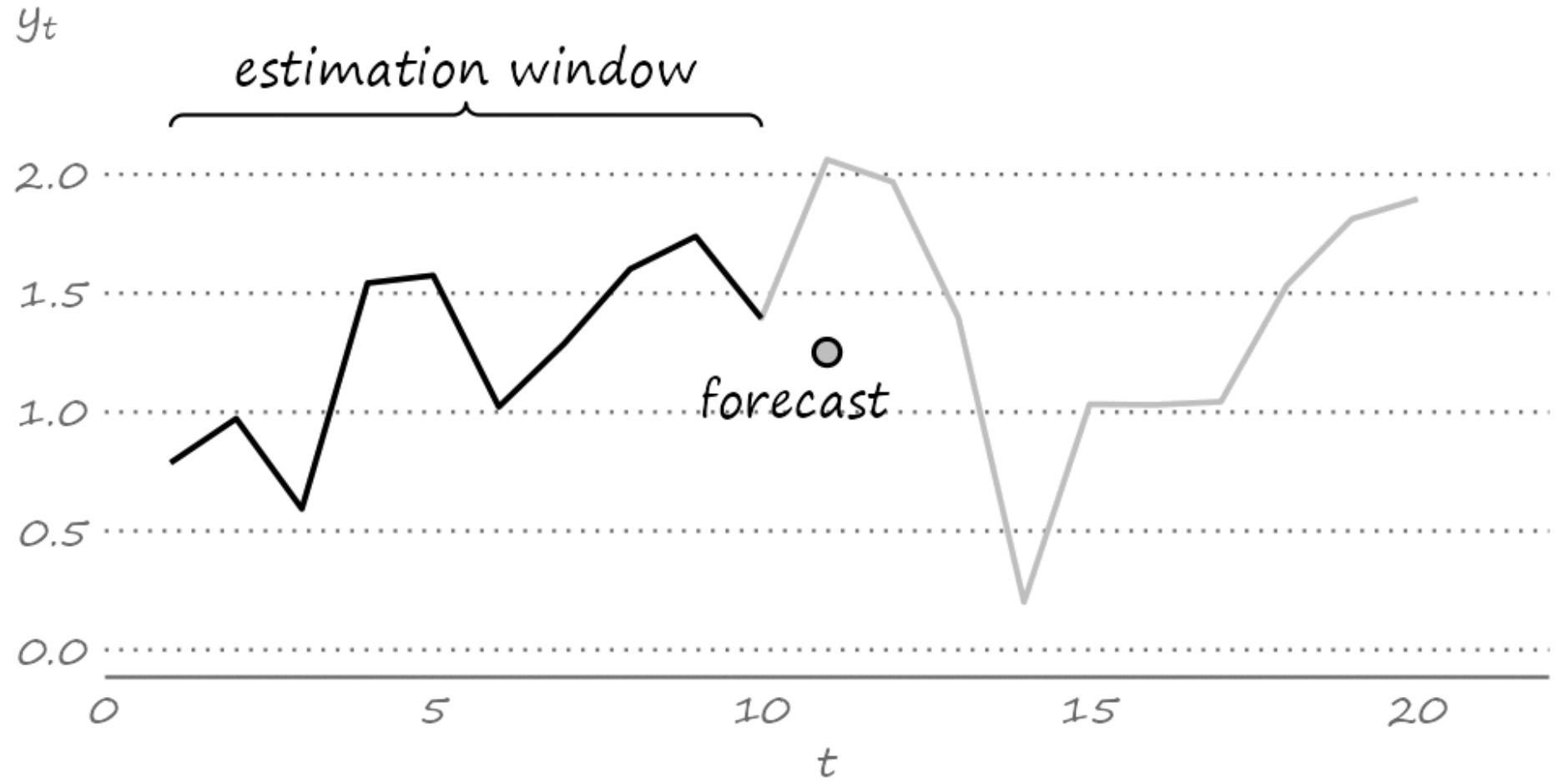
Recursive and rolling windows

As we carry out the pseudo-forecasting routine, we update the information set by including the more recent observations. In so doing, we may choose to keep or discard the more distant observations, which leads to two possible ways of selecting and updating information sets.

