Quality of Tests of Expectation Formation for Revised Data

Paulina Ziembińska*

Abstract

The work contains discussions and simulation analyses of the expectation formation processes, taking account of the data revisions. In particular, it contains results of simulations examining statistical properties of the rationality tests and extrapolation processes, with particular focus on their behaviour in the case of short samples and data with measurement errors. The conclusions indicate that the rationality test based on the optimal regression and the proposed adaptive and accelerating tests are the most efficient and flexible. The tests showcasing best properties have been applied to a new set of macroeconomic forecasts for Poland. The results show that there are no grounds for rejecting the hypothesis on the rationality of forecasts derived from the National Bank of Poland (NBP) and the Organisation for Economic Cooperation and Development; however, this property was rejected for the European Commission. What is more, the comparative analysis indicates that only the national institution (NBP) may potentially aim the final readings of the macroeconomic data as the forecasting target. Finally, it transpires that the extrapolative models, albeit simple and intuitively interpreted, generally fail to correctly explain the forecast formation processes regarding the Polish economy.

Keywords: data revisions, macroeconomic forecasts, Polish economy, rational expectations, expectation processes

JEL Classification: C53, C82, E17, E27, E37

*University of Warsaw, Faculty of Economic Sciences, Warsaw, Poland; e-mail: ziembinskapaulina@gmail.com; ORCID: 0000-0002-8394-4909

Paulina Ziembińska

1   Introduction

One of the key questions in the context of expectations in the theory of economy is the process of forming them by various agents. Many economic models assume the impact of expectations on the ongoing decisions. However, accounting for the expectations within the models requires making assumptions concerning relevant expectation formation processes that should be appropriately calibrated to the empirical data. The main revolution in the theory of expectations took place following the publication of an article by Muth (1961), who defined the term of rational expectations. It forms a basis for the neoclassical economy and for the resultant Lucas critique that puts the impact of agent's expectations on their decisions at the centre of interest. A broader discussion of the place of rational expectations in the history of economy and their evolution can be found in Coibion et al. (2018). Formally, the Muth's hypothesis of rational expectations rests on an assumption that agents build expectations that are coherent with the actual economic processes and make use of the entire set of information available. Their critique mainly concerns the lack of a realistic assumptions concerning the availability of information as well as possibilities (an assumption of knowing and understanding of the entire economic system) and economic justification (omitting costs) of information processing. Abundant models repealing these assumptions have been proposed, especially the ones based on information rigidities: sticky information, noisy information, or rational inattention models (they are more thoroughly discussed in, among others, Coibion and Gorodnichenko, 2015). Extrapolative expectations that assume a linear scheme of learning based on the past data and errors provide a simplified framework. Notably, the model of adaptive expectations based on the error correction mechanism has become a subject of numerous analyses and modifications summarised in Evans and Honkapohja (2012). In particular, the extrapolative models are not informationally efficient, and their errors can be serially correlated – for discussion see e.g. Pesaran and Weale (2006) or Pesaran (1985).

Testability of the expectation models depends on the existence of relevant data to which they could be compared. Specifically, the expectations are non-observable, but the surveys offer some form of approximation. A discussion concerning the observability of expectations and the quality of survey data can be found, for example, in Tomczyk (2011). The literature offers abundant analyses of survey expectations from households, enterprises, and professional forecasters. The recent overview of analyses based on survey expectations can be found in Clements (2019). The most easily available and comprehensive datasets are often maintained by the global central banks. Specifically in Poland, National Bank of Poland runs several surveys among consumers, enterprises and professional forecasters, which provide predictions of the main economic variables. Series of articles by Łyziak and Stanisławska concern detailed analysis of these data, especially focused on the inflation expectations, overview of which can be found in their recent work Łyziak and Sheng (2018) or Baranowski et al. (2021).

Testing the expectation formation models requires not only some data on the expectations themselves, but also on the past readings of forecasted variables and a more broadly perceived state of the economy. Defining the entire set of information available is not trivial though, since most macroeconomic variables are revised largely and multiple times. Almost until the beginning of the 21st century data revisions were mostly ignored - econometricians assumed that they are small and random, and therefore do not affect modelling, inference or forecasting (Croushore, 2011). Only since the release of the first real-time data for the United States by the Federal Reserve Bank of Philadelphia, many studies have been published falsifying the previously adopted simplifying assumptions. Specifically, data revisions are important in several aspects of forecasting. Firstly, they directly affect the forecasting process - data selection, model fit, possibly also expert adjustment. Secondly, the revisions may impact discrepancies observed among forecasts from various sources and increase their perceived uncertainty. Thirdly, for revised data it is unclear which reading to use to define the forecast error, and therefore revisions affect the entire forecast quality evaluation and conclusions about the expectation formation process in general, which is of specific interest in this paper. Recent overview of the related literature can be found in Clements and Galvão (2019) and detailed analysis of the Polish data revisions in Ziembińska (2017) and Ziembińska (2021).

The extrapolative expectation formation processes present a simplified learning model and are dependent only on the past readings and expectations. Therefore, the set of information de facto contains only a variable subject to an analysis, which allows quite straightforward analysis of the impact of revisions on this process. This is one of the objectives of this work. As for the rationality models - they are well-documented in the empirical literature, but their small-sample properties have not been well scrutinised, especially in the context of data with measurement errors resulting from, for example, the abovementioned revisions. Patton and Timmermann (2012) have analysed the properties of rationality tests for a sample of 100 observations in a series of Monte Carlo experiments. Additionally, they took account of several structures of measurement errors found in the data. To the best of my knowledge, similar analyses regarding the extrapolative tests do not exist. In this work I would like to shed light on the understanding and statistical properties of expectation formation tests in the context of data with measurement errors. I am conducting a series of Monte Carlo experiments for different scenarios concerning the revision process and non-optimality of generated forecasts through extending the ones proposed by Patton and Timmermann (2012) by scenarios with non-zero mean measurement error and a shorter sample of 50 observations. When defining the forms of a measurement error analysed as part of the simulation, I am referring to the scale of revisions of the Polish macroeconomic data analysed in Ziembińska (2017). Next, I am testing the results on a set of forecasts of Polish macroeconomic variables prepared by forecasters from large research centres. I have analysed predictions of basic variables concerning national accounts, inflation, and unemployment, derived from the European Commission (EC),

Organisation for Economic Cooperation and Development (OECD) and the National Bank of Poland (NBP).

In the following sections I will discuss rational and extrapolative expectation models and methods of their testing; afterwards, I will present the Monte Carlo experiment studying the properties of the discussed tests. Finally, I will analyse whether the forecasts formulated for the Polish economy are rational and consistent with the extrapolative models. The results and the ensuing discussion aim at providing understanding of the impact of data revision on the considered phenomena.

2 Rational expectations

The existence and form of mathematical representation of the expectation formation processes depend on the understanding of uncertainty in the decision-making process, for example, on distinguishing between actual uncertainty and the risk taken. In the case of expectations formed as point forecasts against the mean square loss function, it is usually assumed that they represent an expected value of the subjective conditional distribution with respect to a set of public and private information available to a given agent. Such a mathematical representation takes account of uncertainty resulting from disagreements in information sets and subjective convictions. This model includes a restrictive version of rational expectations originally formulated by Muth (1961), which assumes that private information has no impact on the expectations and each agent is familiar with an actual "objective" distribution that describes a real model of the economy. A consequence of these assumptions is two properties of expectation errors: they have zero mean and are serially uncorrelated. A broad analysis of mathematical properties of rational expectations can be found, for example, in Pesaran and Weale (2006). However, the literature does not provide lots of discussion on the relation between rational expectations and the quality of data used to evaluate the condition of the economy. In line with the assumptions of the Muth's model and in the context of revised data, the process of forming rational expectations should also consider the revision process as an element of the real model of the economy. However, taking account of the revisions in the process of testing rational expectations is less unambiguous, for example as regards forecasts from survey data.

The most popular method of rationality testing is a regression proposed by Mincer (1969):

$$Y_t = a + b\hat{Y}_{t|t-h} + \epsilon_t \tag{1}$$

and a test based on it (hereinafter: the MZ Test):

$$H_0 : \begin{cases} a = 0 \\ b = 1 \\ \forall t_1 \neq t_2 : Cov(e_{t_1}, e_{t_2}) = 0, \end{cases} \tag{2}$$

where Y_t denotes a forecasted variable for the reference period t, $\hat{Y}_{t|t-h}$ is its forecast prepared in period $t - h$ and $e_t = Y_t - \hat{Y}_{t|t-h}$ denotes the error. I will apply the ϵ_t designation for regression residuals throughout my work. a and b denote model coefficients; similarly - I will apply initial alphabet letters to denote models' coefficients.

Another method for forecasts with numerous horizons was proposed by Patton and Timmermann (2012). They demonstrated that the condition of rationality, when assuming the mean square loss function, implies specific restrictions for the higher moments. Let us denote subsequent times when forecasts $\hat{Y}_{t|t-h_i}$ of the variable Y_t are formed as $1 \leq h_1 < ... < h_k \leq t$ and the corresponding errors as $e_{t|t-h_i}$. Patton and Timmermann (2012) prove that under condition of rationality and mean square loss function (i.e., when $\hat{Y}_{t|t-h_i} = E(Y_t|I_{t-h_i})$), the following conditions occur for the stationary process:

i) Test 1: $E(e_{t|t-(h+1)}^2) \leq E(e_{t|t-h}^2)$ - of the increasing mean square error (MSE);

ii) Test 2: $E(\hat{Y}_{t|t-(h+1)}Y_t) \geq E(\hat{Y}_{t|t-h}Y_t)$ - of the decreasing covariance;

iii) Test 3: $Var(\hat{Y}_{t|t-h} - \hat{Y}_{t|t-(h+1)}) \leq 2Cov(Y_t(\hat{Y}_{t|t-h} - \hat{Y}_{t|t-(h+1)}))$ - covariance restriction;

iv) Test 4: $E(\hat{Y}_{t|t-(h+1)}^2) \geq E(\hat{Y}_{t|t-h}^2)$ - of the decreasing mean squared forecast (MSF);

v) Test 5: $E((\hat{Y}_{t|t-(h+1)} - \hat{Y}_{t|1})^2) \leq E((\hat{Y}_{t|t-h} - \hat{Y}_{t|1})^2)$ - of the increasing mean square forecast revision (MSFR).

Based on these conditions, the authors developed a set of regression tests of multidimensional inequalities. Under the null hypothesis the testing statistics are a weighted average of the chi-square statistics (more details can be found in the original article). Finally, the authors also propose a unidimensional optimal test based on the following regression:

$$Y_t = a + b\hat{Y}_{t|1} + \sum_{i=1}^{k} c_i(\hat{Y}_{t|t-h} - \hat{Y}_{t|t-(h+1)}) + \epsilon_t. \tag{3}$$

The rationality hypothesis is then equivalent to the following conditions: $a = 0$, $b = 1, \forall i : c_i = 1$ (hereinafter: the PT Test). The last approach offers several advantages. First, in their Monte Carlo experiments the authors demonstrate high power and good size of this test. Second, which is significant in the context of the real-time data, the PT Test offers a possibility of replacing the realized value of a variable with its short-term forecast (assuming the optimality of forecasts). The authors point out that this enhances the small-sample properties of the test, especially if we consider the data with significant measurement errors or if the predictive power of the model

Paulina Ziembińska

is low. The lack of necessity to define the actual reading of the variable enables us to examine the rationality of forecasts independently of the revision process.

As regards forecasts with many horizons, I am also checking an alternative version of the MZ Test:

$$e_{t|s_{i+1}} = a + b(\hat{Y}_{t|t-h} - \hat{Y}_{t|t-(h+1)}) + \epsilon_t. \tag{4}$$

A standard rationality test is then equivalent to the hypothesis $H_0 : a = 0$, $b = 0, \forall t_1 \neq t_2 : Cov(\epsilon_{t_1}, \epsilon_{t_2}) = 0$ (Test CG). Coibion and Gorodnichenko (2015) proved that $b > 0$ leads to rejection of a rationality assumption. More complex tests proposed by the authors allow verifying numerous possible models lying behind the lack of rationality, for example of information rigidity, asymmetrical loss function or reputational smoothing of forecasts (cf. for example Bordalo et al., 2020).

In the case when the Y_t variable is subject to revisions, the presented rationality tests may bring about different conclusions, depending on which reading we use. Notably, it might be the revisions that constitute an element of the information set, the consideration of which decides on the optimality of a forecast. In rationality tests we use both the realization of a variable, which can be subject to revisions, and its forecasts, at the formation of which the available information sets may have significantly depended on the revision. Taking the data revision process into account may in general be based on a broader analysis of dependence of the forecast errors on the revision or, in a simplified version, on the comparative analysis of test results for different readings.

3   Extrapolative expectations

The concept of rational expectations is often rejected in empirical studies and criticised for its unrealistic assumptions. Therefore, other expectation formation models have been developed with different degrees of conservatism of assumptions concerning the information set. The class of extrapolative models defined by a linear function of past known readings has been analysed particularly frequently (cf. Pesaran and Weale, 2006):

$$\hat{Y}_{t|t-h}^{extrapolative} = a + \sum_{i=0}^{\infty} b_{i,s} Y_{t-h-i}. \tag{5}$$

The linear model is certainly a simplification of the actual expectation formation process, but it provides better understanding of the mechanism behind the learning process. If the data are subject to revision, the understanding of the extrapolative model becomes less intuitive due to the ambiguity of the Y_{t-h-i} variable. In practice, it is not always possible to maintain a coherent time series, which would be desirable from the theoretical point of view. Therefore revisions introduce inconsistencies in the theoretical models. In practice, the latest available data are used most frequently, which means that the degree to which the revision has been considered varies at different moments. Usually the largely revised data are present at the end of a

sample, whereas the first readings are found mainly at the beginning. The question of how a forecaster treats the revisions and whether they take account of their impact on the subsequent readings of a variable is of key importance for the extrapolative expectations. Let us note that the discussed approach is different than that of Pesaran (1985), where a measurement error is discussed in the context of expectations instead of data that take account of errors in the explained variable. In this section I will provide several examples of extrapolative models, the accompanying intuition, and possible methods of dealing with revisions on that basis. To the best of my knowledge, these models have not been discussed in the context of the real-time data (Arnold (2012) can be treated as an exception). A more detailed overview of the literature can be found in Pesaran (1985). Pesaran and Weale (2006) also introduced some combined models, e.g., an adaptive-regressive (supplementing the adaptive model with autoregressive component), or the adaptive-acceleration.

List of extrapolative models:

i) M1: a naïve model adopts the latest reading available as a prediction of the next one:
$$\hat{Y}_{t|t-h} = Y_{t-h}. \tag{6}$$

ii) M2: a first-degree extrapolative model assumes a certain rate of change of a variable. It adjusts the latest reading by a certain fraction $\alpha \in (-1, 1)$ of the latest observed increment:
$$\hat{Y}_{t|t-h} = Y_{t-h} + \alpha(Y_{t-h} - Y_{t-2h}). \tag{7}$$

For $\alpha = 0$ this equation is reduced to the naïve scheme. A positive α coefficient means that a forecaster expects that the existing trends will remain, while the opposite sign indicates an expectation of trend reversal. We are assuming here that the forecasts are formed at regular intervals, i.e., every h period, but it is easy to extend the proposed models to arbitrarily chosen intervals. It should be noted that all adjustments of forecasts in this model (for longer horizons) depend only on the newly published data (and potentially on the revision, too).

