Economics Bulletin

Volume 43, Issue 1

A bootstrap test of the time-varying efficiency of German growth forecasts

Christian Pierdzioch
Helmut-Schmidt-University Hamburg

Abstract

I use a bootstrap approach to re-examine the time-varying efficiency of growth forecasts for Germany. I argue that, given this small sample of forecasts, the bootstrap approach renders it possible to trace out with more precision than a standard full-sample forecast-efficiency-regression model whether forecasts were efficient at any given point in time. As an empirical application of the bootstrap approach, I present results for six-months-ahead and one-year-ahead growth forecasts published by three German economic research institutes during the sample period 1970$-$2018. The results illustrate that the bootstrap approach, for various configurations of the forecast-efficiency-regression model, yields stronger evidence against forecast efficiency than a conventional full-sample model.

I thank two anonymous reviewers for helpful and constructive comments. The usual disclaimer applies. This research received financial support from the Deutsche Forschungsgemeinschaft (Project: Exploring the experience-expectation nexus in macroeconomic forecasting using computational text analysis and machine learning; Project number: 275693836).

Citation: Christian Pierdzioch, (2023) "A bootstrap test of the time-varying efficiency of German growth forecasts", *Economics Bulletin*, Volume 43, Issue 1, pages 679-687

Contact: Christian Pierdzioch - macroeconomics@hsu-hh.de.

Submitted: October 20, 2022. **Published:** March 30, 2023.

1 Introduction

A common way to test for forecast efficiency is to estimate by the ordinary-least-squares (OLS) technique the following variant of a standard Mincer and Zarnowitz (1969) regression model (Holden and Peel 1990):

$$FE_{t+h|t} = \beta X_t + u_{t+h}, \qquad t = 1, ...T \tag{1}$$

where $FE_{t+h|t}$ denotes the forecast errors, defined by the difference between the forecast of a variable made in period t and the actual value in period $t+h$, X_t denotes some conditioning information that represents a forecaster's information set at the time a forecast is published, u_{t+h} denotes a disturbance term, and β denotes a coefficient (or a coefficient vector) to be estimated. Forecast efficiency with respect to the conditioning information requires that $\beta = 0$. It is common practice among researchers to differentiate between weak and strong forecast efficiency. Weak forecast efficiency requires that the lagged forecast error does not have predictive value for the subsequent forecast error. Strong forecast efficiency requires that any information available to a forecaster at the time a forecast was formed does not help to predict the subsequent forecast error.

Equation (1) sheds light on the question whether forecasts were efficient in the sample period, $t = 1, ...T$, that a researcher studies. A related but somewhat different question is whether forecast efficiency changed during the sample period being studied. Forecast efficiency could change over time because a forecaster accumulated human capital (that is, experience), fashionable economic theories changed, or the sample period comprises periods of a deep recession, a severe financial crisis, or an important regime shift in economic policy. For example, Heilemann and Stekler (2013) report that German growth forecasts became more accurate in the 1980s/1990s as compared to the 1970s, but became less accurate again thereafter. Döpke et al. (2019) report evidence that the Great Recession of 2008/2009 led to a change in forecasters' loss function that reflects, in the case of German growth forecasts, a stronger incentive to underestimate growth.

A natural approach to shed light on the potential time variation in forecast efficiency is to estimate Equation (1) using the full sample of forecasts, and then to plug into the estimated model the time-varying conditioning information. Such an approach, however, typically will yield comparatively wide confidence intervals in a small sample of forecasts, especially so when forecast efficiency changed over time. Adding a panel dimension to the data may provide more precise estimates, provided the cross-sectional dimension is sufficiently large. Another approach is to estimate a time-varying parameter model, or a rolling-estimation window. Such relatively sophisticated approaches, however, are often not feasible when it comes to the study of business-cycle forecasts because the number of forecasts is small.

The German growth forecasts that I study in this research note are no exception in this regard. For this reason, I study the efficiency of these forecasts by means of a simple bootstrap approach that gives a fairly accurate account of the potential time variation of forecast efficiency even though the sample of growth forecasts that I study is small.

2 Method

I implement the bootstrap approach in three steps (for a textbook introduction to bootstrap methods, see Efron and Tibshirani 1994).