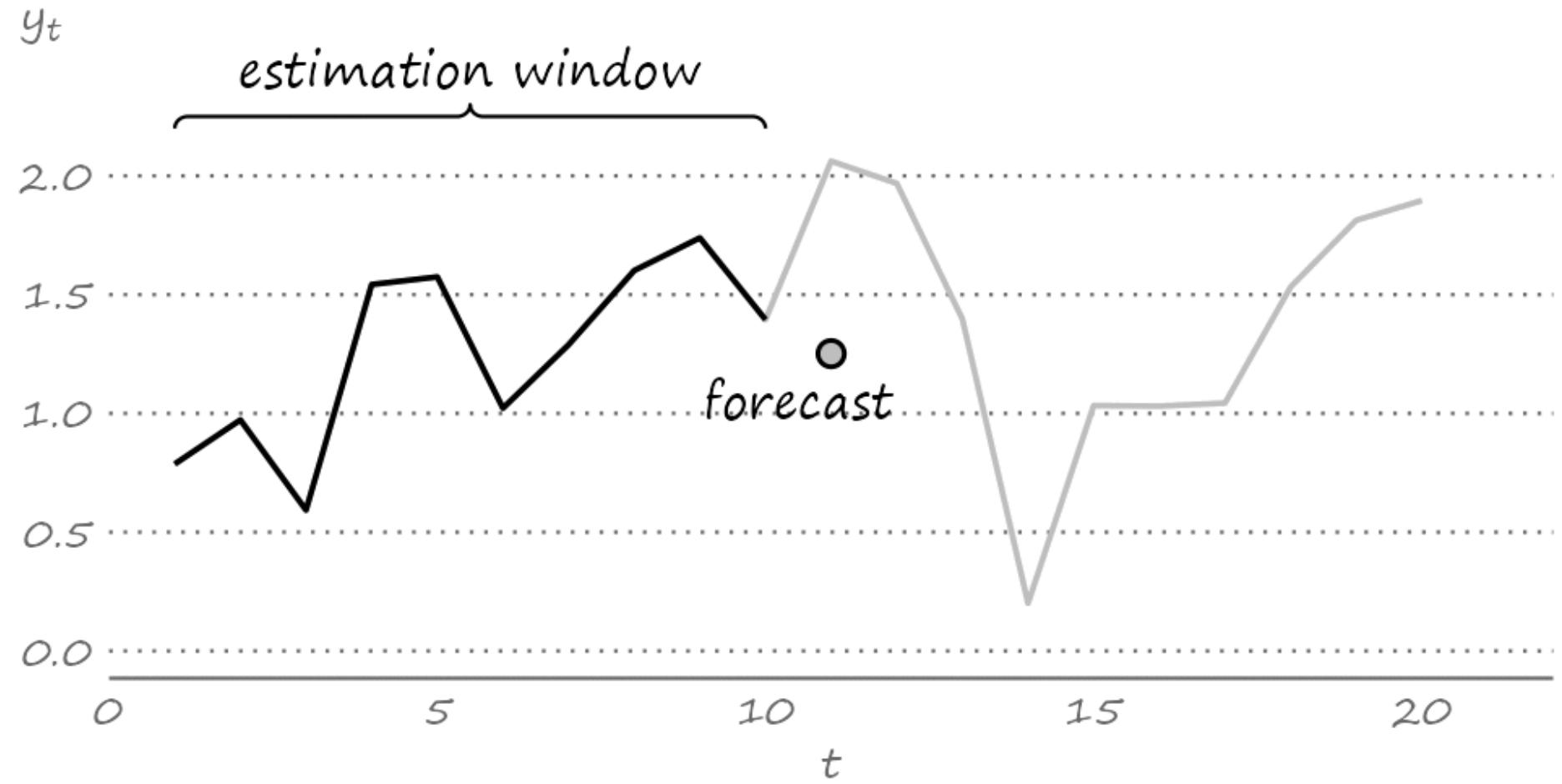
- The **recursive** windows approach uses a sequence of expanding windows to update model estimates and the information set.
- The **rolling** windows approach uses a sequence of windows of the same size to update model estimates and the information set.



Recursive windows



Rolling windows



Recursive vs. rolling windows

These two pseudo-forecasting routines are technically very similar. But they can generate considerably differing sets of forecasts.

The recursive scheme uses increasingly more data to generate successive forecasts. If the data-generating process does not change over time this may be a preferred scheme as it allows us to estimate the model parameters more precisely.

But if the data-generating process changes over time, then the use of all available data may as well harm our forecasts—they will be biased toward the historical past that is no longer relevant. In such instances, the rolling scheme may be a preferred alternative.



Obtaining forecast errors

Suppose we have a time series of a variable y indexed by t , that is, $\{y_t : t = 1, 2, \dots, T\}$, where T is the length of the time series.

To set up a pseudo-forecasting routine, we divide the sample into two parts, the in-sample set containing R observations, such that $R < T$, and the out-of-sample set with the remaining $T - R$ observations.

So, for one-step-ahead forecasts, this would result in a sequence of forecasts: $\{\hat{y}_{s+1|s} : s = R, \dots, T - 1\}$.

Having observed $\{y_{s+1} : s = R, \dots, T - 1\}$ we can, then, obtain forecast errors, $\hat{e}_{s+1|s} = y_{s+1} - \hat{y}_{s+1|s}$, for all $s = R, \dots, T - 1$.