If a forecast aims at correctly prognosing the initial reading, then only a publication for the next reference period will change the prediction value. However, if it aims at the final reading, then subsequent rounds of revisions may influence the adjustment of both Y_{t-h} and Y_{t-2h} (especially for forecasts with a short horizon, i.e., when h frequency is high).

iii) M3: the acceleration model assumes that a change in expectations takes place only if a change is observed in data themselves:
$$\hat{Y}_{t|t-h} = \hat{Y}_{t-h|t-2h} + \alpha(Y_{t-h} - Y_{t-2h}), \tag{8}$$

where $\hat{Y}_{t-h|t-2h} = E(Y_{t-h}|I_{t-2h})$. Similar as in the previous model, all revisions depend on the new data and potentially on the revision.

Paulina Ziembińska

---

  iv) M4: the first-degree adaptive model assumes that a forecaster adapts their previous forecast by a certain fraction of the latest error made:

$$\hat{Y}_{t|t-h} = \hat{Y}_{t-h|t-2h} + \alpha(Y_{t-h} - \hat{Y}_{t-h|t-2h}). \tag{9}$$

In this case the revisions may have an impact on the observed forecast error, therefore the question of how the forecaster defines their objective is of key significance here. If an error is adapted at each subsequent reading of a variable, a similar thing will happen to the forecast.

  v) M5: the mean reversion model assumes that the process will only slightly deviate from the long-term steady state. Mathematically, the forecast is a weighted average of the historical mean and the latest value observed:

$$\hat{Y}_{t|t-h} = \alpha\overline{Y} + \beta Y_{t-h}. \tag{10}$$

Two questions emerge as regards this model in the context of data revisions: what value should be adopted as a long-term mean and whether the last observed reading is updated when the data are revised. If yes, then for the sake of coherence we should assume that the average is calculated also based on the latest readings available.

We can define tests of specific extrapolative models based on the following regressions:

  i) Test 1: $\hat{Y}_{t|t-h} - Y_{t-h} = const + \epsilon_t$

  ii) Test 2: $\hat{Y}_{t|t-h} - Y_{t-h} = const + \alpha(Y_{t-h} - Y_{t-2h}) + \epsilon_t$

  iii) Test 3: $\hat{Y}_{t|t-h} - \hat{Y}_{t-h|t-2h} = const + \alpha(Y_{t-h} - Y_{t-2h}) + \epsilon_t$

  iv) Test 4: $\hat{Y}_{t|t-h} - \hat{Y}_{t-h|t-2h} = const + \alpha(Y_{t-h} - \hat{Y}_{t-h|t-2h}) + \epsilon_t$

  v) Test 5: $\hat{Y}_{t|t-h} = const + \alpha\overline{Y} + \beta Y_{t-h} + \epsilon_t$

and test two conditions:

$$H_0^{bias} : const = 0, \tag{11}$$

$$H_0^{model} : \hat{\alpha} = \alpha, \tag{12}$$

where $\hat{\alpha}$ is an ordinary least squares (OLS) estimate of α (analogical denotations are applied to other parameters of the models). The last hypothesis is different for the M5 scheme:

$$H_0^{model,M5} : \begin{cases} \hat{\alpha} = \alpha \\ \hat{\beta} = 1 - \alpha. \end{cases} \tag{13}$$

Let us note that testing the abovementioned hypotheses makes sense mainly if we have any a priori knowledge concerning the actual expectation formation process and

not if we search for its form. If we are not familiar with this process, which happens always when we analyse forecast data from an unknown model at our disposal, it is more adequate to use a more general test form. Pesaran (1985) postulates to estimate extrapolative models as distributed lag models, which yields consistent estimates even if the data are burdened with measurement errors. Let us consider the following regression:

$$\text{Test 6: } \hat{Y}_{t|t-h} = const + \beta \hat{Y}_{t-h|t-2h} + \gamma_1 Y_{t-h} + \gamma_2 Y_{t-2h} + \epsilon_t. \tag{14}$$

Due to short series, I am not going beyond the second lag. The literature indicates that the lags of higher order are usually not statistically significant and, beyond doubt, hard to interpret. For example, while analysing the inflation expectations, Pesaran (1985) indicates that *"except for one period lagged rate higher order lags do not exert any statistically significant influence on inflation expectations."*

I am testing both this general specification and tests corresponding to the specific models. Let us pay attention to the fact that the discussed regressions do not have a time structure, since they aggregate forecasts with different horizons. Therefore, we should not assume any structure of the residuals (as in Pesaran, 1985). The M1:M5 models are embedded in the model described by Equation (14). Below, I am defining restrictions imposed on the parameters of Equation (14) corresponding to the specific models. Note that, apart from M5, the intercept represents the forecast bias in all models – I am considering it in the estimation, and I am separately testing the $const = 0$ hypothesis. The bias may, for example, correspond to excessive optimism or pessimism displayed by a forecaster, who may formulate expectations in line with the extrapolative scheme and then adjust them based on behavioural factors.

$$H_0^{naive} : \begin{cases} \beta = \gamma_2 = 0 \\ \gamma_1 = 1; \end{cases} \tag{15}$$

$$H_0^{extra} : \begin{cases} \beta = 0 \\ \gamma_1 = 1 - \gamma_2; \end{cases} \tag{16}$$

$$H_0^{acc} : \begin{cases} \beta = 1 \\ \gamma_1 = -\gamma_2; \end{cases} \tag{17}$$

$$H_0^{adapt} : \begin{cases} \beta = 1 - \gamma_1 \\ \gamma_2 = 0; \end{cases} \tag{18}$$

$$H_0^{rev} : \begin{cases} \beta = \gamma_2 = 0 \\ \frac{const}{\bar{x}} = \gamma_1 + 1. \end{cases} \tag{19}$$

The above discussion aims at indicating problems with interpretation regarding basic extrapolative models in a situation where the data are subject to revision. The

objective behind testing these models is to explain how a forecaster understands the data generating process and how they learn from the data and own errors.

4   The Monte Carlo experiment

In this part I would like to shed light on statistical properties of all expectation tests discussed above. I am conducting a series of Monte Carlo experiments for different scenarios concerning the revision process and non-optimality of generated forecasts, at the same time extending the ones proposed by Patton and Timmermann (2012) by scenarios with non-zero mean measurement error and a shorter sample of 50 observations. First, I will characterise the experiment to discuss the power and size of defined tests. The test power is a probability of not making type II error: adopting the null hypothesis when it is in fact false. The test size is a probability of making type I error, namely rejecting the true null hypothesis.

4.1   Structure

Following the standards adopted in the literature and structure of the models fitted to the Polish real-time data (cf. Ziembińska, 2017), I am generating the main process as $Y_t \sim AR(1)$:

$$Y_t = \mu_Y + \phi(Y_{t-1} - \mu_Y) + \epsilon_t, \text{ for: } t = 1, ..., T; \epsilon_t \sim N(0, 1). \tag{20}$$

I am taking into consideration the following specifications: $\mu = 0.75$, $\phi = \{-0.5, 0.5\}$ and $T = \{50, 100, 1000\}$. In fact, I am generating adequately longer data and, following the structuring of appropriate series of forecasts, I am cutting off the last T observations that represent input data for the tests. This approach is different from the one applied by Patton and Timmermann (2012), where the sample was in fact appropriately shorter.

Next, I am introducing the measurement error (for example, resulting from the revision process):

$$\tilde{Y}_t = Y_t + \psi_t, \tag{21}$$

where $\psi_t \sim N(\mu_\psi, \sigma_\psi^2)$ with a mean: $\mu_\psi \in \{0, -0.1\mu_Y, 0.1\mu_Y\}$ and standard deviation: $\sigma_\psi \in \{0, \sqrt{\mu_Y}\sigma_Y, 2\sqrt{\mu_Y}\sigma_Y\}$. Compared to the literature, I am additionally analysing the behaviour of tests in the case of non-zero mean revisions which – as shown in Ziembińska (2017) – are not uncommon in the Polish macroeconomic publications.

I will be considering the optimal forecast (against the mean square error minimisation criterion):

$$\hat{Y}_{t|t-h} = \mu_Y + \phi^h(Y_{t-h} - \mu_Y) \tag{22}$$

and five models of extrapolative forecasts described by Equations (6)–(10).

As far as the extrapolative and acceleration schemes are concerned, I will analyse $\alpha \in \{-0.3, -0.1, 0.1, 0.3\}$, while for the adaptive and mean reversion: $\alpha \in \{0.2, 0.5, 0.7\}$. Where necessary, I am using the unconditional mean μ_Y as the initial forecasts. I am considering $h = 1, 4, 8$ horizons. Let us note that the discussed tests are based only on the extrapolative structure, which in practice means that in the case of no variability of the measurement error (namely, for $\sigma_\psi = 0$), the equations will be fully fitted by definition. Of course, we do not observe such a situation in practice since the real data do not derive from an ideal theoretical model and since a forecaster may not be familiar with that process. Therefore, for the extrapolative models we are omitting the case where $\sigma_\psi = 0$.

Wishing to analyse the power of tests, we need to generate sub-optimal forecasts, that is, those for which the tests should reject the null hypothesis. Let us define it as follows (cf. Patton and Timmermann, 2012),

$$\hat{Y}^{sub}_{t|t-h} = \hat{Y}_{t|t-h} + \sigma_{\xi,h}\xi_{t,t-h}, \tag{23}$$

where $\xi_{t,t-h} \sim N(0,1)$ and $\sigma_{\xi,h} \in \{\sqrt{\mu_Y}\sigma_Y, \frac{2(h-1)}{7}\sqrt{\mu_Y}\sigma_Y\}$. The latter specification corresponds to a situation where a standard deviation of the forecast error increases with its horizon (for $h = \{4, 8\}$).

4.2   Results

Tables 1, 2, and 3 present the results of forecast rationality tests. Panels for $\mu_\psi = 0$ and $T = 100$ replicate some results obtained by Patton and Timmermann (2012). Both the size and power of tests are highly similar to the ones presented in the original work. Test 3 is the only exception and demonstrates much lower power. Other results confirm the conclusions reached by the authors: the MZ test rejects the null hypothesis too frequently, especially for longer horizons; tests 1–5 based on inequalities have a much smaller size than the nominal one; the PT test has the best properties, especially when we use a forecast with the shortest horizon (*proxy*) instead of the realization of a variable; test 4 (of the decreasing MSF) has high power in the case where the forecast error increases with its horizon. When it comes to the PT and CG tests, it is worth paying attention to the monotonicity of power in relation to the measurement error variance in the case where the forecast error increases with its horizon. Only about 10% power in the case of the lack of a measurement error indicates that it is responsible for rejecting the null hypothesis (and not the test efficiency). Summing up, in the case of a zero mean measurement error, the PT test, based on the "optimal regression", and the CG test have the best performance; whereby if the forecast error increases with its horizon, the test 4, based on the property of decreasing mean square forecasts, provides best results.

Let us now analyse how these numbers change under the influence of shortening the sample to 50 observations. First, as regards tests 1–5, the size increases and is closer

Table 1: Size of the rationality tests

| | h = 4 | | | | | | | | | h = 8 | | | | | | | | |
| | $\mu_\psi = -0.1$ | | | $\mu_\psi = 0$ | | | $\mu_\psi = 0.1$ | | | $\mu_\psi = -0.1$ | | | $\mu_\psi = 0$ | | | $\mu_\psi = 0.1$ | | |
T Test	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)
50 T1	9.20	7.50	7.30	3.50	3.20	2.50	51.00	53.50	51.30	18.90	20.80	19.80	13.10	10.20	13.00	45.70	46.80	47.40
50 T2	4.60	2.60	2.20	5.20	2.70	1.50	6.20	3.40	2.70	18.20	11.50	8.30	17.60	8.90	7.90	17.20	12.50	7.90
50 T2 proxy	7.30	3.30	2.60	5.90	2.00	2.90	8.70	7.70	5.50	17.90	11.80	7.90	14.10	8.50	5.40	20.10	16.50	11.40
50 T3	17.40	10.20	7.10	6.60	2.40	1.80	72.30	68.90	60.60	20.40	13.40	8.20	10.10	5.20	4.20	69.20	61.70	51.80
50 T3 proxy	19.90	12.00	8.40	10.40	6.10	4.10	42.30	35.40	27.80	25.10	17.00	10.60	18.90	9.80	7.60	43.90	33.60	26.40
50 T4	3.70	2.50	1.80	6.00	2.60	1.90	1.40	0.80	1.50	10.70	7.00	3.50	11.20	7.60	3.90	6.40	5.00	2.80
50 T5	7.50	7.40	6.80	1.10	1.60	1.10	27.80	30.00	27.60	16.60	20.10	17.20	10.60	9.20	11.00	37.10	37.10	34.70
50 PT	42.90	43.10	43.10	12.10	9.60	9.20	98.60	99.40	98.80	32.10	30.10	29.00	7.70	7.20	8.20	92.70	93.40	93.50
50 PT proxy	24.00	24.90	26.50	9.10	8.50	11.20	55.70	60.00	56.40	15.60	18.00	17.50	8.40	7.60	9.40	38.90	37.30	36.20
50 CG	42.00	43.40	42.10	12.90	9.40	8.70	99.10	99.30	99.00	32.30	29.20	29.90	8.60	7.70	8.90	93.70	94.90	94.00
50 MZ min h	62.20	63.80	63.60	15.30	9.40	10.60	100.00	99.90	99.90	64.30	63.30	66.30	13.90	8.70	12.40	100.00	99.90	100.00
50 MZ	85.90	87.20	85.70	62.80	58.50	62.70	99.80	99.90	100.00	90.30	91.80	92.40	72.70	74.40	75.00	100.00	99.90	100.00

Table presents empirical size of the rationality tests based on 1000 simulations and asymptotic critical values. 10% nominal size is assumed. T denotes the sample size. (0), (1) and (2) correspond to the respective variability of the measurement errors: $\sigma_\psi \in \{0, \sqrt{\mu_Y}\sigma_Y, 2\sqrt{\mu_Y}\sigma_Y\}$. 'proxy' denotes the situation in which we use the forecast with the shortest horizon instead of the actual realization. The table continues on the next page.