1. I sample a proportion of x% training data without replacement from the data. I estimate Equation (1) on the sampled training data.

2. I use the estimation results to predict the $(1-x)$% hold-out (test) data and store the predictions. Because I sample in the first step without replacement, the fixed proportion of hold-out data can be interpreted as quasi "out-of-sample" data.

3. I repeat the first two steps 1,000 times (increasing the number of repetitions to 10,000 leads to qualitatively similar results).

This three-step bootstrap approach gives me, for every period of time, $t = 1, ...T$, a large number of predictions of every $FE_{t+h|t}$ in the small sample of forecasts that I study. In this regard, it should be noted that these predictions are made based on the hold-out data, which captures the idea that "new" information on a predictor should not have predictive value in case forecasts are efficient. The bootstrap approach, hence, sheds light on the question whether it would have been possible to predict the local forecast error conditional on the forecast-efficiency model estimated on the boostrapped data and conditional on the hold-out data. This is an important difference to the standard approach, which uses the full sample of data to make pure in-sample predictions of forecast errors.

I use the sampling distribution of the predicted forecast errors generated by applying the bootstrap to compute (i) the mean prediction of $FE_{t+h|t}$ for every $t = 1, ...T$, and, (ii) the 90% and 95% confidence intervals of the predicted $FE_{t+h|t}$ for every $t = 1, ...T$. I call the mean prediction of $FE_{t+h|t}$, $t = 1, ...T$, the local prediction of forecast efficiency (LPFE), and I reject local forecast efficiency when the bootstrapped confidence intervals do not include zero. I set $x = 75\%$ in a baseline scenario, but I also study how varying x% affects my results.

I use the R language and environment for statistical computing (R Core Team 2022) to code up the bootstrap approach.

3 Data and Results

I study the pooled six-months-ahead (published mid-year) annual forecasts and one-year-ahead (published at the turn of a year) annual forecasts of GDP growth published by three leading German economic research institutes between 1970 and 2018.[1] Studying the efficiency of this sample

[1] Behrens et al. (2019) refer to these forecasts as "$q2$" forecasts and "$q4$" forecasts. The economic research institute are (in alphabetical order): German Institute for Economic Research (DIW Berlin), ifo Institute and Kiel Institute for the World Economy. It should also be noted that the German statistical office has used GDP as the lead indicator for economic growth to GDP since 1992, and GNP before. The data I study in this research consist of GDP forecasts and the corresponding realizations throughout the entire sample period, so that this change in official statistical concepts does not affect my results.

Table 1: Summary statistics

Forecasts	Institute 1	Institute 2	Institute 3	Pool
		Number of forecasts		
six-months-ahead	37	42	37	116
one-year-ahead	47	44	41	132
		Mean-squared-prediction error		
six-months-ahead	0.81	0.51	0.76	0.68
one-year-ahead	2.11	1.57	1.15	1.63

of German growth forecasts is interesting because various facets of these forecasts have been extensively studied in a row of recent papers (e.g., Behrens et al. 2019, Foltas and Pierdzioch 2022). The forecast error is defined as the difference between a forecast and the realized (first-release) growth rate. The data account for German reunification. In total, I can study (after accounting for lagged data) 116 six-months-ahead forecasts and 132 one-year-ahead forecasts. The frequency of forecasts differs across the institutes and is not constant over time.

I report, for every research institute and the pooled data, the number of forecasts and the full-sample mean-squared prediction error in Table 1. The mean-squared prediction error exhibits a noticeable degree of heterogeneity across the three economic research institutes (which, however, is not the focus of this research note).[2]