Measures of forecast accuracy

The most commonly applied accuracy measures are the mean absolute forecast error (MAFE) and the root mean squared forecast error (RMSFE):

$$\text{MAFE} = \frac{1}{T - h - R + 1} \sum_{s=R}^{T-h} |\hat{e}_{s+h|s}|$$

$$\text{RMSFE} = \sqrt{\frac{1}{T - h - R + 1} \sum_{s=R}^{T-h} \hat{e}_{s+h|s}^2}$$

where h is the forecast horizon.



Key features of the forecast accuracy measures

The MAPE assumes absolute loss; the RMSFE assumes quadratic loss.

The lower is the measure of the forecast accuracy, the better is the model or the method used in generating forecasts.

Unlike some of the better-known in-sample goodness-of-fit measures (e.g., R-squared), these accuracy measures, on their own, have little meaning. They gain the meaning when we compare two (or more) models or methods of forecasting.



Measures of forecast adequacy

MAFE and RMSFE (or other similar measures) are *relative* measures designed to compare forecasts. These are the mean values (or monotonic transformations thereof) of some loss functions.

There also exist *absolute* measures designed to evaluate each forecast individually, i.e., irrespective to any other forecasts.

These are measures of forecast *unbiasedness* and *efficiency*, together forming the measure of forecast *adequacy* (or *rationality*).



Forecast unbiasedness

A forecast is unbiased if the forecast error is zero, on average.

This condition mimics that of the residuals of a model. But the residuals are, by construction, guaranteed to be mean-zero. This is not the case with forecast errors.

We want forecast errors to be mean-zero. Otherwise, we probably have an issue with the model or the method that we chose for generating the forecasts.



Misspecified model can yield biased forecasts

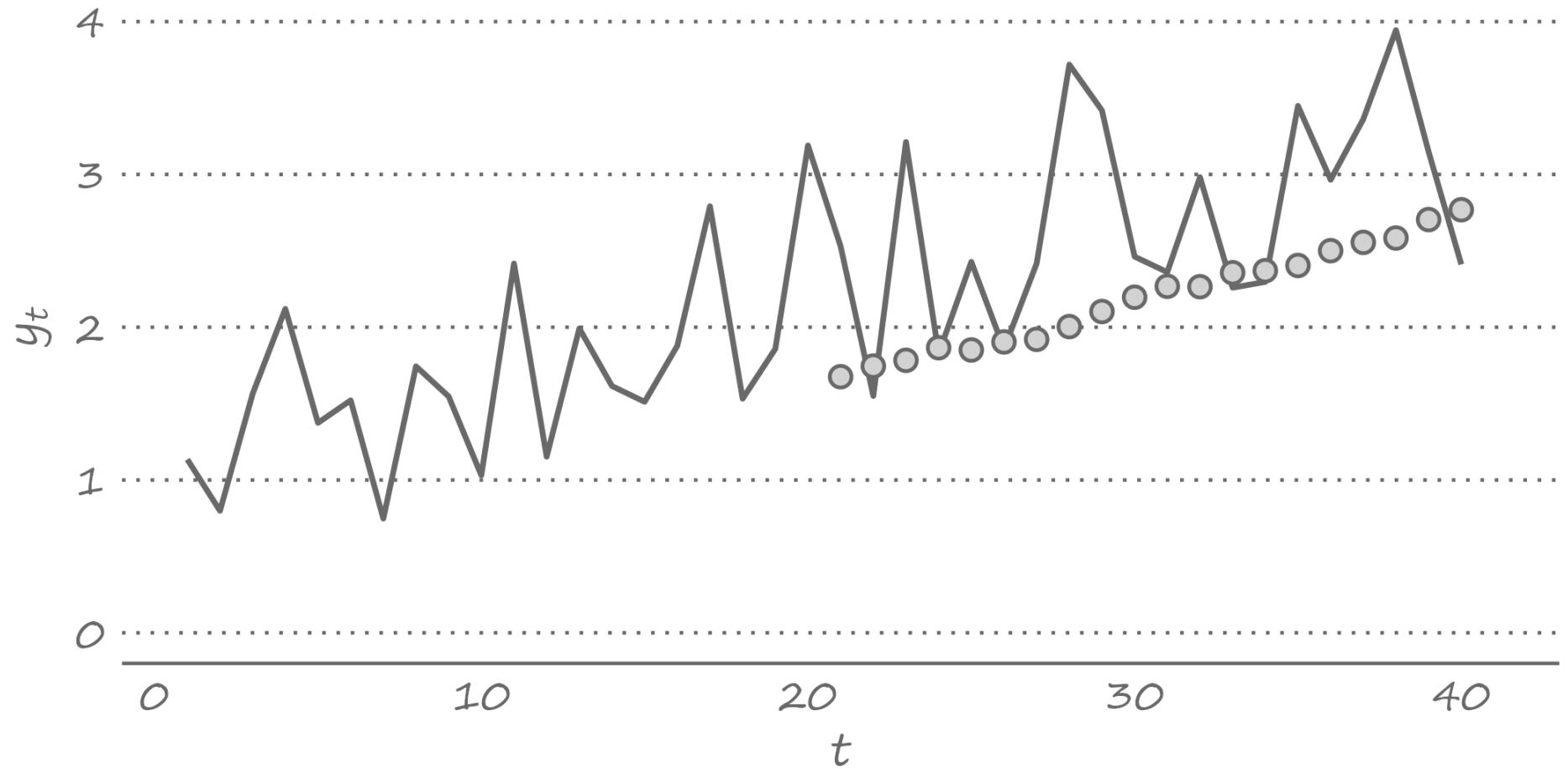
Consider a positively trending time series. And suppose, due to ignorance or incompetence, we did not account for the trend. Instead, we assumed that an appropriate model was the simple mean of the observed series.