Table 1: Size of the rationality tests cont.

| | | h = 4 | | | | | | | | | h = 8 | | | | | | | | |
| | | $\mu_\psi = -0.1$ | | | $\mu_\psi = 0$ | | | $\mu_\psi = 0.1$ | | | $\mu_\psi = -0.1$ | | | $\mu_\psi = 0$ | | | $\mu_\psi = 0.1$ | | |
T	Test	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)
100	T1	7.90	6.70	4.60	1.40	1.20	1.00	69.50	69.30	71.90	12.90	11.30	12.80	6.80	6.20	5.80	59.10	61.70	57.50
100	T2	3.50	1.60	0.90	3.40	1.50	0.50	2.40	1.10	0.60	13.30	6.20	3.40	4.20	6.50	2.60	11.90	5.40	4.10
100	T2 proxy	3.80	1.70	0.90	4.20	1.20	0.70	6.00	3.30	2.60	14.30	8.50	3.10	4.70	6.90	1.90	17.10	10.00	5.70
100	T3	19.80	11.00	5.40	3.40	1.00	0.70	83.20	82.90	83.50	18.30	8.30	5.90	5.50	1.80	0.60	88.20	84.60	79.00
100	T3 proxy	23.80	12.30	6.50	9.50	2.50	1.60	59.50	45.30	37.30	25.90	13.00	6.70	6.10	7.20	2.40	54.30	41.10	31.10
100	T4	2.60	0.80	0.60	6.40	1.70	1.00	0.50	0.10	0.00	9.00	3.90	1.30	3.70	4.90	1.20	4.30	2.10	1.50
100	T5	5.20	5.60	4.60	0.30	0.50	0.40	40.80	37.30	39.40	12.20	11.00	11.40	5.90	5.40	5.60	38.20	40.00	35.80
100	PT	78.30	79.40	78.60	14.40	9.50	11.80	100.00	100.00	100.00	62.40	63.50	64.50	11.20	9.10	10.40	100.00	100.00	99.90
100	PT proxy	51.90	48.30	49.50	10.50	8.40	11.20	92.30	89.60	90.60	36.60	34.80	36.80	11.20	10.00	11.50	78.80	78.90	79.40
100	CG	80.60	81.20	80.50	14.90	8.70	11.70	100.00	100.00	100.00	63.40	62.80	66.60	15.40	10.10	10.20	100.00	100.00	99.90
100	MZ min h	88.20	88.90	90.90	16.40	9.30	13.00	100.00	100.00	100.00	90.10	89.90	90.50	18.40	10.60	10.60	100.00	100.00	100.00
100	MZ	95.70	95.20	94.20	44.10	39.40	41.10	100.00	100.00	100.00	97.10	96.70	96.70	76.10	73.60	73.20	100.00	100.00	100.00

Table 2: Power of the rationality tests for $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$

T	Test	h = 4									h = 8								
		$\mu_\psi = -0.1$			$\mu_\psi = 0$			$\mu_\psi = 0.1$			$\mu_\psi = -0.1$			$\mu_\psi = 0$			$\mu_\psi = 0.1$		
		(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)
50	T1	10.80	15.30	37.70	9.80	16.20	17.00	10.00	17.00	36.60	28.60	32.20	45.50	29.00	32.20	37.10	27.70	34.50	48.00
50	T2	11.20	9.90	11.20	10.10	10.00	10.50	8.60	8.40	8.10	26.90	30.00	30.10	25.50	25.70	28.30	22.20	24.10	23.60
50	T2 proxy	12.60	13.70	11.60	12.90	13.00	13.90	11.10	12.00	13.60	31.30	32.40	29.40	30.70	28.70	31.00	25.30	26.90	27.80
50	T3	73.00	69.80	59.30	72.30	74.20	62.00	73.40	73.10	65.80	62.90	57.30	55.10	65.30	61.60	55.40	68.20	66.60	57.50
50	T3 proxy	68.30	72.00	74.10	72.50	73.20	73.20	73.90	71.40	75.20	49.60	49.60	56.90	54.70	55.80	57.30	59.60	60.80	63.60
50	T4	14.40	13.10	7.30	10.70	8.40	6.10	11.90	8.30	4.80	31.70	29.00	23.60	28.00	26.60	21.60	23.90	25.30	18.00
50	T5	13.00	15.00	18.50	13.30	13.90	17.30	12.40	14.80	18.70	28.00	30.70	34.40	27.40	32.70	35.40	28.70	30.60	34.70
50	PT	99.10	99.00	100.00	99.10	99.80	100.00	98.90	99.50	100.00	92.30	97.30	99.80	94.10	97.70	99.90	94.50	98.10	99.80
50	PT proxy	99.60	99.30	98.40	99.70	99.80	98.20	99.80	99.10	98.60	96.00	95.60	92.20	96.30	96.70	91.20	96.40	94.00	91.20
50	CG	95.20	98.00	100.00	97.30	98.60	100.00	96.80	99.10	99.90	90.10	95.60	99.70	92.80	96.10	99.90	92.60	96.80	99.90
50	MZ min h	99.80	99.90	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.80	100.00	100.00	100.00	99.90	100.00	100.00	100.00	100.00
50	MZ	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table presents empirical power of the rationality tests in the scenario where prediction errors do not depend on the forecast horizon. It is based on 1000 simulations and asymptotic critical values. T denotes the sample size. (0), (1) and (2) correspond to the respective variability of the measurement errors: $\sigma_\psi \in \{0, \sqrt{\mu_Y}\sigma_Y, 2\sqrt{\mu_Y}\sigma_Y\}$. 'proxy' denotes the situation in which we use the forecast with the shortest horizon instead of the actual realization. The table continues on the next page.

Table 2: Power of the rationality tests for $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$ cont.

| | | h=4 | | | | | | | | | h=8 | | | | | | | | |
| | | $\mu_\psi = -0.1$ | | | $\mu_\psi = 0$ | | | $\mu_\psi = 0.1$ | | | $\mu_\psi = -0.1$ | | | $\mu_\psi = 0$ | | | $\mu_\psi = 0.1$ | | |
T	Test	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)
100	T1	6.90	11.60	45.50	6.20	10.10	26.50	6.60	13.20	47.40	18.00	23.90	52.50	17.80	21.60	19.00	15.90	20.30	51.90
100	T2	8.00	7.20	7.70	5.40	4.50	6.50	4.80	4.50	4.60	19.40	17.50	17.20	13.20	14.00	16.70	13.10	12.80	12.70
100	T2 proxy	10.50	10.60	12.20	8.80	8.00	10.50	7.40	7.50	7.70	21.70	20.50	21.90	17.80	19.40	18.40	16.90	17.60	17.10
100	T3	44.50	33.20	12.90	51.40	43.40	17.20	57.30	49.40	21.70	30.50	23.10	10.50	39.40	35.70	13.10	49.10	39.80	19.20
100	T3 proxy	29.30	30.70	38.60	35.20	39.70	44.20	40.90	47.30	50.50	10.80	13.90	24.10	15.70	18.70	23.60	22.60	26.40	32.70
100	T4	11.90	6.10	3.00	6.70	6.60	2.30	5.00	5.20	1.70	21.80	18.80	13.80	15.80	14.50	12.80	14.30	13.40	11.20
100	T5	10.30	12.50	20.10	9.40	11.90	16.90	10.30	12.60	17.90	19.70	19.90	25.60	19.40	20.70	19.60	18.20	22.60	27.80
100	PT	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
100	PT proxy	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
100	CG	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.90	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
100	MZ min h	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
100	MZ	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 3: Power of the rationality tests for $\sigma_{\xi,h} = \frac{2(h-1)}{7}\sqrt{\mu_Y}\sigma_Y$

| | h=4 | | | | | | | | | h=8 | | | | | | | | |
| | $\mu_\psi = -0.1$ | | | $\mu_\psi = 0$ | | | $\mu_\psi = 0.1$ | | | $\mu_\psi = -0.1$ | | | $\mu_\psi = 0$ | | | $\mu_\psi = 0.1$ | | |
T Test	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)
50 T1	0.70	1.00	21.40	0.30	0.40	1.80	0.50	1.40	21.90	0.30	0.60	9.20	0.60	0.70	7.90	0.30	0.60	6.90
50 T2	8.30	8.60	9.20	8.10	5.90	6.10	5.10	5.40	6.30	27.00	23.90	25.70	22.20	22.40	22.00	18.80	20.80	20.40
50 T2 proxy	13.40	13.10	14.50	10.20	10.60	10.30	8.10	9.20	9.80	31.90	29.90	30.10	25.90	27.90	27.90	24.80	26.10	28.90
50 T3	32.10	47.40	76.90	26.50	40.20	75.70	14.90	31.30	71.60	30.80	28.90	25.20	36.00	34.90	30.90	43.50	42.60	36.80
50 T3 proxy	70.50	70.70	74.20	70.60	68.50	69.10	63.20	63.70	67.00	31.00	33.30	38.50	32.20	35.50	46.80	35.60	38.90	49.30
50 T4	55.70	44.40	28.80	47.50	38.90	22.50	36.60	33.00	20.70	84.60	86.50	95.70	85.10	90.10	98.80	88.60	93.20	90.70
50 T5	0.00	0.10	2.10	0.00	0.20	2.70	0.00	0.70	2.70	0.10	0.20	1.30	0.30	0.10	1.20	0.30	0.40	1.30
50 PT	12.30	45.80	99.10	11.90	40.30	99.00	10.20	42.40	99.50	12.20	35.30	92.80	10.10	31.30	94.70	9.40	33.10	93.90
50 PT proxy	59.80	65.80	74.00	61.50	64.60	72.10	61.80	65.70	76.90	44.70	46.30	55.50	40.10	45.10	54.20	45.70	50.90	54.20
50 CG	12.60	45.60	98.90	11.10	37.20	98.80	10.20	40.60	99.20	11.60	30.80	92.70	10.10	30.90	93.20	9.90	29.40	94.20
50 MZ min h	14.50	62.90	99.80	12.00	59.50	99.90	8.30	65.10	100.00	14.60	65.70	100.00	9.90	67.50	100.00	9.20	65.90	99.90
50 MZ	100.00	100.00	100.00	100.00	100.00	100.00	99.90	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table presents empirical power of the rationality tests in the scenario where prediction errors do depend on the forecast horizon. It is based on 1000 simulations and asymptotic critical values. T denotes the sample size. (0), (1) and (2) correspond to the respective variability of the measurement errors: $\sigma_\psi \in \{0, \sqrt{\mu_Y}\sigma_Y, 2\sqrt{\mu_Y}\sigma_Y\}$. "proxy" denotes the situation in which we use the forecast with the shortest horizon instead of the actual realization. The table continues on the next page.

Table 3: Power of the rationality tests for $\sigma_{\xi,h} = \frac{2(h-1)}{7}\sqrt{\mu_Y}\,\sigma_Y$ cont.

| | | h = 4 | | | | | | | | h = 8 | | | | | | | | |
| | | $\mu_\psi = -0.1$ | | | $\mu_\psi = 0$ | | | $\mu_\psi = 0.1$ | | | $\mu_\psi = -0.1$ | | | $\mu_\psi = 0$ | | | $\mu_\psi = 0.1$ | | |
T	Test	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)	(0)	(1)	(2)
100	T1	0.20	0.30	36.10	0.00	0.10	5.20	0.00	0.20	35.30	0.00	0.00	10.30	0.00	0.10	9.90	0.00	0.00	9.90
100	T2	4.70	5.80	6.80	4.10	4.80	4.40	3.30	3.80	2.80	14.00	17.30	16.30	13.70	12.90	15.20	11.00	12.00	12.10
100	T2 proxy	8.50	7.80	7.60	6.20	6.40	8.00	6.30	4.50	5.90	17.80	18.20	22.20	15.60	17.20	18.50	14.30	14.30	15.60
100	T3	44.20	67.10	71.00	29.70	53.60	75.10	20.00	37.40	79.20	23.40	22.70	21.60	26.40	23.80	22.50	31.00	27.70	23.30
100	T3 proxy	66.60	65.90	65.50	73.20	73.00	68.80	73.60	74.80	77.60	30.00	30.00	30.30	30.00	30.10	30.40	31.00	30.00	31.20
100	T4	69.90	63.90	40.70	67.90	55.30	35.50	59.20	46.20	31.40	90.00	90.10	90.40	90.00	90.20	90.80	90.20	90.10	91.10
100	T5	0.00	0.00	1.70	0.00	0.00	1.00	0.00	0.00	1.10	0.10	0.00	0.30	0.00	0.00	0.10	0.00	0.00	0.10
100	PT	15.40	79.30	100.00	10.30	77.30	100.00	11.20	79.70	100.00	11.50	63.70	100.00	10.10	61.70	100.00	10.50	64.50	100.00
100	PT proxy	94.50	94.60	97.20	91.10	95.10	96.90	92.70	95.70	97.50	82.40	86.20	90.60	79.10	86.60	91.60	80.90	84.40	92.90
100	CG	16.10	78.10	100.00	9.00	76.50	100.00	10.70	79.10	100.00	11.90	61.90	100.00	10.70	58.70	100.00	11.60	60.90	100.00
100	MZ min h	18.70	88.80	100.00	10.40	89.80	100.00	10.30	91.90	100.00	16.60	89.10	100.00	10.20	90.80	100.00	12.00	91.90	100.00
100	MZ	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

to or even exceeds the nominal 10% (particularly for longer horizons), while their power changes only slightly. When it comes to the PT, CG, and MZ tests, the size basically does not change under the influence of the sample reduction (except for the aggregated MZ test for $h = 4$), similarly as the 100% power for a constant forecast error. The size distortion observed for the MZ test is well documented in the literature. As regards certain tests analysed here, we are observing a similarly increasing size or power distortion, which can challenge the consistency of these tests. However, we need to emphasize that in most cases the differences are not significant, the analysed sample lengths are small (max. 100 observations) and theoretical critical values have been applied (replicating those used most frequently in practice). Potentially, they can also result from a calculation error, but the applied number of simulation repetitions (1000) constitutes a standard in the literature (compare e.g. Patton and Timmermann, 2012). In a situation where the forecast error increases along with its horizon, the power of the MZ test does not change, but for the PT and CG tests the probability of not making a type II error decreases. Notably, for the optimal test with a proxy, the power decreases from over 90% to 60-70% for a shorter horizon and from 80-90% to 40-50% for the longer horizon (depending on the distribution of a measurement error). Additionally, good properties of test 4 remain regardless of the sample length. Finally, let us discuss how the properties of the rationality tests change when the measurement error has a non-zero mean $\mu_\psi \neq 0$. When it comes to inequality tests, the size changes depending on the sign of the error mean. In particular, as regards the positive mean error in tests 1, 3, and 5, it increases to 40-70%, thus considerably exceeding the nominal level. This is an intuitive result for test 3, where the mean effect does not level in a simple way, but nevertheless, it is surprising for other tests. When it comes to tests based on regressions, the non-zero measurement error causes a far more frequent rejection of the null hypothesis, in line with the expectations. Therefore causing the increase in the size of tests, regardless of an exact form of the error and the forecast horizon. Again, it is worth noting the asymmetry between the positive and negative error mean – the latter has lower impact on the weakening of the regressive tests' properties. The results concerning the power of tests do not change considerably for various values of the mean measurement error, but it should be noted that they are often monotonic in relation to its variance – the higher the variability of error (e.g., revision), the more frequently we will reject the null hypothesis.