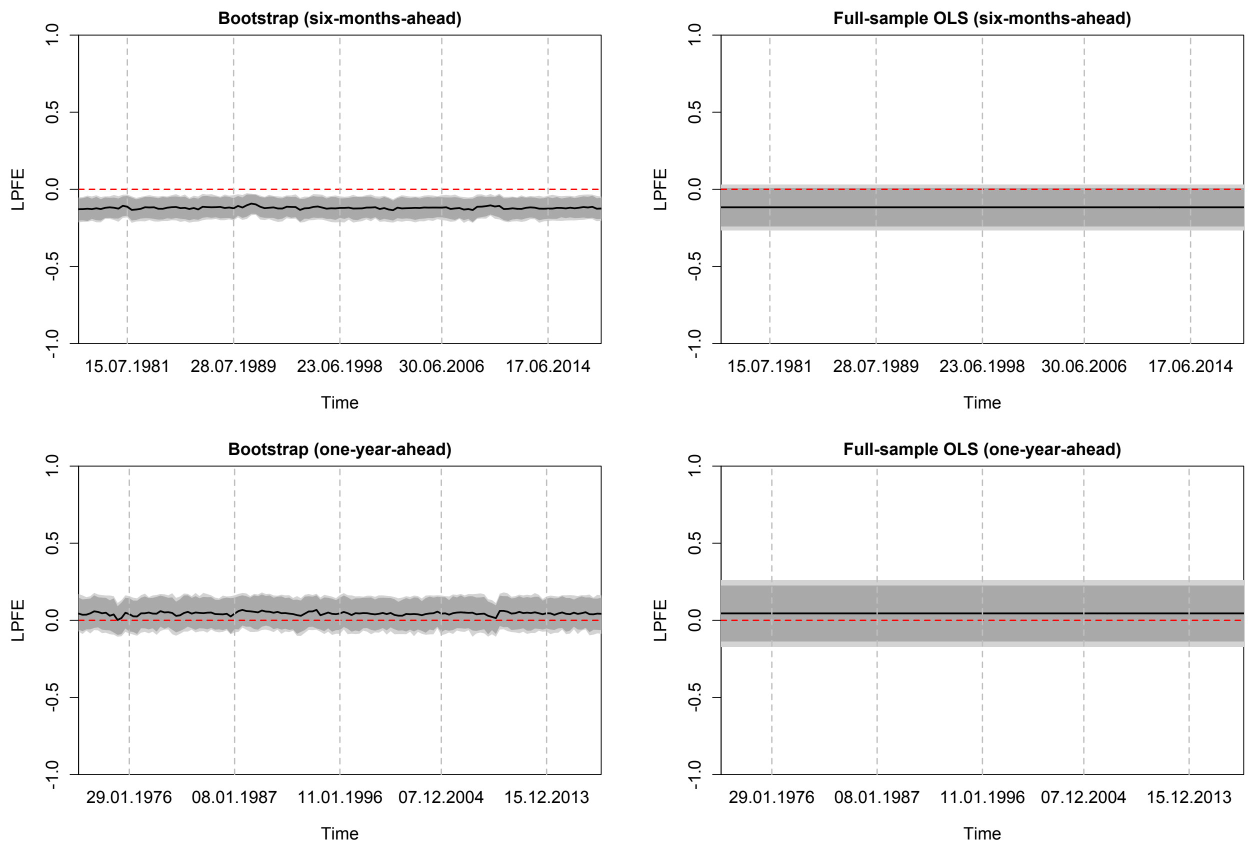
As for the conditioning information, I consider three cases. First, I regress the forecast error on a constant. In this case, forecast efficiency is equivalent to the unbiasedness of forecasts, and it requires that the constant is not significantly different from zero. Second, I regress the forecast error on a constant and the (institute-specific) lagged forecast error. This second case, hence, amounts to a test of weak forecast efficiency. Third, I regress the forecast error on a constant and the short-term interest rate (the money market rate; I assume a forecast formation/publication lag of one month). Using the short-term interest rate as conditioning information has the advantage that information on the short-term interest rate is readily available to a forecaster, and it is not subject to data revisions. The third case is a test of strong forecast efficiency.

I plot in Figure 1 the results for the first case. I plot in the left panels the bootstrap results, and in the right panels for comparison purpose the full-sample OLS results. The upper row shows the results for the six-months-ahead forecasts, and the lower row shows the results for the one-year-ahead forecasts. The results show that the average local predicted six-months-ahead (one-year-ahead) forecast error (LPFE) is negative (positive). The full-sample OLS confidence intervals are wider than the bootstrapped confidence intervals.[3] While I cannot reject unbiasedness of forecasts based on the full-sample OLS estimates, the bootstrap results show that I cannot reject unbiasedness only for the one-year-ahead forecasts. The bootstrap results imply that the extent of (un-)biasedness of the forecasts was stable over time.