The forecasts from this model will likely underestimate the realized future values of the time series. That is, the forecasts will be biased.

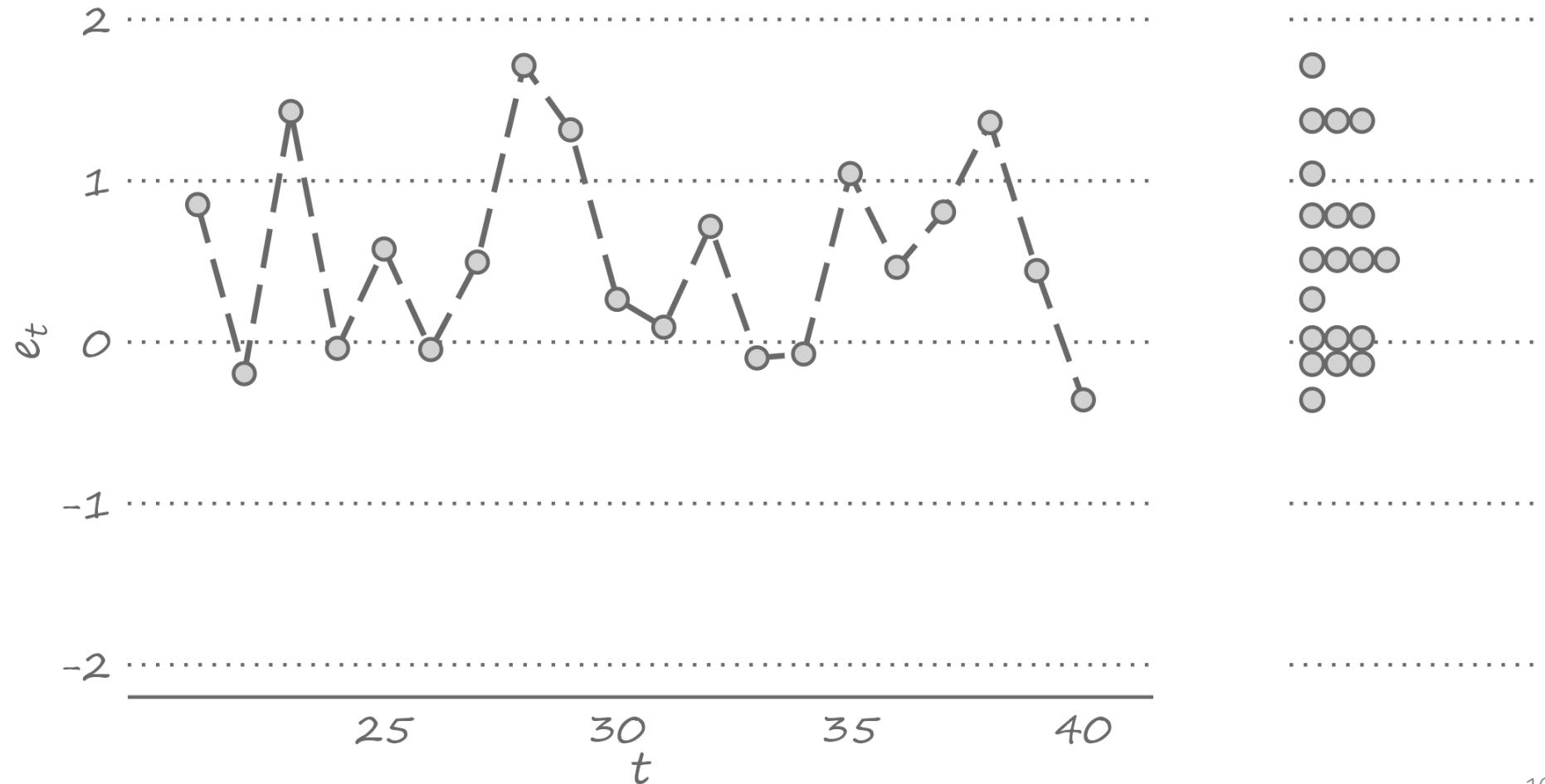
The following graphs illustrate this (using the rolling windows approach).



Biased forecasts



Biased forecast errors



Biased forecast errors

It is obvious the mean of these forecast errors is some positive value, which is considerably different from zero.