Let us now move on to the discussion of results of extrapolative tests. Tables 4 and 5 present results for the naïve scheme for samples consisting of 50, 100, and 1000 observations. As regards the H_0^{bias} test: the size is similar to the nominal 10% only if the measurement error has a zero mean. For smaller samples, in the case of a non-zero measurement error mean, the bias test has a better size when this error is more volatile (i.e., when σ_ψ is higher), but then its power considerably decreases. Conclusions for the bias test are identical for all discussed models and generally do not offer good properties. Particularly, regardless of the size of the analysed sample and at the lack of a measurement error, our suboptimality condition causes the rejection

Table 4: Size of Test 1 (naïve)

| | | | $T = 50$ | | | | | | $T = 100$ | | | | | | $T = 1000$ | | | | | |
| | | | H_0^{bias} | | | $H_0^{naïve}$ | | | H_0^{bias} | | | $H_0^{naïve}$ | | | H_0^{bias} | | | $H_0^{naïve}$ | | |
ϕ	μ_ψ	σ_ψ	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8
0.5	0	$\sqrt{\mu_Y}\sigma_Y$	11.30	11.10	11.00	9.60	9.70	11.50	11.50	12.20	8.80	9.00	11.60	10.40	9.90	10.60	10.00	9.20	9.30	12.10
0.5	0	$2\sqrt{\mu_Y}\sigma_Y$	10.10	7.70	10.60	10.60	9.30	11.00	8.10	10.20	9.80	8.90	10.60	10.30	9.60	11.90	9.70	10.60	9.00	10.30
0.5	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	40.60	37.60	38.10	11.40	9.80	9.50	63.60	67.30	62.30	8.70	8.80	9.60	100.00	100.00	100.00	12.30	9.20	11.20
0.5	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	19.80	18.20	17.70	10.60	10.10	8.40	27.20	27.30	25.60	9.30	9.80	10.00	92.40	94.60	93.40	9.10	8.40	11.20

Table presents empirical size of Test 1 (naïve) with hypotheses (11) (H_0^{bias}) and (15) ($H_0^{naïve}$). It is based on 1000 simulations and asymptotic critical values. 10% nominal size is assumed. T denotes the sample size.

Table 5: Power of Test 1 (naïve)

ϕ	T	μ_ψ	σ_ψ	H_0^{bias}					H_0^{naive}				
				$\sigma_{\xi,h}=\sqrt{\mu_Y\sigma_Y}$			$\sigma_{\xi,h}=\frac{2(h-1)}{7}\sqrt{\mu_Y\sigma_Y}$		$\sigma_{\xi,h}=\sqrt{\mu_Y\sigma_Y}$			$\sigma_{\xi,h}=\frac{2(h-1)}{7}\sqrt{\mu_Y\sigma_Y}$	
				h=1	h=4	h=8	h=4	h=8	h=1	h=4	h=8	h=4	h=8
0.5	50			8.70	9.80	9.20	11.70	11.80	10.90	11.50	9.10	8.50	10.10
0.5	100			11.30	11.70	9.70	9.50	9.50	10.40	9.90	10.10	9.70	7.90
0.5	1000			9.50	10.10	8.10	10.40	9.70	10.30	10.20	9.30	11.00	10.10
0.5	50	0	$\sqrt{\mu_Y\sigma_Y}$	9.60	10.00	11.30	10.40	9.90	11.60	10.80	10.20	9.90	9.90
0.5	100	0	$\sqrt{\mu_Y\sigma_Y}$	12.00	10.30	9.90	9.30	9.40	10.30	8.80	8.20	8.20	10.40
0.5	1000	0	$\sqrt{\mu_Y\sigma_Y}$	10.90	9.80	10.40	9.50	9.20	10.10	10.00	8.00	9.50	10.40
0.5	50	0	$2\sqrt{\mu_Y\sigma_Y}$	10.60	9.50	10.20	10.50	10.50	8.00	8.70	9.10	11.40	11.20
0.5	100	0	$2\sqrt{\mu_Y\sigma_Y}$	10.70	9.70	10.10	9.20	9.70	10.70	8.60	11.70	9.60	10.40
0.5	1000	0	$2\sqrt{\mu_Y\sigma_Y}$	10.70	8.50	10.50	9.40	9.00	10.70	8.10	9.70	9.60	12.20
0.5	50	$0.1\mu_Y$	$\sqrt{\mu_Y\sigma_Y}$	19.30	23.40	19.80	25.90	13.40	9.80	9.20	8.70	10.40	10.00
0.5	100	$0.1\mu_Y$	$\sqrt{\mu_Y\sigma_Y}$	31.80	31.70	32.50	34.40	16.20	11.30	10.00	9.50	10.60	9.60
0.5	1000	$0.1\mu_Y$	$\sqrt{\mu_Y\sigma_Y}$	97.60	97.90	98.10	99.80	68.80	10.40	11.30	10.50	8.70	11.30
0.5	50	$0.1\mu_Y$	$2\sqrt{\mu_Y\sigma_Y}$	11.30	14.50	13.40	14.50	11.90	10.00	10.70	9.40	11.20	10.60
0.5	100	$0.1\mu_Y$	$2\sqrt{\mu_Y\sigma_Y}$	18.50	16.80	17.70	19.30	13.60	9.40	9.80	9.90	9.50	9.00
0.5	1000	$0.1\mu_Y$	$2\sqrt{\mu_Y\sigma_Y}$	71.30	69.60	76.20		40.40	8.70	9.00	10.00	10.50	9.80

Table presents empirical power of Test 1 (naïve) with hypotheses (11) (H_0^{bias}) and (15) (H_0^{naive}). It is based on 1000 simulations and asymptotic critical values. T denotes the sample size.

of the null hypothesis only for around 10% of samples. Higher values of this test's power (especially for the sample of 1000 observations) only reflect the non-zero mean measurement error, thus there is also a remarkably high probability of rejecting the true null hypothesis. The aggregated test based on Equation (12) demonstrates particularly good properties in terms of size, but its power is very low in most cases. The distribution of estimated coefficients in this regression is centred in accordance with the hypothesis H_0^{naive}, regardless of the forecast error, which makes us reject it too seldom. Therefore, my conclusion is that in the case of the naïve model, even with the zero mean forecast error the discussed tests have too low power to reject the false null hypothesis and should not be used.

Tables 6, 7, and 8 present results for the first order extrapolative model. As regards the bias test, it is worth paying attention to differences resulting from the α parameter. For the positive coefficient (denoting the expected trend continuation), the bias test H_0^{bias} has a much smaller size than for the negative coefficient. It translates into a test size that is smaller than the nominal 10% for $\alpha > 0$ and zero mean measurement error. It should be noted that for the non-zero measurement error the size distortion of this test (similarly as for the aggregated test) increases with the length of sample. It results from the higher concentration of the intercept distribution around the non-zero measurement mean (cf. Figure A.1 in the Appendix).

Comparison of the size and power of tests H_0^{model} across the extrapolative models indicates that they do not perform very well – the proportion of rejected null hypotheses basically does not depend on the imposed forecast error. This conclusion is coherent for various levels of the measurement error, sample length and forecast horizons. It indicates that the imposed forecast error increases the uncertainty of estimation enough to make it difficult to obtain statistical significance. As regards the aggregated test based on Equation (14), a reasonable size – smaller than 10% – received for shorter samples also reflects the lack of statistical significance. It is worth noting that the higher the absolute value of coefficient α, the worse the statistical properties. A comparison of the results of size and power of the aggregated test (H_0^{extra}) suggests that the size is often greater than the power, which reveals that the imposed measurement error increases the uncertainty of estimation enough to make it difficult to obtain statistical significance.

Due to such weak properties observed, I will analyse the estimation results in more detail. The analysed charts are presented in the Appendix. Figures A.1 and A.2 show examples (for $\phi = 0.5$, $h = 4$ and $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$) of distributions of estimated intercept and coefficient α for tests H_0^{bias} i H_0^{model}. The charts clearly demonstrate that a non-zero measurement error is responsible for the rejection of the no-bias hypothesis, regardless of the imposed forecast error. For the H_0^{model} tests, the distributions are remarkably similar for all analysed configurations. The imposed suboptimality of forecasts changes the volatility of $\hat{\alpha}$ only slightly, which means that the size and power of a test depend mainly on the sample length (the shorter the sample, the less of the test's statistical significance). When it comes to the aggregated

Table 6: Size of Test 2 (extrapolative)

| | | | T = 50 | | | | | | | | | T = 100 | | | | | | | | |
| | | | H_0^{bias} | | | H_0^{model} | | | H_0^{extra} | | | H_0^{bias} | | | H_0^{model} | | | H_0^{extra} | | |
α	μ_ψ	$\sigma_\psi \sigma_Y$	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8
-0.1	0	$\sqrt{\mu_Y}\sigma_Y$	12.10	14.10	12.40	10.30	8.80	9.10	13.80	13.20	13.70	11.70	13.10	14.20	9.60	10.30	7.40	21.60	19.60	21.50
-0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	13.20	12.40	11.90	8.30	8.30	7.40	13.20	14.60	14.90	14.10	13.20	14.10	8.60	8.90	8.20	22.70	22.40	22.10
-0.1	0.1μ_Y	$\sqrt{\mu_Y}\sigma_Y$	49.60	48.60	42.70	10.10	8.10	8.40	12.30	14.20	15.20	65.70	68.00	70.00	8.60	7.40	7.10	21.90	21.40	22.30
-0.1	0.1μ_Y	$2\sqrt{\mu_Y}\sigma_Y$	23.60	22.80	22.90	9.90	8.70	8.30	14.30	13.60	15.50	31.30	32.50	32.30	9.50	10.30	8.00	20.00	21.10	21.50
-0.3	0	$\sqrt{\mu_Y}\sigma_Y$	20.10	20.40	19.60	7.60	5.30	6.40	63.60	60.70	64.00	21.00	18.90	21.30	7.00	4.80	4.50	93.30	91.70	91.10
-0.3	0	$2\sqrt{\mu_Y}\sigma_Y$	20.60	20.60	19.20	6.90	5.00	4.30	68.40	69.00	69.40	22.40	22.00	20.40	6.50	4.50	3.00	94.90	94.60	95.30
-0.3	0.1μ_Y	$\sqrt{\mu_Y}\sigma_Y$	56.50	54.90	55.70	6.60	4.70	5.40	63.20	63.00	65.30	75.60	79.60	79.90	7.60	3.50	3.50	91.50	92.60	92.30
-0.3	0.1μ_Y	$2\sqrt{\mu_Y}\sigma_Y$	30.70	31.30	28.90	6.30	6.10	4.50	68.10	69.50	70.80	40.70	42.60	38.20	8.20	4.30	4.50	96.10	94.90	95.20
0.1	0	$\sqrt{\mu_Y}\sigma_Y$	5.70	5.80	8.10	8.20	10.10	12.30	14.20	12.70	14.10	5.90	8.60	6.50	9.40	11.30	9.60	18.70	19.30	20.10
0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	8.30	7.30	5.90	10.70	11.30	12.10	15.30	13.90	14.30	8.30	8.10	7.50	11.40	12.80	10.70	21.90	17.40	18.10
0.1	0.1μ_Y	$\sqrt{\mu_Y}\sigma_Y$	33.00	31.40	36.30	9.70	10.30	11.70	15.10	15.40	15.30	59.10	58.40	60.80	11.20	9.40	10.40	20.50	18.90	19.90
0.1	0.1μ_Y	$2\sqrt{\mu_Y}\sigma_Y$	13.60	13.90	16.20	8.00	10.80	14.10	15.20	13.60	14.80	22.40	19.50	21.40	9.80	10.40	12.50	19.70	19.50	20.40
0.3	0	$\sqrt{\mu_Y}\sigma_Y$	2.90	3.20	3.40	11.00	13.30	14.50	35.10	38.40	33.20	3.30	3.20	3.10	11.70	13.00	12.90	59.70	57.10	57.00
0.3	0	$2\sqrt{\mu_Y}\sigma_Y$	1.90	3.60	4.40	10.90	15.20	15.10	34.00	33.40	36.40	3.30	3.00	2.20	9.90	13.20	13.20	60.00	63.80	60.10
0.3	0.1μ_Y	$\sqrt{\mu_Y}\sigma_Y$	22.60	20.60	19.60	12.20	14.90	13.00	37.10	32.10	35.20	42.50	41.00	39.90	9.70	13.80	13.00	58.90	57.80	56.20
0.3	0.1μ_Y	$2\sqrt{\mu_Y}\sigma_Y$	6.40	8.60	7.20	11.90	12.60	12.60	36.00	32.80	34.00	12.50	12.70	11.70	11.80	14.20	12.70	60.50	63.10	61.60

Table presents empirical size of Test 2 (extrapolative) with hypotheses (11) (H_0^{bias}),(12) (H_0^{model}) and (16) (H_0^{extra}). It is based on 1000 simulations and asymptotic critical values. 10% nominal size is assumed. T denotes the sample size. The table continues on the next page.

Table 6: Size of Test 2 (extrapolative) cont.

α	μ_ψ	σ_ψ	H_0^{bias}			H_0^{model}			H_0^{extra}		
			h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8
-0.1	0	$\sqrt{\mu_Y}\sigma_Y$	11.80	14.00	14.60	9.70	9.40	6.30	92.50	92.50	93.80
-0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	16.40	12.90	10.90	10.00	8.00	8.80	93.80	92.70	92.90
-0.1	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	100.00	100.00	100.00	8.10	6.20	6.80	92.70	92.80	93.00
-0.1	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	95.60	94.90	94.60	8.10	7.40	8.10	92.90	93.10	91.70
-0.3	0	$\sqrt{\mu_Y}\sigma_Y$	21.10	20.70	24.00	5.80	4.50	4.50	100.00	100.00	100.00
-0.3	0	$2\sqrt{\mu_Y}\sigma_Y$	20.30	22.80	20.00	6.70	5.40	3.60	100.00	100.00	100.00
-0.3	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	100.00	100.00	100.00	6.70	4.20	4.00	100.00	100.00	100.00
-0.3	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	96.50	97.70	96.70	6.70	5.10	4.10	100.00	100.00	100.00
0.1	0	$\sqrt{\mu_Y}\sigma_Y$	7.80	5.80	6.30	9.30	12.90	10.00	83.70	82.80	82.30
0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	7.20	6.60	7.10	11.40	12.90	9.60	83.50	82.00	82.80
0.1	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	100.00	100.00	100.00	11.70	12.60	12.90	82.10	82.70	82.70
0.1	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	91.70	91.40	92.50	9.90	11.20	11.30	85.10	81.30	79.80
0.3	0	$\sqrt{\mu_Y}\sigma_Y$	3.30	2.70	2.50	11.40	14.10	12.40	100.00	100.00	100.00
0.3	0	$2\sqrt{\mu_Y}\sigma_Y$	3.30	2.60	3.10	9.70	12.40	12.30	100.00	100.00	100.00
0.3	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	100.00	100.00	100.00	11.60	14.90	14.90	100.00	100.00	100.00
0.3	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	80.40	84.00	83.00	13.20	13.00	12.40	100.00	100.00	100.00

The H_0^{bias}, H_0^{model}, H_0^{extra} columns are grouped under $T = 1000$.