[2]Döhrn and Schmidt (2011) report results for a broader panel of economic research institutes and international organizations, and for GDP and its components. They study the sample period 1991−2008 and emphasize the importance of the length of the forecast horizon for absolute forecast accuracy, while institutional factors appear to play a minor role for absolute forecast accuracy.

[3]Because the estimates are based on of a full-sample regression of the forecast error on a constant only, the estimates (and the confidence intervals) depicted in the right-hand-side panels of Figure 1 are time invariant.

Figure 1: A constant is the conditioning variable



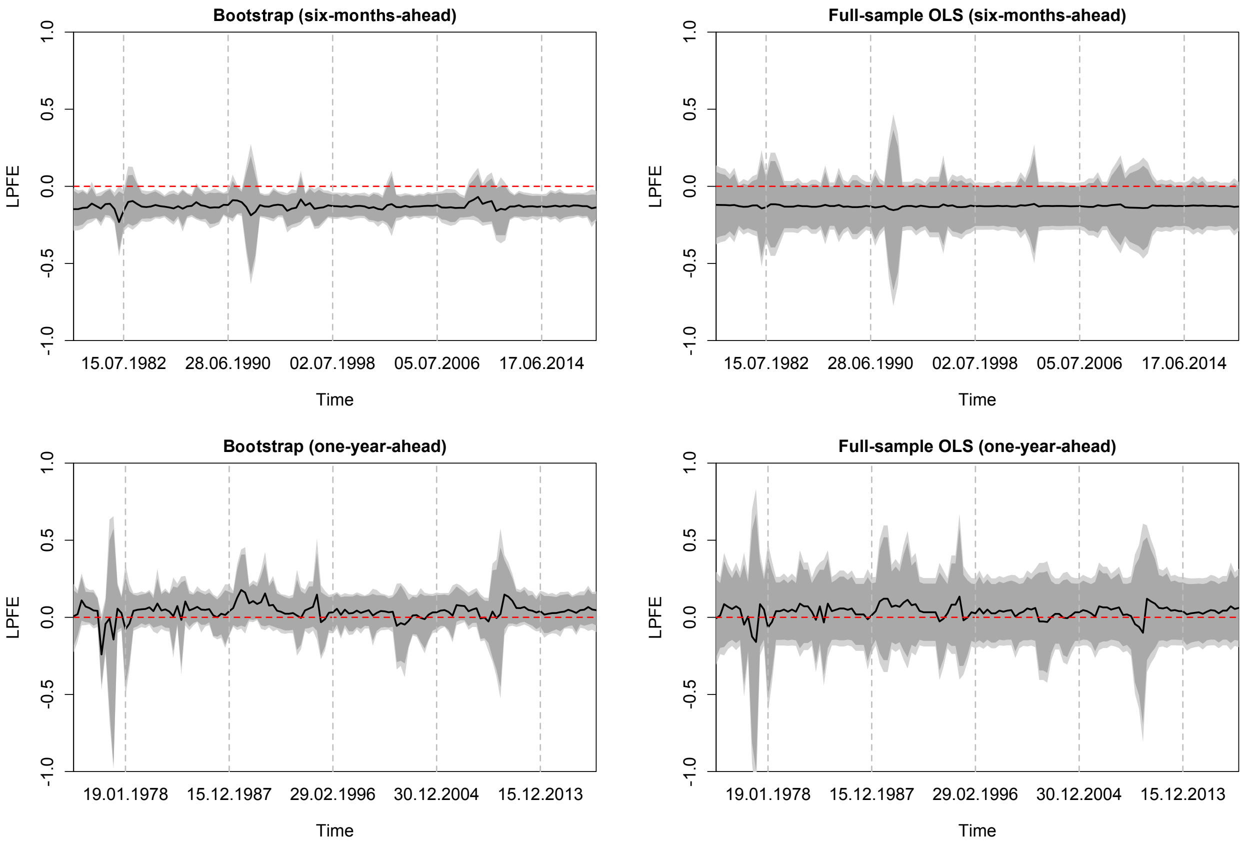
Bootstrap: The solid line represents the local predicted forecast error (LPFE). OLS: The solid line represents the predicted forecast error. The gray (dark gray) areas represent the 95% (90%) confidence intervals.