Of course, we are well aware of the issue - in generating the forecasts, we have ignored the presence of a trend in the time series.



Biased forecast usually are less accurate

The bias, in this instance, contributes to the lack of accuracy of the applied model. The RMSFE, based on these 20 forecast errors, is **0.80**.

Had there been no bias, the RMSFE, based on the bias-corrected 20 forecast errors, would have been **0.59**.

Biased forecasts usually (albeit not necessarily) are less accurate than their unbiased counterparts.



Forecast efficiency

A forecast is efficient if it incorporates all the relevant information that is available at the time when the forecast is made.

This is akin to the concept of efficient markets. When a market is efficient, the available information is not helpful in predicting a change in the price of a stock, for example. Likewise, when a forecast is efficient, the forecast should not be of use in explaining the forecast error.



Misspecified model can yield forecast inefficiency

Consider a series that follows a mean-reverting process. That is, any time the series deviate from the long-run mean, some forces pull the series back toward this long-run mean.

Think of a market equilibrium price, for example. When, for some random reason, the price increases, producers follow by increasing their supply, and consumers by reducing their demand - both forces pulling the realized prices toward the equilibrium, that is, the long-run mean.



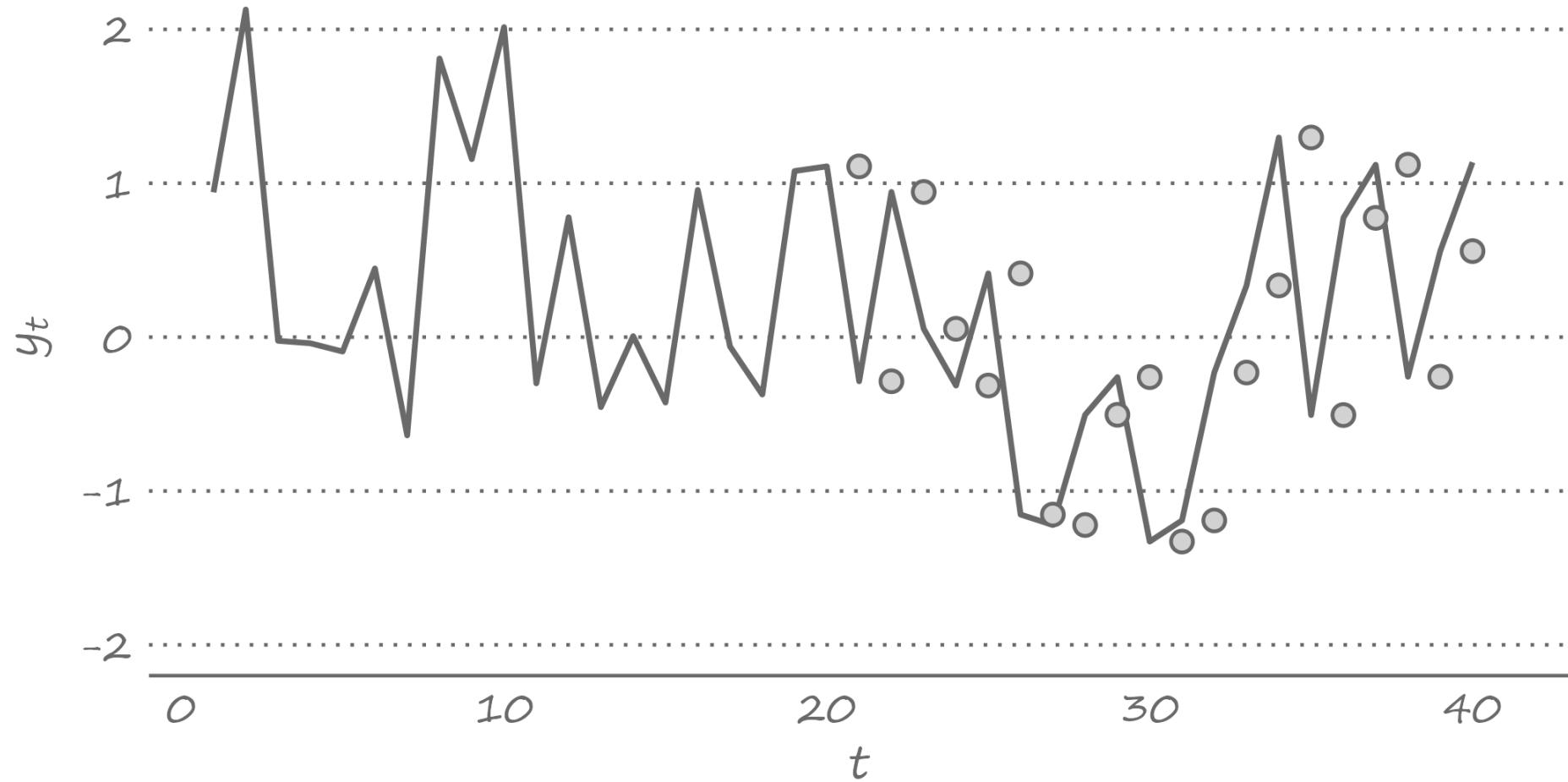
Misspecified model can yield forecast inefficiency

Suppose we neglected the presence of a mean-reverting process, and assumed that the random walk best characterizes the time series dynamics. That is, we assumed that our best guess about the next realization of the variable is its most recent observed realization.