P. Ziembińska

Paulina Ziembińska

Table 7: Power of Test 2 (extrapolative) for $T = 1000$

N	α	μ_ψ	σ_ψ	\multicolumn{9}{c}{$\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$}	\multicolumn{6}{c}{$\sigma_{\xi,h} = \frac{2(h-1)}{7}\sqrt{\mu_Y}\sigma_Y$}
				\multicolumn{3}{c}{H_0^{bias}}	\multicolumn{3}{c}{H_0^{model}}	\multicolumn{3}{c}{H_0^{extra}}	\multicolumn{2}{c}{H_0^{bias}}	\multicolumn{2}{c}{H_0^{model}}	\multicolumn{2}{c}{H_0^{extra}}
				h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=4	h=8	h=4	h=8	h=4	h=8
1000	-0.1	0	0	10.00	10.90	11.20	9.10	9.50	9.10	10.80	9.20	10.00	9.00	9.80	10.00	12.60	7.90	7.40
1000	-0.1	0	$\sqrt{\mu_Y}\sigma_Y$	9.80	9.80	10.00	11.60	11.50	10.00	22.90	25.60	23.60	9.80	10.30	11.80	10.00	32.10	10.30
1000	-0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	10.70	11.40	11.10	10.90	9.70	7.90	37.70	39.50	36.10	13.10	11.20	8.90	10.20	44.90	12.30
1000	-0.1	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	98.70	97.70	98.10	7.90	9.20	9.80	22.10	23.20	21.70	99.50	68.20	11.10	10.80	28.80	9.60
1000	-0.1	$0.1\mu_Y$	0	100.00	100.00	100.00	11.40	8.90	10.50	11.50	12.00	10.70	100.00	92.70	9.00	9.60	9.00	9.40
1000	-0.1	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	72.70	72.80	71.50	9.50	8.90	9.00	33.70	34.90	36.40	79.60	40.20	9.80	9.30	44.20	14.30
1000	-0.3	0	0	9.30	9.90	10.50	9.20	9.60	11.40	8.80	8.70	10.50	9.00	10.20	9.90	9.30	10.30	10.60
1000	-0.3	0	$\sqrt{\mu_Y}\sigma_Y$	11.70	13.90	13.10	9.50	9.80	8.30	80.80	79.50	80.70	13.50	12.00	9.60	10.50	92.30	19.30
1000	-0.3	0	$2\sqrt{\mu_Y}\sigma_Y$	13.90	14.10	14.00	8.20	7.20	8.00	98.20	98.60	98.70	14.70	11.50	7.30	10.10	99.80	33.10
1000	-0.3	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	98.80	98.80	98.70	10.10	8.90	8.30	75.10	73.80	75.10	99.60	69.60	9.00	9.70	89.20	18.70
1000	-0.3	$0.1\mu_Y$	0	100.00	100.00	100.00	10.00	10.00	10.00	10.00	9.20	11.70	100.00	91.20	10.40	11.90	10.10	10.60
1000	-0.3	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	72.90	73.20	75.70	8.40	7.90	7.90	96.40	97.90	97.00	84.40	40.20	7.00	9.50	99.60	30.90

Table presents empirical power of Test 2 (extrapolative) with hypotheses (11) (H_0^{bias}), (12) (H_0^{model}), and (16) (H_0^{extra}). It is based on 1000 simulations and asymptotic critical values. Results for sample size $T = 1000$. The table continues on the next page.

Table 7: Power of Test 2 (extrapolative) for $T = 1000$ cont.

| | | | | $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$ | | | | | | | | | $\sigma_{\xi,h} = \frac{2(h-1)}{\tau}\sqrt{\mu_Y}\sigma_Y$ | | | | | |
| | | | | H_0^{bias} | | | H_0^{model} | | | H_0^{extra} | | | H_0^{bias} | | H_0^{model} | | H_0^{extra} | |
N	α	μ_ψ	σ_ψ	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=4	h=8	h=4	h=8	h=4	h=8
1000	0.1			9.70	9.60	9.00	11.40	9.70	9.50	10.50	10.10	9.50	9.40	10.00	11.80	10.60	8.10	8.90
1000	0.1	0	$\sqrt{\mu_Y}\sigma_Y$	9.00	8.80	8.20	11.80	9.80	10.50	26.70	27.80	27.20	8.90	9.80	11.00	8.90	31.30	13.70
1000	0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	6.70	7.80	8.70	12.70	10.30	10.60	38.20	39.10	38.20	7.00	11.40	10.60	10.40	45.90	17.00
1000	0.1	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	97.70	97.70	97.40	10.20	9.40	10.60	23.80	27.10	23.90	98.90	66.40	10.10	9.90	32.90	11.60
1000	0.1	$0.1\mu_Y$	0	100.00	100.00	100.00	8.80	10.20	10.10	11.10	8.90	9.50	100.00	91.20	8.60	7.50	10.20	11.40
1000	0.1	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	67.90	71.70	65.50	9.30	11.20	10.10	38.20	34.50	34.40	71.40	36.60	10.90	10.40	43.00	16.60
1000	0.3			10.80	8.60	10.10	10.00	9.60	8.50	9.80	8.80	8.50	10.30	8.40	9.80	10.00	9.40	8.40
1000	0.3	0	$\sqrt{\mu_Y}\sigma_Y$	5.10	5.80	6.80	8.70	11.80	10.00	94.80	93.20	91.90	7.30	8.40	11.60	10.20	97.60	34.00
1000	0.3	0	$2\sqrt{\mu_Y}\sigma_Y$	6.00	5.70	6.50	11.70	10.70	9.80	99.40	99.70	99.50	3.80	7.70	10.50	10.00	99.70	57.80
1000	0.3	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	95.80	97.00	96.60	11.50	13.50	11.00	93.90	91.00	90.80	98.60	64.50	12.50	10.80	96.70	29.10
1000	0.3	$0.1\mu_Y$	0	100.00	100.00	100.00	10.00	10.60	10.20	9.30	9.90	10.00	100.00	92.00	10.30	10.50	9.50	10.00
1000	0.3	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	58.90	60.30	60.50	10.60	13.90	11.00	99.00	99.00	99.20	65.30	31.80	12.80	10.30	100.00	55.10

Paulina Ziembińska

Table 8: Power of Test 2 (extrapolative) for $T = 100$

N	α	μ_ψ	σ_ψ	$\sigma_{\xi,h}=\sqrt{\mu_Y}\sigma_Y$ H_0^{bias} h=1	h=4	h=8	H_0^{model} h=1	h=4	h=8	H_0^{extra} h=1	h=4	h=8	$\sigma_{\xi,h}=\frac{2(h-1)}{7}\sqrt{\mu_Y}\sigma_Y$ H_0^{bias} h=4	h=8	H_0^{model} h=4	h=8	H_0^{extra} h=4	h=8
100	-0.1			10.00	10.60	11.50	9.90	10.60	9.40	8.40	10.90	9.90	11.00	10.60	10.30	9.40	8.70	11.10
100	-0.1	0	$\sqrt{\mu_Y}\sigma_Y$	11.70	10.20	9.50	11.20	10.90	10.20	11.20	10.90	9.60	11.10	11.80	8.80	9.30	10.20	8.70
100	-0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	12.20	12.90	12.70	8.90	8.60	10.40	10.00	12.70	11.70	12.20	11.50	9.40	7.50	12.70	9.70
100	-0.1	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	34.00	35.50	33.30	9.90	9.20	8.80	9.80	11.00	11.70	37.60	18.80	8.20	9.70	10.70	10.10
100	-0.1	$0.1\mu_Y$	0	59.40	57.50	59.00	10.10	11.10	11.50	11.50	10.40	9.50	71.90	26.30	9.20	10.10	9.40	9.10
100	-0.1	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	18.80	19.10	20.70	9.30	9.30	9.70	12.10	12.20	11.10	20.00	13.80	8.20	10.40	13.10	12.00
100	-0.3			10.30	9.70	10.40	10.90	9.00	10.80	11.30	9.60	9.60	9.70	8.90	11.30	10.10	9.50	10.50
100	-0.3	0	$\sqrt{\mu_Y}\sigma_Y$	14.00	12.10	13.70	7.60	8.40	8.80	16.80	18.50	18.80	11.40	10.10	10.00	9.40	21.70	10.90
100	-0.3	0	$2\sqrt{\mu_Y}\sigma_Y$	13.60	11.90	12.40	9.30	6.90	8.10	25.50	27.20	24.10	16.90	10.40	7.00	9.60	38.80	12.80
100	-0.3	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	38.30	35.30	33.90	10.00	9.50	8.20	16.80	15.50	15.90	41.30	18.90	7.20	10.30	19.60	11.50
100	-0.3	$0.1\mu_Y$	0	61.70	59.50	61.20	10.00	10.80	10.00	9.40	8.80	10.00	69.30	24.10	9.50	9.90	10.20	11.30
100	-0.3	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	22.40	25.90	24.00	9.00	7.30	8.40	24.50	22.90	26.00	28.20	14.30	7.20	9.20	35.40	10.20

Table presents empirical power of Test 2 (extrapolative) with hypotheses (11) (H_0^{bias}), (12) (H_0^{model}) and (16) (H_0^{extra}). It is based on 1000 simulations and asymptotic critical values. Results for sample size $T = 100$. The table continues on the next page.

Table 8: Power of Test 2 (extrapolative) for $T = 100$ cont.

Columns 5–13 belong to $\sigma_{\xi,h} = \sqrt{\mu_Y}\,\sigma_Y$; columns 14–19 belong to $\sigma_{\xi,h} = \frac{2(h-1)}{7}\sqrt{\mu_Y}\,\sigma_Y$.

N	α	μ_ψ	σ_ψ	H_0^{bias} h=1	h=4	h=8	H_0^{model} h=1	h=4	h=8	H_0^{extra} h=1	h=4	h=8	H_0^{bias} h=4	h=8	H_0^{model} h=4	h=8	H_0^{extra} h=4	h=8
100	0.1	0	0	12.10	10.00	9.90	9.90	11.80	10.70	10.20	10.70	9.50	9.80	10.10	10.00	11.30	11.00	10.20
100	0.1	0	$\sqrt{\mu_Y}\sigma_Y$	7.90	8.80	9.20	11.80	8.50	12.80	10.90	12.80	12.20	7.30	10.10	10.50	11.90	12.80	9.30
100	0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	8.20	8.80	8.40	9.90	11.10	10.80	13.70	12.60	12.90	8.70	8.70	11.40	11.20	12.40	11.30
100	0.1	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	30.50	28.50	29.80	10.80	10.00	10.20	11.20	10.90	10.90	33.70	16.90	12.10	11.40	13.20	10.40
100	0.1	$0.1\mu_Y$	0	62.20	61.80	62.10	9.60	11.20	8.10	11.00	9.50	8.70	69.20	24.10	9.80	8.40	9.60	11.20
100	0.1	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	15.70	15.90	16.30	9.80	10.40	10.80	13.60	14.00	12.50	17.10	13.30	10.60	10.70	12.40	10.50
100	0.3	0	0	10.70	12.80	11.90	9.90	11.10	8.40	10.10	10.10	11.60	8.90	7.80	9.50	12.30	8.70	11.20
100	0.3	0	$\sqrt{\mu_Y}\sigma_Y$	5.40	6.40	7.10	13.00	14.60	11.20	24.80	24.30	24.30	5.40	7.80	10.60	9.10	27.90	12.10
100	0.3	0	$2\sqrt{\mu_Y}\sigma_Y$	5.00	4.50	5.70	9.20	12.80	12.10	34.00	31.80	32.30	4.50	7.80	12.10	10.50	40.10	17.10
100	0.3	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	25.00	24.90	26.50	11.30	12.50	13.90	23.50	23.00	23.10	16.30	12.80	12.80	9.30	24.60	12.90
100	0.3	$0.1\mu_Y$	0	57.40	62.00	62.20	9.80	9.00	11.30	10.80	9.90	8.70	71.80	22.70	10.60	9.70	10.20	10.10
100	0.3	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	12.30	10.50	12.60	12.20	12.90	11.70	31.40	32.80	30.80	10.10	14.10	10.90	11.00	36.30	13.90

test, a distribution of intercept from Equation (14) is centred around zero. Rejection of the null hypothesis H_0^{extra} in most cases (regardless of the forecast error) results from the shift of the coefficient $\hat{\beta}$ distribution centre in relation to zero to ca. -0.15, which implies that $\hat{\gamma}_1 + \hat{\gamma}_2 \approx 0.15$. An exception is a situation where the measurement error variance $\sigma_\psi^2 = 0$, when distribution $\hat{\beta}$ is centred around zero and causes non-rejection of the joined hypothesis H_0^{extra}. Reduction of the sample size considerably increases dispersion of the estimated parameters. These results reveal that these tests should essentially not be used, regardless of a specification.

Tables 9 and 10 present results for the adaptive model. In line with the expectations, the H_0^{bias} test rejects the true null hypothesis at the nominal level only for a zero mean measurement error, but even then, the test power does not exceed 60% and declines with the forecast horizon. Sizes of H_0^{model} and H_0^{adapt} tests are similar to the nominal level, regardless of the sample length, form of the measurement error and forecast horizon. One should note that the power of these tests strongly depends on the α coefficient, which defines how strongly a forecaster adapts their last prediction by a realized error. The higher the adjustment level, the less frequently we reject a false null hypothesis for both tests. The H_0^{model} test has higher power than the test based on Equation (14) for a longer horizon and for a more volatile measurement error in case of higher adaptive coefficients. The forecast horizon and sample length have only a slight impact on the power of discussed tests for a constant forecast error (in relation to the horizon) and for low adaptive coefficients. With higher α the longer horizon and a smaller sample reduce the test power. Overall, the H_0^{model} test has good properties for the adaptive model, regardless of the forecast horizon and the sample length.

Let us now analyse the results for the acceleration model. Irrespective of the form of measurement error or forecast error (we are only analysing situations where the latter has a zero mean), we do not reject H_0^{bias} in any of the considered cases (therefore, the results are not presented in a table). This is not a negative feature of the test and we can presume that it results from the fact that on both sides of the estimated equation (Test 3) we observe differences derived from the same process – either forecasts or actual values. This implies that the mean measurement error with a non-zero value that can disturb other bias tests has been eliminated owing to the structure of variables used for the estimation. The H_0^{model} test has low power of 12-20%, regardless of the sample size and the analysed forecast error. Distribution of the $\hat{\alpha}$ coefficient is more dispersed for short samples and suboptimal forecasts, but it still concentrates around a theoretical value (cf. Figure 1). The H_0^{acc} test also has weak properties – the null hypothesis is nearly always rejected, regardless of parameters. A more detailed analysis of distribution of coefficients in Equation (12) shows that such weak properties of this test may result from taking an intercept into account. Therefore, we are conducting an additional experiment neglecting the intercept in the estimation. The results are presented in Tables 11 and 12. The power of thus formulated test is close to 100% for all analysed schemes. What is interesting is that

Table 9: Size of Test 4 (adaptive)

α	μ_ψ	σ_ψ	$T=50$ H_0^{bias} h=1	h=4	h=8	H_0^{model} h=1	h=4	h=8	H_0^{adapt} h=1	h=4	h=8	$T=100$ H_0^{bias} h=1	h=4	h=8	H_0^{model} h=1	h=4	h=8	H_0^{adapt} h=1	h=4	h=8
0.2	0	$\sqrt{\mu_Y}\sigma_Y$	12.00	11.00	10.40	10.00	9.80	9.50	12.20	10.90	9.90	10.90	12.90	9.40	10.60	12.40	11.80	10.30	11.30	10.50
0.2	0	$2\sqrt{\mu_Y}\sigma_Y$	10.00	8.10	10.90	12.00	9.00	10.50	11.50	11.70	11.20	9.60	10.20	9.00	8.20	10.40	9.00	10.60	13.10	12.00
0.2	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	40.70	41.10	39.40	10.70	8.80	10.40	11.40	11.50	11.40	64.20	65.40	64.80	9.70	11.10	10.20	11.00	11.90	12.50
0.2	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	18.30	16.40	19.10	8.30	8.30	9.20	13.30	14.80	10.80	23.00	27.00	24.40	8.10	10.50	11.90	11.00	12.50	12.50
0.5	0	$\sqrt{\mu_Y}\sigma_Y$	11.70	11.10	10.70	9.70	9.90	9.80	11.70	11.90	10.90	10.80	13.00	8.90	9.80	10.40	10.10	9.50	11.60	10.10
0.5	0	$2\sqrt{\mu_Y}\sigma_Y$	10.30	7.80	10.40	10.10	9.30	9.40	9.90	10.10	10.90	9.30	11.10	9.70	9.60	10.40	10.20	10.40	13.00	10.80
0.5	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	40.80	43.40	40.00	9.50	8.50	10.30	11.20	10.00	10.00	64.50	65.00	65.10	10.70	11.00	10.20	11.40	11.40	12.00
0.5	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	19.30	16.50	19.30	9.20	8.50	9.40	12.60	11.20	11.20	23.40	27.10	23.70	11.30	8.40	11.90	11.90	11.90	12.00
0.7	0	$\sqrt{\mu_Y}\sigma_Y$	9.60	10.70	9.20	9.70	10.30	10.90	10.20	10.40	10.40	10.40	9.90	9.40	10.30	10.30	7.90	10.30	10.30	10.10
0.7	0	$2\sqrt{\mu_Y}\sigma_Y$	9.80	10.80	9.80	9.30	10.90	9.70	10.30	10.70	10.70	9.20	9.80	10.90	8.70	7.70	10.90	9.60	11.20	11.20
0.7	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	37.20	39.80	40.00	10.10	9.10	9.10	9.00	11.10	11.10	62.90	61.70	62.30	8.70	8.70	8.80	9.40	11.20	11.20
0.7	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	17.20	17.80	17.30	9.90	8.40	10.20	11.70	11.10	11.10	27.20	29.70	24.40	8.80	9.50	9.90	9.90	10.10	8.40

Table presents empirical size of Test 2 (extrapolative) with hypotheses (11) (H_0^{bias}), (12) (H_0^{model}) and (18) (H_0^{adapt}). It is based on 1000 simulations and asymptotic critical values. 10% nominal size is assumed. T denotes the sample size.