I plot in Figure 2 the results for the second case. In this case, I test for weak forecast efficiency. The full-sample OLS confidence intervals are wider than the bootstrap confidence intervals and imply that I cannot reject weak forecast efficiency for both the six-months-ahead and the one-year-ahead forecasts. In contrast, I find evidence against weak forecast efficiency for the six-months-ahead forecasts when I consider the boostrap results. Taken together, the bootstrap results imply that the six-months-ahead forecasts are neither unbiased nor weakly efficient.

I plot in Figure 3 the results of a test of strong forecast efficiency.[4] The test results imply that I cannot reject strong forecast efficiency based on the full-sample OLS estimates, while the bootstrap results yield signs of time-varying strong forecast efficiency. The six-months-ahead forecasts were inefficient most of the time in a strong sense until around 2007, where the local predicted forecast error was consistently negative. The local prediction forecast error became smaller (in absolute size) thereafter and forecasts no longer violated strong forecast efficiency. Similarly, the bootstrap results provide evidence of time-varying strong efficiency of the one-year-ahead forecasts. The local predicted forecast error was mainly positive before 1996 for the one-year-ahead forecasts, and thereafter started turning negative. As a result, the bootstrap results imply that I cannot reject

[4]The so-called zero-lower-bound period does not affect the test results (results are not reported for reasons of space).