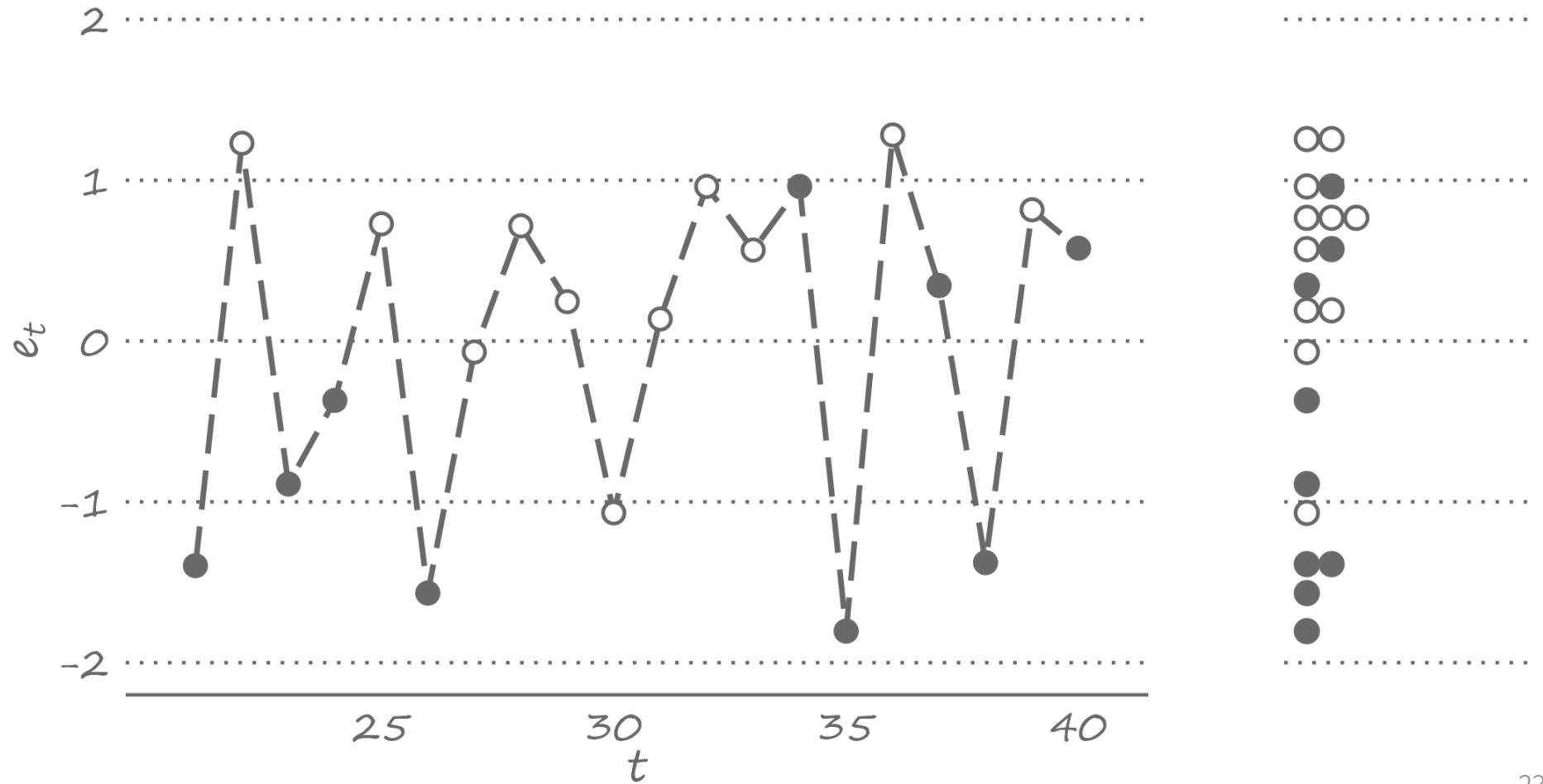
This assumption will result in forecasts that overestimate the realizations in the (extreme) positive range of the time series, and underestimate them in the (extreme) negative range of the time series.



Inefficient forecasts



Inefficient forecast errors



Inefficient forecast errors

As it were the case with the biased forecasts, the presence of inefficient forecasts also points to our lack understanding of the model used to generate forecasts.

We are, of course, well-aware of this deficiency in our modeling effort - we completely ignored that the forecast in a given period depends on whether the most recent observed realization is above or below its long-run mean.