Table 10: Power of Test 4 (adaptive) for $T = 100$

The two outer column groups correspond to $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$ and $\sigma_{\xi,h} = \frac{2(h-1)}{\tau}\sqrt{\mu_Y}\sigma_Y$. Within each, columns are grouped under H_0^{bias}, H_0^{model}, H_0^{adapt}.

N	α	μ_ψ	σ_ψ	H_0^{bias} h=1	h=4	h=8	H_0^{model} h=1	h=4	h=8	H_0^{adapt} h=1	h=4	h=8	H_0^{bias} h=4	h=8	H_0^{model} h=4	h=8	H_0^{adapt} h=4	h=8
100	0.2	0	0	0.20	0.50	0.80	100.00	99.60	99.80	99.90	100.00	100.00	0.10	4.80	98.70	100.00	100.00	97.10
100	0.2	0	$\sqrt{\mu_Y}\sigma_Y$	2.50	2.70	3.30	100.00	99.90	100.00	98.10	99.00	99.60	1.70	6.90	99.80	100.00	99.80	86.50
100	0.2	0	$2\sqrt{\mu_Y}\sigma_Y$	10.20	7.00	6.80	100.00	100.00	100.00	86.70	93.00	90.50	6.10	7.80	100.00	100.00	93.90	61.80
100	0.2	$0.1\mu_Y$	0	13.50	12.00	13.30	100.00	100.00	100.00	97.90	99.20	99.30	13.40	11.10	99.90	100.00	99.40	85.60
100	0.2	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	9.00	10.30	10.20	100.00	99.70	99.70	99.70	100.00	100.00	9.70	13.60	98.60	100.00	100.00	96.00
100	0.2	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	14.30	15.70	13.60	100.00	100.00	100.00	83.90	90.00	89.40	14.00	13.80	100.00	100.00	94.00	56.90
100	0.5	0	0	2.30	2.00	1.20	94.20	87.00	86.70	87.00	94.40	94.60	1.30	6.00	79.80	99.30	95.40	76.60
100	0.5	0	$\sqrt{\mu_Y}\sigma_Y$	7.20	4.30	5.10	96.50	90.80	91.60	63.40	76.10	76.30	4.70	7.60	85.00	99.30	74.10	55.90
100	0.5	0	$2\sqrt{\mu_Y}\sigma_Y$	10.70	7.50	7.40	97.10	95.50	95.90	38.90	48.70	47.60	7.10	8.40	90.40	99.90	45.40	34.80
100	0.5	$0.1\mu_Y$	0	22.80	20.70	20.80	95.80	93.50	91.30	62.80	74.50	77.50	22.60	12.40	85.50	99.60	78.00	53.60
100	0.5	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	29.90	29.70	27.40	96.40	89.60	87.20	85.10	94.20	96.30	33.90	17.80	82.10	99.60	95.50	73.70
100	0.5	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	14.50	17.70	13.80	97.10	95.60	95.80	37.40	46.10	49.70	15.30	14.00	90.40	99.90	47.80	31.30
100	0.7	0	0	4.70	2.80	3.90	63.80	50.10	46.70	58.30	64.50	68.00	3.00	6.30	48.50	79.30	67.20	43.60
100	0.7	0	$\sqrt{\mu_Y}\sigma_Y$	6.60	6.20	7.00	59.30	49.10	52.20	31.10	38.80	38.10	7.00	8.20	42.20	80.40	38.60	29.60
100	0.7	0	$2\sqrt{\mu_Y}\sigma_Y$	11.50	9.80	6.90	58.90	54.60	53.20	18.70	21.90	22.70	8.70	9.30	46.20	80.30	21.60	18.80
100	0.7	$0.1\mu_Y$	0	26.90	24.30	26.60	59.60	53.10	54.30	28.20	39.30	39.50	24.60	13.30	46.80	82.80	36.80	29.10
100	0.7	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	44.30	42.30	38.10	64.80	51.30	52.80	57.60	61.80	64.10	53.10	18.70	45.80	79.40	66.70	41.80
100	0.7	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	17.80	15.60	13.70	60.70	56.90	55.90	17.60	19.90	22.90	15.10	11.90	46.70	82.80	22.60	18.00

Table presents empirical power of Test 4 (adaptive) with hypotheses (11) (H_0^{bias}), (12) (H_0^{model}) and (18) (H_0^{adapt}). It is based on 1000 simulations and asymptotic critical values. Results for sample size $T = 100$.

Table 11: Power of Test 3 (acceleration) - estimation with no intercept

| | | | $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$ | | | | | | | | | $\sigma_{\xi,h} = \frac{2(h-1)}{7}\sqrt{\mu_Y}\sigma_Y$ | | | | | |
| | | | $T=50$ | | | $T=100$ | | | $T=1000$ | | | $T=50$ | | $T=100$ | | $T=1000$ | |
α	μ_ψ	σ_ψ	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=1	h=4	h=1	h=4
-0.1	0	0	99.10	99.40	99.50	100.00	100.00	100.00	100.00	100.00	100.00	97.30	100.00	100.00	100.00	100.00	100.00
-0.1	0	$\sqrt{\mu_Y}\sigma_Y$	99.80	99.90	99.90	100.00	100.00	100.00	100.00	100.00	100.00	99.50	100.00	100.00	100.00	100.00	100.00
-0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
-0.1	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	100.00	99.90	99.80	100.00	100.00	100.00	100.00	100.00	100.00	99.70	100.00	100.00	100.00	100.00	100.00
-0.1	$0.1\mu_Y$	0	98.80	98.90	99.40	100.00	100.00	100.00	100.00	100.00	100.00	98.20	100.00	100.00	100.00	100.00	100.00
-0.1	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.1	0	0	99.80	99.50	99.70	100.00	100.00	100.00	100.00	100.00	100.00	99.30	100.00	100.00	100.00	100.00	100.00
0.1	0	$\sqrt{\mu_Y}\sigma_Y$	99.90	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.1	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	99.90	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.1	$0.1\mu_Y$	0	99.60	99.90	99.80	100.00	100.00	100.00	100.00	100.00	100.00	99.30	100.00	100.00	100.00	100.00	100.00
0.1	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table presents empirical power of Test 3 (acceleration) with hypothesis (17) (H_0^{acc}) for estimation without intercept. It is based on 1000 simulations and asymptotic critical values.

the test size is highly similar to the nominal size for a short sample and large variance of the measurement error. Even though we too often reject the true null hypothesis at a sample consisting of 1000 elements, the aggregated test without intercept has considerably better properties.

Table 12: Size of Test 3 (acceleration) - estimation with no intercept

			$T = 50$			$T = 100$			$T = 1000$		
α	μ_ψ	σ_ψ	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8
-0.1	0	$\sqrt{\mu_Y}\sigma_Y$	0.50	2.40	3.10	1.60	5.30	4.90	60.30	72.60	72.60
-0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	7.70	11.90	12.30	21.00	30.00	26.50	99.30	99.40	99.90
-0.1	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	0.50	2.30	2.40	0.80	3.70	2.90	49.50	62.80	65.20
-0.1	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	4.10	8.50	8.60	15.00	22.20	22.20	98.70	99.10	99.10
0.1	0	$\sqrt{\mu_Y}\sigma_Y$	1.20	2.30	3.70	2.80	7.80	7.10	79.50	89.10	88.40
0.1	0	$2\sqrt{\mu_Y}\sigma_Y$	15.70	16.70	14.10	34.00	37.40	40.20	100.00	99.90	100.00
0.1	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	0.30	2.10	2.60	2.10	4.40	5.00	68.20	81.60	81.00
0.1	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	10.70	14.60	10.10	25.30	31.80	30.20	100.00	99.50	99.90

Table presents empirical size of Test 3 (accaleration) with hypothesis (17) (H_0^{acc}) for estimation without intercept. It is based on 1000 simulations and asymptotic critical values. 10% nominal size is assumed. T denotes the sample size.

Table 13 presents results for the mean-reversion scheme. At first, let us note that this is the only scheme with a different equation structure for H_0^{bias} and H_0^{model} tests that have a non-differentiated variable on the left. As regards both these tests, the size is similar as the nominal 10%. However, we should pay attention to the fact that it increases alarmingly with an increase of the sample size. The probability of rejecting a true null hypothesis also increases with an increase of an adjustment coefficient α. The power of both these tests is incredibly low and does not exceed 20% even for a sample of 1000 elements (detailed results are available on request). The H_0^{rev} test size indicates that it nearly always rejects the null hypothesis, regardless of the forecast error. Similar as with the previous scheme, I will check whether removal of intercept from the estimated equation will improve the properties of the aggregated test. Tables 14 and 15 present those results for H_0^{model} and H_0^{rev} tests. Obviously, the removal of intercept implies that the test size will increase in a situation of a non-zero mean measurement error. Indeed, the results for the H_0^{model} test indicate this. Let us note that the properties of the aggregated test have considerably improved, and its size is basically similar as the nominal one, regardless of the measurement error form and the forecast error. Unfortunately, the removal of intercept has triggered a considerable decline in the power of both tests and indicates that they should not be used at large.

Table 13: Size of Test 5 (mean-reversion)

α	μ_ψ	σ_ψ	$T=50$									$T=100$								
			H_0^{bias}			H_0^{model}			H_0^{rev}			H_0^{bias}			H_0^{model}			H_0^{rev}		
			h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8
0.2	0	$\sqrt{\mu_Y}\sigma_Y$	9.70	11.30	9.30	9.50	10.70	10.50	100.00	100.00	100.00	12.20	9.50	9.90	11.80	11.20	11.90	100.00	100.00	100.00
0.2	0	$2\sqrt{\mu_Y}\sigma_Y$	10.20	9.30	11.50	11.40	10.60	12.10	91.60	91.50	90.60	9.10	10.60	9.80	8.70	10.20	10.60	99.70	99.80	100.00
0.2	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	9.00	10.20	10.90	11.40	10.00	10.20	100.00	100.00	100.00	9.40	10.60	11.30	9.50	10.30	11.10	100.00	100.00	100.00
0.2	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	10.10	10.00	10.20	9.40	8.90	9.80	89.30	89.00	87.70	8.60	10.60	10.50	9.50	10.00	11.70	99.50	99.40	99.50
0.5	0	$\sqrt{\mu_Y}\sigma_Y$	7.60	11.40	9.90	9.40	11.00	10.30	100.00	100.00	100.00	10.30	9.40	12.30	10.50	9.00	11.30	100.00	100.00	100.00
0.5	0	$2\sqrt{\mu_Y}\sigma_Y$	10.20	10.30	9.50	11.00	11.30	8.60	90.20	91.90	90.30	10.50	12.10	12.40	9.30	10.20	11.60	99.80	100.00	99.60
0.5	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	8.80	10.90	10.70	9.00	12.40	11.30	99.90	100.00	100.00	10.20	10.80	11.30	10.60	10.90	10.00	100.00	100.00	100.00
0.5	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	10.10	11.20	9.80	8.90	9.70	10.30	84.30	84.20	86.40	11.20	10.30	10.00	11.20	10.40	9.80	99.30	99.30	99.60
0.7	0	$\sqrt{\mu_Y}\sigma_Y$	9.80	9.10	13.50	9.60	10.00	12.50	99.90	100.00	100.00	12.40	11.50	12.30	11.10	12.70	12.10	100.00	100.00	100.00
0.7	0	$2\sqrt{\mu_Y}\sigma_Y$	10.80	9.50	8.60	10.50	9.00	9.00	90.20	91.70	90.20	12.30	13.30	10.50	11.30	10.60	12.20	99.50	99.80	99.80
0.7	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	10.70	11.00	11.00	11.20	10.70	10.80	99.80	100.00	100.00	11.90	12.50	16.80	12.70	12.70	14.30	100.00	100.00	100.00
0.7	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	12.00	8.60	10.60	10.70	8.80	11.00	79.90	79.60	77.70	12.40	13.80	14.70	13.00	13.80	12.60	97.70	96.90	97.10

Table presents empirical size of Test 5 (mean-reversion) with hypotheses (11) (H_0^{bias}), (12) (H_0^{model}) and (19) (H_0^{rev}). It is based on 1000 simulations and asymptotic critical values. 10% nominal size is assumed. T denotes the sample size. The table continues on the next page.

Table 13: Size of Test 5 (mean-reversion) cont.