Figure 2: The lagged forecast error is the conditioning variable

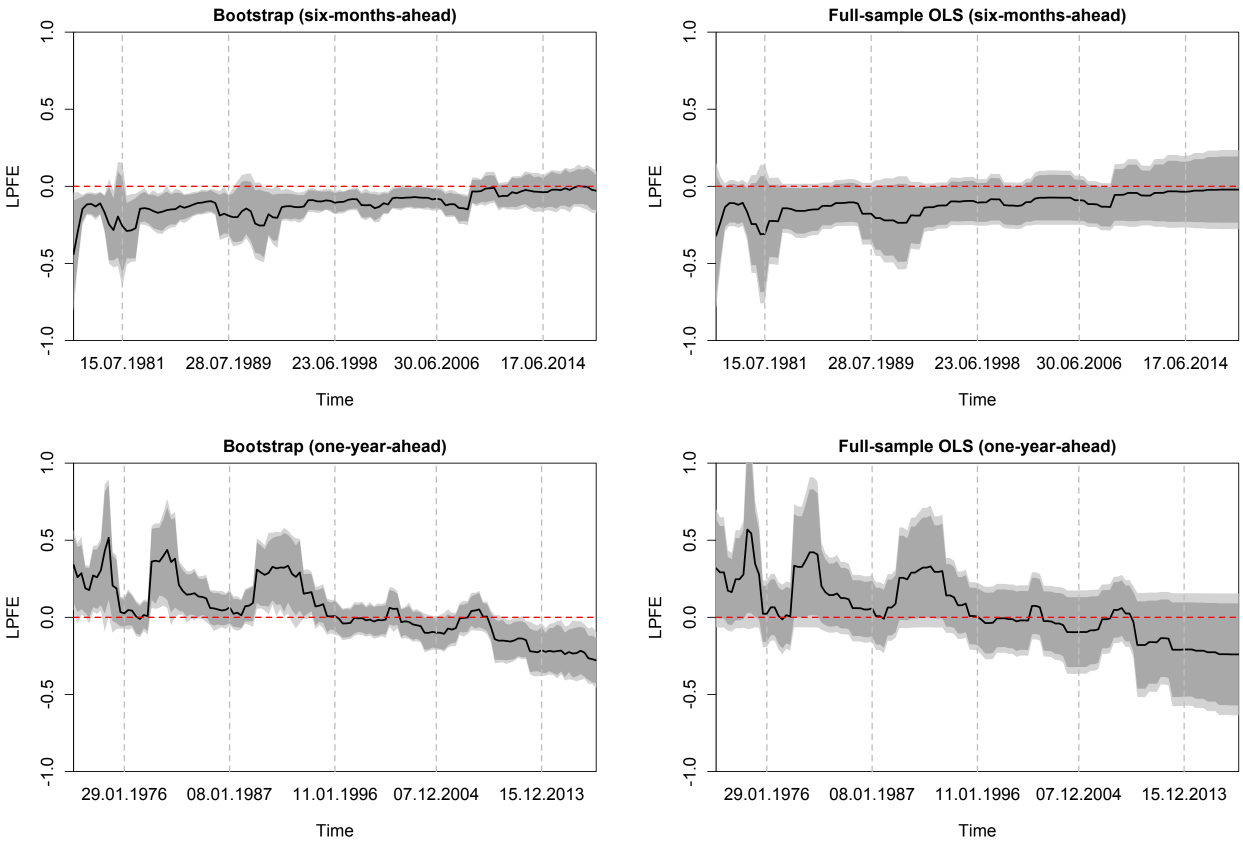


Bootstrap: The solid line represents the local predicted forecast error (LPFE). OLS: The solid line represents the predicted forecast error. The gray (dark gray) areas represent the 95% (90%) confidence intervals.

strong forecast efficiency only during two intervals of time around 1976 and 1987, and between 1996 and 2009. Hence, after the Great Financial Crisis and the Great Recession the one-year-ahead forecasts became inefficient in a strong sense, but this inefficiency differed from the inefficiency that prevailed earlier in the sample because the local predicted forecast error was negative rather than positive.

Besides the number of simulation runs, the proportion of bootstrapped data is the only hyper-parameter of the bootstrap approach. It, therefore, is important to inspect how a variation in the proportion of bootstrapped data affects the results of the forecast-efficiency tests. Decreasing the proportion of bootstrapped data used to estimate the forecast-efficiency-regression model implies that the dispersion of the local predictions of the forecast errors made with an increasing proportion of hold-out data is getting larger. Hence, intuition suggests that lowering the proportion of boot-strapped data widens the confidence interval and makes it harder to reject the null hypothesis of local forecast efficiency. The results that I plot in Figure 4 show that this intuition is correct. The figure plots the proportion of times that the bootstrap approach leads to a rejection of local forecast efficiency (that is, how often the confidence interval for the predicted local forecast error does not include zero) as a function of the proportion of bootstrapped data. I also compare the results with

Figure 3: The short-term interest rate is the conditioning variable



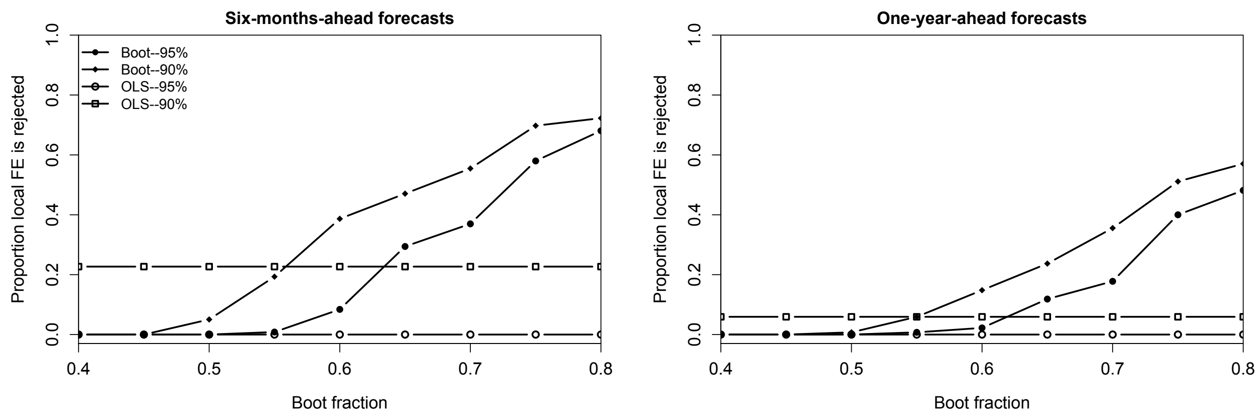
Bootstrap: The solid line represents the local predicted forecast error (LPFE). OLS: The solid line represents the predicted forecast error. The gray (dark gray) areas represent the 95% (90%) confidence intervals.