Inefficient forecast usually are less accurate

The inefficiency contributes to the lack of accuracy.

The RMSFE, based on the forecasts from the misspecified model, is **1.25**.

The RMSFE, based on the forecasts from the correctly specified model, would have been **1.11**.

Inefficient forecasts are less accurate than their efficient counterparts.



Mincer-Zarnowitz regression

The unbiasedness and efficiency are hypotheses that can be tested.

We can do this in a basic regression setting:

$$y_{t+h} = \alpha_0 + \alpha_1 \hat{y}_{t+h|t} + v_{t+h}, \quad t = R, \dots, T-h,$$

where h is the length of the forecast horizon, and $R + h, \dots, T$ is the out-of-sample segment over which the forecasts are generated and evaluated; v_{t+h} is an error term (of this regression model) that is assumed to be independent and identically distributed random variable with zero mean and constant variance.



Mincer-Zarnowitz regression

The null hypothesis $H_0 : (\alpha_0 = 0, \alpha_1 = 1)$ is that for the joint test of unbiasedness and efficiency. Under this null hypothesis, v_{t+h} represents the forecast error.

More generally, the forecast error is:

$$\hat{e}_{t+h|t} = \alpha_0 + \beta_1 \hat{y}_{t+h|t} + v_{t+h},$$

where $\beta_1 = (\alpha_1 - 1)$.

When $\alpha_0 = \beta_1 = 0$, the forecast error is mean-zero and uncorrelated with the forecast.



Mincer-Zarnowitz regression

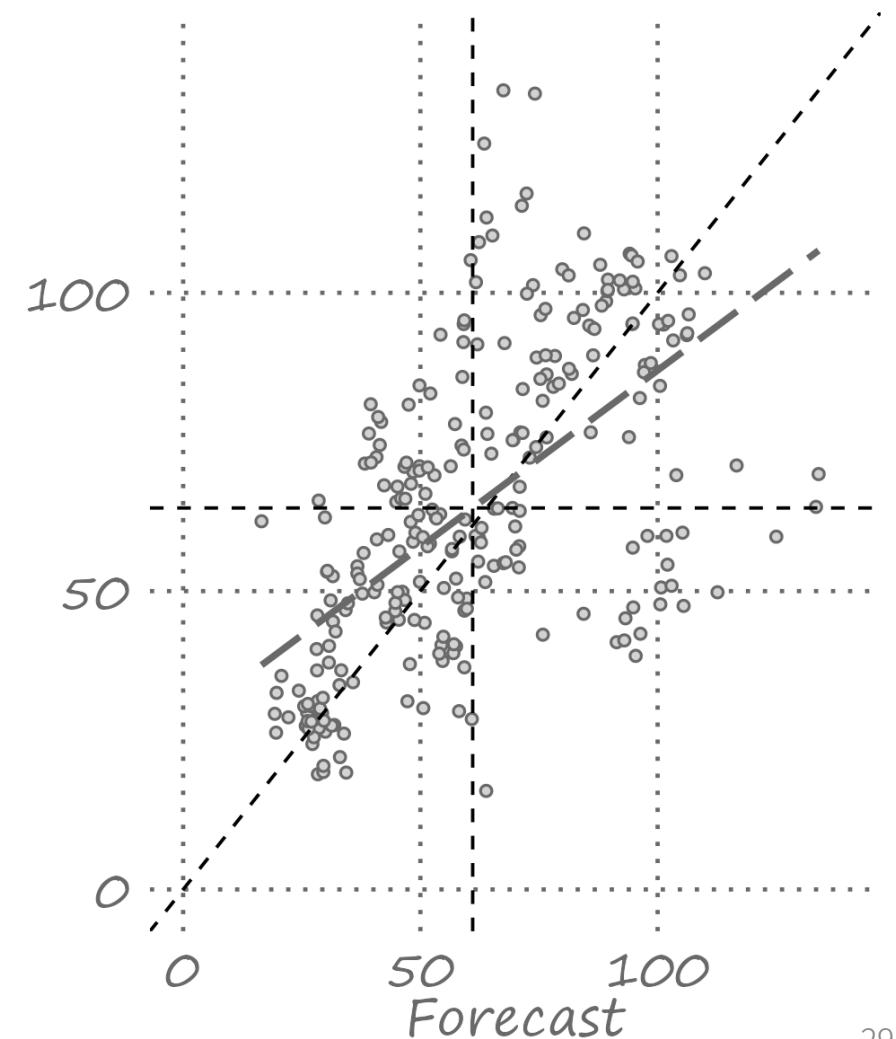
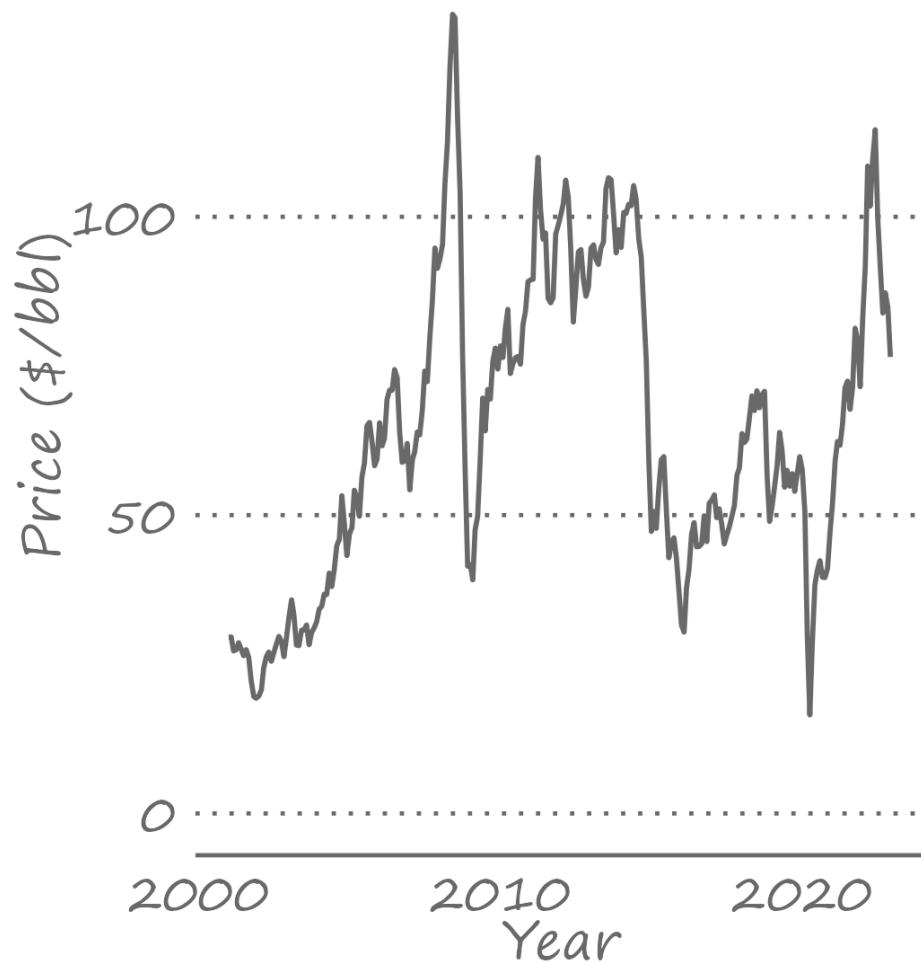
Suppose we are interested in forecasting monthly crude oil prices. We decided, for some arbitrary reason, to use the 12-month lag as a forecast for the month price.