			$T = 1000$								
			H_0^{bias}			H_0^{model}			H_0^{rev}		
α	μ_ψ	σ_ψ	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8
0.2	0	$\sqrt{\mu_Y}\sigma_Y$	11.50	10.10	8.40	12.30	9.20	9.40	100.00	100.00	100.00
0.2	0	$2\sqrt{\mu_Y}\sigma_Y$	12.90	10.60	11.30	11.00	10.50	10.30	100.00	100.00	100.00
0.2	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	11.00	13.10	11.20	9.40	9.80	8.80	100.00	100.00	100.00
0.2	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	9.50	11.70	10.30	10.10	11.20	9.70	100.00	100.00	100.00
0.5	0	$\sqrt{\mu_Y}\sigma_Y$	15.00	15.90	15.00	14.80	14.20	14.80	100.00	100.00	100.00
0.5	0	$2\sqrt{\mu_Y}\sigma_Y$	14.90	15.20	16.20	13.50	14.00	13.60	100.00	100.00	100.00
0.5	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	17.30	17.20	15.50	12.70	15.30	13.10	100.00	100.00	100.00
0.5	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	15.70	18.00	14.50	14.00	15.80	12.90	100.00	100.00	100.00
0.7	0	$\sqrt{\mu_Y}\sigma_Y$	26.70	26.70	24.20	23.50	23.80	22.70	100.00	100.00	100.00
0.7	0	$2\sqrt{\mu_Y}\sigma_Y$	24.50	25.10	27.50	22.20	23.40	25.10	100.00	100.00	100.00
0.7	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	27.40	27.40	27.00	22.20	21.60	22.30	100.00	100.00	100.00
0.7	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	28.30	25.60	25.70	24.20	21.00	21.60	100.00	100.00	100.00

Table 14: Size of Test 5 (mean-reversion) - estimation with no intercept

| | | | T = 50 | | | | | | T = 100 | | | | | | T = 1000 | | | | | |
| | | | H_0^{model} | | | H_0^{rev} | | | H_0^{model} | | | H_0^{rev} | | | H_0^{model} | | | H_0^{rev} | | |
α	μ_ψ	σ_ψ	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8
0.2	0	$\sqrt{\mu_Y}\sigma_Y$	13.00	14.60	12.90	8.90	9.20	9.50	14.90	14.30	17.00	9.40	10.30	9.80	14.40	16.30	14.80	10.30	9.50	10.00
0.2	0	$2\sqrt{\mu_Y}\sigma_Y$	11.90	12.80	16.60	9.00	8.80	9.30	12.40	14.70	13.70	9.40	7.50	11.60	13.00	16.70	13.10	10.00	9.30	8.40
0.2	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	47.10	47.60	44.10	10.60	9.20	10.10	69.90	69.60	67.00	9.70	10.70	10.00	100.00	100.00	100.00	11.30	10.00	8.50
0.2	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	25.50	20.30	23.90	10.90	8.80	7.70	33.20	31.60	29.80	9.00	11.50	10.50	95.90	95.30	95.30	10.20	9.50	10.20
0.5	0	$\sqrt{\mu_Y}\sigma_Y$	29.70	28.10	31.40	8.70	13.40	14.30	33.90	36.00	35.90	12.00	13.20	14.30	33.60	31.00	33.70	9.40	11.30	10.70
0.5	0	$2\sqrt{\mu_Y}\sigma_Y$	28.50	28.80	32.90	10.80	8.80	10.40	32.90	36.40	34.50	9.60	10.60	12.20	33.60	32.40	31.90	10.90	9.70	10.00
0.5	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	70.90	68.20	67.90	9.10	10.30	13.60	85.80	86.90	86.00	9.90	14.00	13.00	100.00	100.00	100.00	10.70	11.20	10.10
0.5	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	44.90	46.40	41.60	12.10	10.00	12.60	53.50	58.40	55.00	10.60	11.00	10.70	99.20	98.70	99.30	11.10	9.60	8.90

Table presents empirical size of Test 5 (mean-reversion) with hypotheses (12) (H_0^{model}) and (19) (H_0^{rev}) for estimation without intercept. It is based on 1000 simulations and asymptotic critical values. 10% nominal size is assumed. T denotes the sample size.

Table 15: Power of Test 5 (mean-reversion) - estimation with no intercept for $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$

			$T=50$						$T=100$						$T=1000$					
			H_0^{model}			H_0^{rev}			H_0^{model}			H_0^{rev}			H_0^{model}			H_0^{rev}		
α	μ_ψ	σ_ψ	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8
0.2	0		9.60	8.60	11.30	8.80	9.00	10.60	9.30	7.90	10.00	10.70	9.90	10.20	10.90	8.40	8.90	10.40	9.80	10.40
0.2	0	$\sqrt{\mu_Y}\sigma_Y$	12.20	10.30	13.80	8.00	10.00	9.00	11.70	10.20	13.90	11.60	10.80	9.10	10.10	13.00	9.90	9.40	10.10	9.50
0.2	0	$2\sqrt{\mu_Y}\sigma_Y$	10.60	12.10	12.80	8.70	9.50	10.40	13.40	14.40	10.50	9.20	10.10	10.60	12.80	12.50	12.50	10.40	10.40	10.10
0.2	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	19.40	18.30	20.10	8.50	10.20	8.20	27.00	29.70	28.20	8.50	11.10	8.90	97.20	97.30	96.00	11.20	8.60	10.20
0.2	$0.1\mu_Y$	0	34.40	27.40	31.60	8.40	10.70	9.90	53.60	49.20	48.50	9.90	10.60	8.50	100.00	100.00	100.00	10.40	10.30	9.60
0.2	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	15.90	14.20	17.00	9.30	9.60	10.40	19.70	19.30	15.80	11.10	10.20	8.90	67.70	68.50	65.30	9.50	9.70	10.20
0.5	0		9.80	8.80	12.40	11.40	8.90	10.10	9.40	11.30	13.20	10.30	9.00	11.80	9.80	9.70	9.50	10.10	10.30	9.40
0.5	0	$\sqrt{\mu_Y}\sigma_Y$	11.30	13.10	13.80	9.60	9.70	10.70	13.90	13.50	13.90	8.60	10.70	8.60	13.80	14.40	14.60	10.00	10.60	9.50
0.5	0	$2\sqrt{\mu_Y}\sigma_Y$	16.30	14.50	15.80	8.80	10.70	10.50	16.00	17.10	16.70	10.30	9.60	9.80	17.10	14.90	15.60	10.20	12.40	9.30
0.5	$0.1\mu_Y$	$\sqrt{\mu_Y}\sigma_Y$	22.20	22.30	25.30	9.60	9.30	9.70	34.00	32.50	31.80	9.40	9.60	9.30	98.00	98.70	97.50	9.60	11.10	10.30
0.5	$0.1\mu_Y$	0	31.30	29.10	32.80	8.10	10.00	10.50	49.70	50.80	49.90	10.40	10.10	10.30	100.00	100.00	100.00	8.20	10.30	10.00
0.5	$0.1\mu_Y$	$2\sqrt{\mu_Y}\sigma_Y$	17.90	17.80	19.50	9.00	8.80	9.80	24.70	24.80	22.80	7.70	9.40	11.10	75.80	76.80	74.90	10.10	10.10	9.50

Table presents empirical power of Test 5 (mean-reversion) with hypotheses (12) (H_0^{model}) and (19) (H_0^{rev}) for estimation without intercept. It is based on 1000 simulations and asymptotic critical values. T denotes the sample size.

Figure 1: Distribution of $\hat{\alpha}$ when testing H_0^{model} for the acceleration model (h=4, $\alpha = 0.3$)
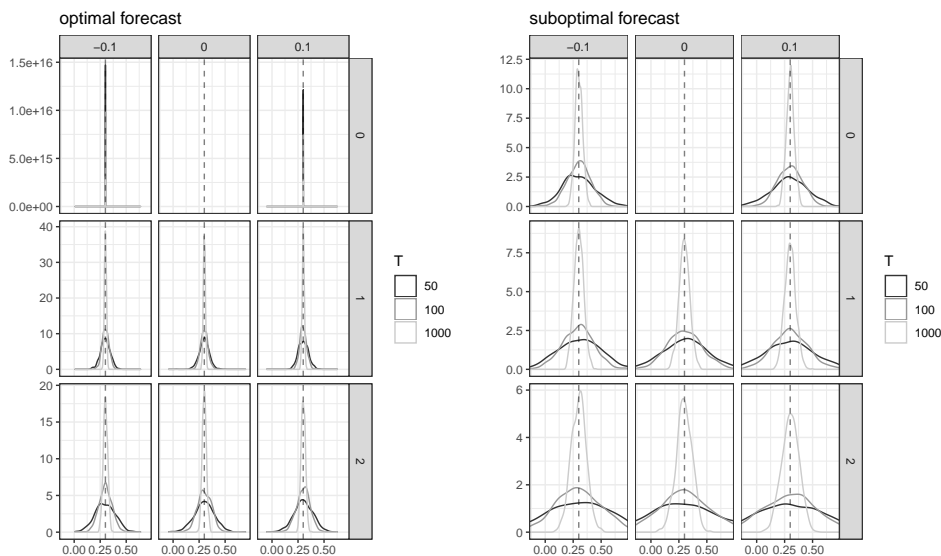
Plots show the distribution of $\hat{\alpha}$ when testing H_0^{model} for the acceleration model across 1000 Monte Carlo samples. Columns correspond to the level of mean measurement error ($\mu_\psi \in \{0, -0.1\mu_Y, 0.1\mu_Y\}$), rows to its standard deviation ($\sigma_\psi \in \{0, \sqrt{\mu_Y}\sigma_Y, 2\sqrt{\mu_Y}\sigma_Y\}$). Suboptimal forecast with $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$ is assumed.

5 Polish macroeconomic forecasts

This section aims at analysing the expectation formation process for the set of Polish macroeconomic data prepared by forecasters of large research centres. To this end, a new set of forecasts of basic variables has been prepared based on two international sources: the European Commission (EC) and the Organisation for Economic Cooperation and Development (OECD) and the National Bank of Poland (NBP). The data cover the forecasts of annual growth rates of national accounts and inflation as well as harmonised unemployment rate for the period between as early as 1998 and end-2020. A detailed description of that data set is included in the Appendix.

Based on the results of simulations, we can arrive at a conclusion that two rationality tests have the best properties: a test based on an optimal regression (PT test) and the CG test; as well as tests of adaptive scheme (H_0^{model}) and acceleration scheme (H_0^{acc} for estimation without intercept). The H_0^{model} test requires familiarity with the

Paulina Ziembińska

adaptive process, which remains unknown when we have only expectations data at our disposal. Therefore, it is necessary to apply the H_0^{adapt} test, which has lower power in case of the longer forecast horizon and a more variable measurement error. In a situation where there are no grounds for rejecting the model based on regression (14),

Table 16: Expectations formation tests for the Polish macroeconomic forecasts

Forecast					Rationality tests			Acceleration test		Adaptive test	
Source	Variable	N	PT	PT proxy	CG		adj.R2	H_0^{acc} pvalue	adj.R2	H_0^{adapt} pvalue
EC	Exports nsa yy	71	0	0.00	0.5		58.54	0.00	-0.61	0.00
EC	GDP deflator	61	0	0.00	0.29		83.49	0.00	6.62	0.00
EC	GDP nsa yy	78	1	1.00	0.09	*	82.38	0.00	20.50	0.00
EC	GFCF nsa yy	71	1	0.00	0.05	*	42.12	0.00	-0.35	0.00
EC	Imports nsa yy	71	0	0.00	0.61	*	43.36	0.00	0.41	0.00
EC	Private cons. nsa yy	71	0	0.00	0.43		81.87	0.00	20.89	0.00
EC	Public cons. nsa yy	58	0	0.00	0		77.13	0.00	32.06	0.00
EC	Harm. unemployment rate	60	0	0.00	0.1	**	98.31	0.00	80.25	0.21
NBP	CPI yy (avg)	69	1	1.00	0.8		72.83	0.00	31.21	0.00
NBP	Exports nsa yy	77	1	1.00	0.36	*	54.44	0.00	-1.93	0.00
NBP	GDP nsa yy	77	1	1.00	0.15	*	83.15	0.00	6.28	0.00
NBP	GFCF nsa yy	77	1	1.00	0.13		24.16	0.00	0.35	0.00
NBP	Imports nsa yy	77	1	1.00	0.16	*	31.54	0.00	0.44	0.00
NBP	Private cons. nsa yy	77	1	1.00	0.02		84.58	0.00	9.01	0.00
NBP	Public cons. nsa yy	74	1	1.00	0.11	*	76.49	0.00	29.39	0.00
NBP	Harm. unemployment rate	80	1	1.00	0		96.09	0.00	58.76	0.01
OECD	CPI yy (avg)	58	1	1.00	0.52		64.70	0.00	-3.66	0.00
OECD	Exports nsa yy	76	1	1.00	0.87		82.07	0.00	13.44	0.00
OECD	GDP deflator	69	1	1.00	0.1		88.64	0.00	65.50	0.41
OECD	GDP nsa yy	76	1	1.00	0.43		86.11	0.00	23.96	0.00
OECD	GFCF nsa yy	76	1	1.00	0		37.70	0.00	-1.79	0.00
OECD	Imports nsa yy	76	1	0.53	0.15	*	65.07	0.00	0.93	0.00
OECD	Private cons. nsa yy	76	1	1.00	0.49		88.19	0.00	38.75	0.00
OECD	Public cons. nsa yy	38	1	1.00	0	**	70.83	0.00	30.64	0.00
OECD	Harm. unemployment rate	41	1	0.00	0.05	*	98.09	0.00	82.88	0.36

Table presents p-values from the rationality tests: PT, PT with proxy and CG. Rationality tests relate to a shorter sample of forecasts prepared up to the end of 2016. For acceleration and adaptive tests adjusted R-square is presented together with the p-values of the respective tests. N denotes number of observations in the estimation. Presented results are calculated for the final readings (i.e. after data revisions). We separately calculate results for the initial readings. Stars denote situations when conclusions for a given test are different for the initial data assuming 0.01 ('***'), 0.5 ('**') and 0.1 ('*') significance level.

it is possible to estimate the adaptive parameter $\hat{\alpha}$. Otherwise, we will be unable to determine its significance. Table 16 presents results of these tests for the analysed set

of Polish macroeconomic data pooled across all forecast horizons. The p-values are presented based on relevant tests for errors calculated against the final reading. I have marked with asterisks situations where a conclusion derived from a specific test would be different when using the initial readings (the number of asterisks corresponds to the level of significance).

First, let us note that as regards the rationality tests the conclusions differ depending on the applied test and that especially the PT test proposed by Patton and Timmermann (2012) provides unambiguous results. When it comes to this test, the revision process also has considerably lower significance – conclusions are identical for the initial and final readings. As regards most analysed variables, the PT test rejects the hypothesis of rationality for forecasts derived from the European Commission, but not for NBP and OECD. In several cases the use of a short-term forecast as a proxy changes the conclusions of the PT test. It is also worth noting that tests applying a proxy do not necessarily yield more conservative results. It reveals a significant role of the revision process in the forecasting process as well as its equivocal character. In particular, it may attest to the fact that the forecasters are not convinced which reading they forecast. The results show that only NBP can potentially treat final readings as a forecasting target – the CG test indicates that for variables forecasted by NBP the rationality hypothesis is more seldom rejected in relation to the final reading. As regards extrapolative models, one should note good fit of acceleration models to the Polish forecasts – the determination coefficient often reaches 70-80%. However, both tests reject the null hypothesis for nearly all variables. An exception is the GDP deflator forecasted by OECD and unemployment rate predicted by the European Commission, but the adaptive coefficient estimated for these predictions is not statistically different from zero (and for the first variable negative), which poses interpretation problems (in general, a negative coefficient would indicate a strong expectation of a trend change).

6   Summary

The work includes a discussion on the expectation formation processes, taking account of the processes of revision, and a broad range of applied tests and their comparative analysis. To the best of my knowledge, this topic has not been considered in this context and has not been tested, not only regarding Poland.