the proportion of times that the OLS approach leads to a rejection of local forecast efficiency (horizontal lines). In line with intuition, the bootstrap approach leads more often to a rejection of local forecast efficiency when the proportion of bootstrapped data increases (and, of course, when 90% confidence bands are being used). The bootstrap approach rejects local forecast efficiency more often than the OLS approach when the proportion of bootstrapped data is larger than roughly 55% and 65% (depending on whether I study the six-months-ahed or one-year-ahead forecasts). These are proportions commonly used in bootstrap simulations. When the number of forecasts available for an empirical analysis is somewhat larger than the small number of forecasts that I study in this research note, however, it is interesting to consider whether cross-validation techniques can be used to choose the proportion of bootstrapped data.

4 Concluding Remarks

A classic topic in the literature on business-cycle forecasts is whether such forecasts are efficient and whether forecast efficiency changed over time. While various techniques are available to study time-varying forecaster behavior and time-varying forecast efficiency, I have used to this end in

Figure 4: The influence of the proportion of bootstrapped data



The horizontal axis shows the proportion of bootstrapped data. The vertical axis shows the percentage of times local forecast efficiency (FE) is rejected (that is, how often a confidence interval for the LPFE does not include zero). The short-term interest rate is the conditioning variable

this research note a simple bootstrap approach to re-examine the forecast efficiency of a small sample of German growth forecasts. I have shown that the bootstrap approach, depending on the proportion of data being bootstrapped, yields narrower confidence intervals for the local predicted forecast error, leading in some cases to different conclusions regarding the time-varying (local) forecast efficiency than a standard full-sample OLS approach. Hence, my results show that it could be useful to add the bootstrap approach to the array of tools that empirical researchers routinely use to study forecast efficiency.

While I have applied the bootstrap approach to study the efficiency of a small sample of German growth forecasts, it is straightforward to apply the approach in future research to study the potentially time-varying efficiency (or other potentially changing properties) of business-cycle forecasts for other periods of time and other countries. Application of the bootstrap approach in such contexts will not necessarily lead to fundamentally different results as compared to the results researchers obtain by applying other standard approaches to test for forecast efficiency, but the bootstrap approach may help to sharpen evidence of forecast efficiency (or lack thereof) and, thus, deepen our understanding of how forecast efficiency and the process of business-cycle forecasting have evolved over time.

References

Behrens, C., Pierdzioch, C., and M. Risse (2019) “Do German Economic Research Institutes Publish Efficient growth and Inflation Forecasts? A Bayesian analysis” *Journal of Applied Statistics* **47**, 698–723.

Döhrn,R., and Schmidt, C. M. (2022) “Information or Institution? On the Determinants of Forecast Accuracy” *Jahrbücher für Nationalökonomie und Statistik* **231**, 9–27.

Döpke, J., Fritsche, U., and Müller, K. (2019) “Has Macroeconomic Forecasting Changed After the Great Recession? – Panel-Based Evidence on Accuracy and Forecaster Behaviour from Germany" *Journal of Macroeconomics* **62**, 103–135.

Foltas, A., and Pierdzioch, C. (2022) "On the Efficiency of German Growth Forecasts: An Empirical Analysis Using Quantile Random Forests and Density Forecasts" *Applied Economics Letters* **29**, 1644–1653.

Efron, B., and Tibshirani, R.J. (1994) *An Introduction to the Bootstrap.* New York, Chapman and Hall.

Heilemann, U., and Stekler, H. O. (2013) "Has the Accuracy of Macroeconomic Forecasts for Germany Improved?" *German Economic Review* **14**, 235–253.

Holden, K., and Peel, D.A. (1990) "On Testing for Unbiasedness and Efficiency of forecasts" *The Manchester School* **58**, 120–127.

Mincer, J. A., and Zarnowitz, V. (1969) "The Evaluation of Economic forecasts" in *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance* by J. A. Mincer, Ed., National Bureau of Economic Research: New York, 3–46.

R Core Team (2022) "R: A Language and Environment for Statistical Computing". R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.