Using a pseudo-forecasting routine, we can generate a set of out-of-sample forecasts for all but the first twelve periods of the available data.

The left panel of the following graph presents this series.



Crude oil prices and Mincer-Zarnowitz regression



Test of unbiasedness

We test unbiasedness of the one-step-ahead forecast by setting up a regression:

$$\hat{e}_{t+1|t} = \beta_0 + v_{t+1},$$

where, in this instance, $e_{t+1} = y_{t+1} - y_{t-11}$. The test of unbiasedness is equivalent to that of $H_0 : \beta_0 = 0$.

As it turns out, the test statistic is **0.895**, and thus we fail to reject the null hypothesis of unbiasedness.



Test of efficiency

We test efficiency of the one-step-ahead forecast by testing the null hypothesis of $H_0 : \beta_1 = 0$. The test statistic is **-2.633**, which, in absolute terms, is greater than the critical value at 1% significance level, and thus means that we reject the null hypothesis of efficiency.



Joint test of unbiasedness and efficiency

We can test forecast adequacy, which is a simultaneous test of unbiasedness and efficiency, by testing the null hypothesis of $H_0 : \alpha_0 = 0, \beta_1 = 0$.

The F-statistic associated with this null hypothesis is **3.824**, which is greater than the critical value of **3.00**. Thus we reject the null hypothesis of forecast adequacy.

The lack of forecast adequacy is primarily driven by the lack of forecast efficiency, as evidenced by the respective test.



Mincer-Zarnowitz regression

In the Mincer-Zarnowitz world, adequacy implies that

1. forecasts are, on average, equal to the realized observations, so that the vertical and horizontal lines intersect with the 45° line through the origin, and
2. there is, on average, a one-to-one relationship between the realized observations and the forecasts, resulting in the 45° slope of the regression line.

Moreover, (2) given (1) would result in the regression line overlapping with the line of perfect forecasts.



Key takeaways



- We can evaluate forecasts by using the first portion of the available time series to generate the forecasts and then comparing them to the second portion of the time series.
- We examine the adequacy of the forecasts by checking whether the forecast errors are, on average, zero and the forecast errors are uncorrelated with the forecasts.
- We assess relative accuracy of the forecasts by comparing their measures of forecast accuracy.