The analysis of properties of the rationality tests has confirmed the results known from the literature, demonstrating that the standard Mincer-Zarnowitz test rejects the null hypothesis too frequently, especially for longer forecasting horizons; tests based on inequalities have smaller size compared to the nominal; a test based on optimal regression proposed by Patton and Timmermann (2012) has the best properties, particularly when we use a forecast with the shortest horizon as a proxy for the actual reading of a variable. Finally, the decreasing mean square forecast test has high power in a situation where the forecast error increases with its horizon. These

conclusions indicate that the test based on optimal regression is the most efficient and flexible. The value add of the presented analysis is an answer to a question of how the tests behave at short samples and in a situation of non-zero mean measurement error arising, for example, from data revisions. In the case of the selected optimal test, the size basically does not change under the influence of sample reduction, similarly as its 100% power at a constant forecast error (in relation to the horizon). In a situation where the forecast error increases with its horizon, the test power declines from over 90% to 60-70% for the shorter horizon and from 80-90% to 40-50% for the longer horizon. Conversely, a non-zero measurement error causes a much more frequent rejection of the null hypothesis, thus the increase of the size of all regression tests, irrespective of an exact form of this error and the forecast horizon, however it does not trigger any significant changes of their power. It is worth noting the asymmetry between the positive and the negative mean error – the latter has smaller impact on the weakening of these tests' properties. What is interesting, a similar relationship is not observed for the discussed extrapolative tests.

Among the extrapolative tests, the adaptive model stands out. The sizes of H_0^{model} and H_0^{adapt} tests are similar as the nominal level, regardless of the sample size, form of the measurement error and forecast horizon. The H_0^{model} test has higher power than the test based on Equation (14) at a longer horizon and at a more variable measurement error for higher adaptive coefficients α. In general, the higher the adaptation of the previous forecast by the executed error, the less frequently we reject the false null hypothesis for both tests.

The proposed tests have worse properties for other extrapolative models. Notably, we often observe an increase in the size or power distortion along with an increase in the sample size, which can question consistency of these tests. In the case of the naïve scheme, the discussed tests have too low a power at a zero mean forecast error to be able to reject the false null hypothesis. When it comes to the extrapolative scheme of the first order, it is worth noting the differences in test properties arising from the sign and scale of the α parameter, which denotes the expectation of trend continuation (positive) or change (negative). The higher the absolute value of the α coefficient (especially for positive values), the worse the statistical properties of the tests. The aggregated test size is often higher than the test power, which indicates that the imposed forecast error increases the estimation uncertainty enough to make it difficult to obtain statistical significance and implies that the test should not be used. As regards the acceleration scheme, we have never rejected the no-bias hypothesis, which results from the test construction. The H_0^{model} test for this scheme has low power, regardless of the sample size and the forecast error considered, but the aggregated H_0^{acc} test demonstrates quite good properties when we remove intercept from the equation – the power of thus formulated test amounts to nearly 100% for all analysed configurations.

Summing up, based on these results, we can conclude that the best properties are showcased by the two rationality tests: based on the optimal regression (PT test)

and the CG tests; as well as the adaptive and acceleration tests (H_0^{model} and H_0^{acc} for estimation without intercept). The H_0^{model} test requires familiarity with the adaptive process, which remains unknown when we have only the survey expectations at our disposal. Therefore, it is necessary to apply the aggregated test with lower power. In the case where there are no grounds for rejecting a model based on regression (14), it is possible to estimate the adaptive parameter $\hat{\alpha}$.

Finally, the selected tests showcasing good properties have been applied to the analysis of the process of forming Polish macroeconomic forecasts from the large research centres. Predictions of the basic variables concerning national accounts, inflation, and unemployment, derived from the OECD, the European Commission and NBP, have been analysed. The conclusions for the rationality tests differ depending on the applied test. For most analysed variables, the PT test rejects the hypothesis of rationality of forecasts derived from the European Commission, but not for the predictions from NBP and OECD. The conclusions are coherent for many variables derived from a certain source, which seems to make sense in a situation where forecasts originate from similar models. Particularly as regards NBP, I rejected the rationality hypothesis more seldom when I compared the predictions to the final reading, which indicates that the national institution may potentially take final readings as the forecast target. Finally, it reveals that the extrapolative models, albeit simple and intuitively interpreted, generally fail to correctly explain the processes of forming the forecasts of the Polish economy.

7   Acknowledgements

References

[1] Arnold E. A., The role of revisions and uncertainty in professional forecasts, (2012), [in:] National Bank of Poland Workshop "Are We Really Forward-Looking?".

[2] Baranowski P., Doryń W., Łyziak T., Stanisławska E., (2021), Words and deeds in managing expectations: Empirical evidence from an inflation targeting economy. *Economic Modelling* 95, 49–67.

[3] Batchelor R., (2001), How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus, *Applied Economics* 33(2), 225–235.

[4] Bordalo P., Gennaioli N., Ma Y., Shleifer A., (2020), Overreaction in macroeconomic expectations, *American Economic Review* 110(9), 2748–82.

[5] Clements M. P., (2019), *Macroeconomic survey expectations*, Springer.

[6] Clements M. P., Galvão A. B., (2019), Data revisions and real-time forecasting, [in:] *Oxford Research Encyclopedia of Economics and Finance.*

[7] Coibion O., Gorodnichenko Y., (2015), Information rigidity and the expectations formation process: A simple framework and new facts, *American Economic Review* 105(8), 2644–78.

[8] Coibion O., Gorodnichenko Y., Kamdar R., (2018), The formation of expectations, inflation, and the Phillips curve, *Journal of Economic Literature* 56(4), 1447–91.

[9] Croushore D., (2011), Frontiers of real-time data analysis, *Journal of Economic Literature* 49(1), 72–100.

[10] Evans G. W., Honkapohja S., (2012), *Learning and expectations in macroeconomics*, Princeton University Press.

[11] Łyziak T., Sheng X., (2018), Disagreement in consumer inflation expectations, Technical report, Narodowy Bank Polski, Economic Research Department.

[12] Mincer J. A., (1969), Economic forecasts and expectations: Analysis of forecasting behavior and performance, *NBER Books.*

[13] Muth J. F., (1961), Rational expectations and the theory of price movements, *Econometrica* (29), 315-–335.

[14] Patton A. J., Timmermann A., (2012), Forecast rationality tests based on multi-horizon bounds, *Journal of Business & Economic Statistics* 30(1), 1–17.

[15] Pesaran M. H., (1985), Formation of inflation expectations in British manufacturing industries, *The Economic Journal* 95(380), 948–975.

[16] Pesaran M. H., Weale M., (2006), Survey expectations, *Handbook of economic forecasting* 1, 715–776.

[17] Timmermann A., (2007), An evaluation of the world economic outlook forecasts, *IMF Staff Papers*, 1–33.

[18] Tomczyk E., (2011), *Oczekiwania w ekonomii. Idea, pomiar analiza*, Oficyna Wydawnicza SGH.

[19] Ziembińska P., (2017), Znaczenie rewizji danych w procesie modelowania, *Wiadomości Statystyczne* 669(2), 5–30.

[20] Ziembińska P., (2021), The importance of data revisions for statistical inference, *Wiadomości Statystyczne. The Polish Statistician* 66(02), 7–24.

Appendix A Additional charts

Figure A.1: Distribution of the intercept in the H_0^{bias} test for the extrapolative model (h=4)
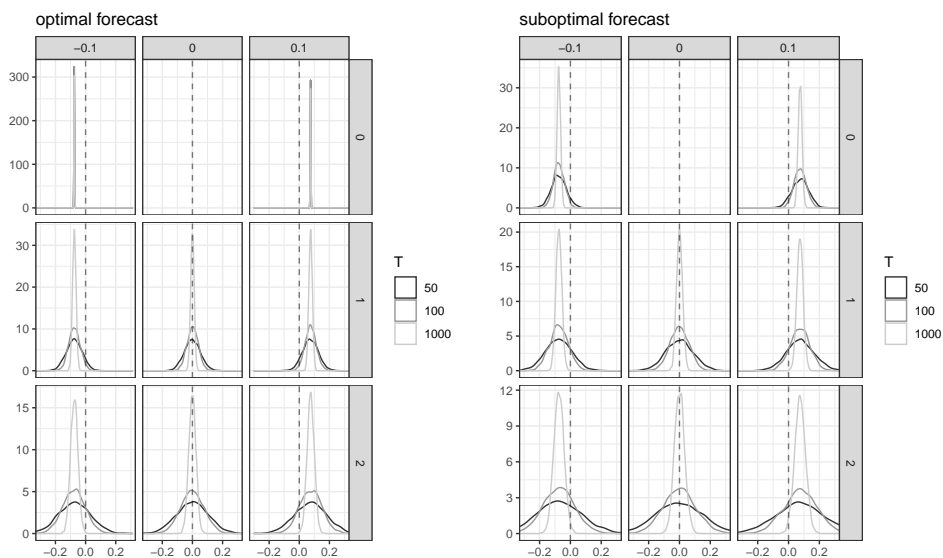
Distribution of the intercept in the H_0^{bias} test for the extrapolative model (h=4) for 1000 Monte Carlo simulations. Columns correspond to the level of mean measurement error ($\mu_\psi \in \{0, -0.1\mu_Y, 0.1\mu_Y\}$), rows to its standard deviation ($\sigma_\psi \in \{0, \sqrt{\mu_Y}\sigma_Y, 2\sqrt{\mu_Y}\sigma_Y\}$). Suboptimal forecast with $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$ is assumed.
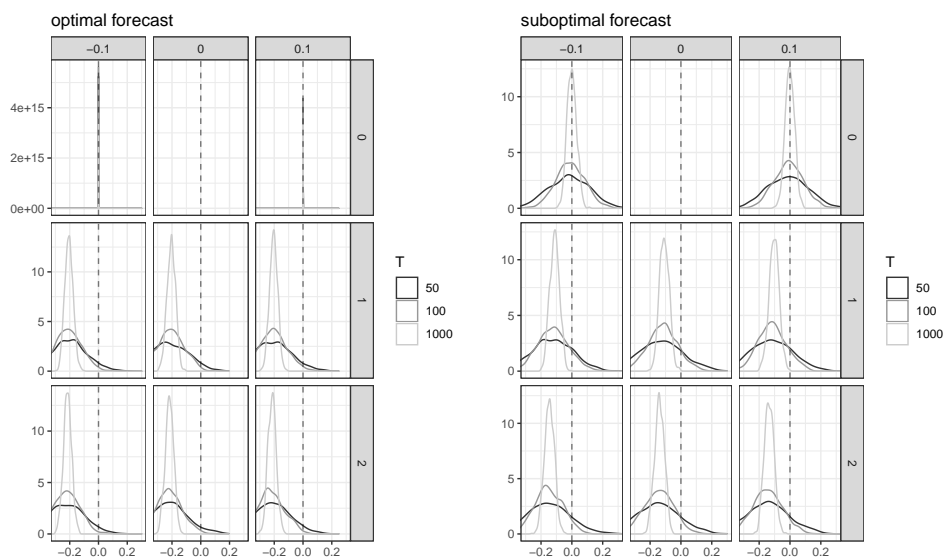
Figure A.2: Distribution of $\hat{\alpha}$ in the H_0^{model} test for the extrapolative model (h=4, $\alpha = 0.3$)
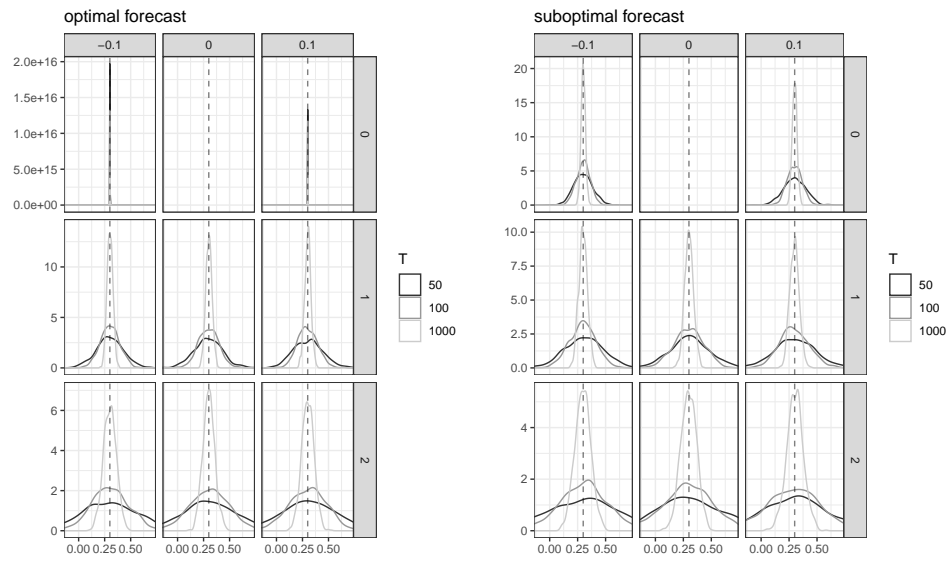
Distribution of $\hat{\alpha}$ in the H_0^{model} test for the extrapolative model (h=4, $\alpha = 0.3$) for 1000 Monte Carlo simulations. Columns correspond to the level of mean measurement error ($\mu_\psi \in \{0, -0.1\mu_Y, 0.1\mu_Y\}$), rows to its standard deviation ($\sigma_\psi \in \{0, \sqrt{\mu_Y}\sigma_Y, 2\sqrt{\mu_Y}\sigma_Y\}$). Suboptimal forecast with $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$ is assumed.
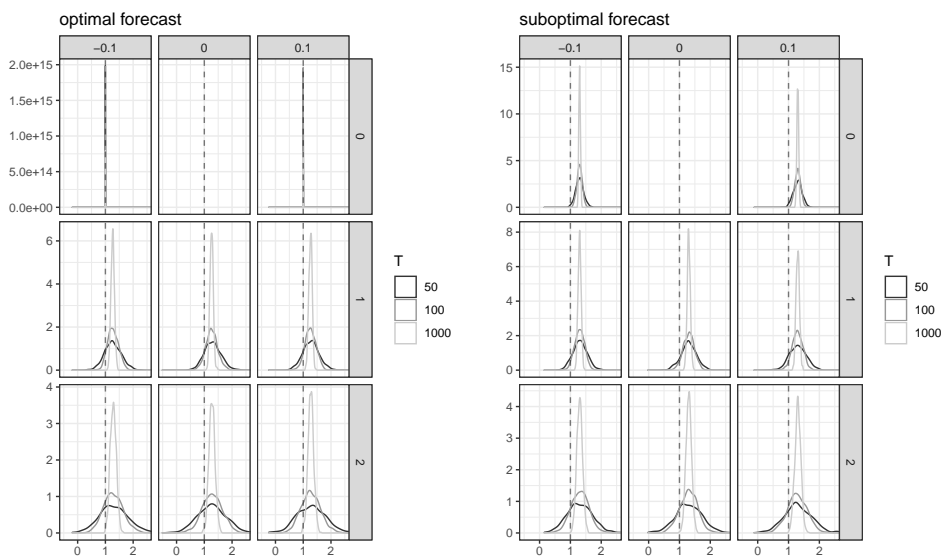
Figure A.3: Distribution of $\hat{const}$ in the H_0^{extra} test (h=4, $\alpha = 0.3$)
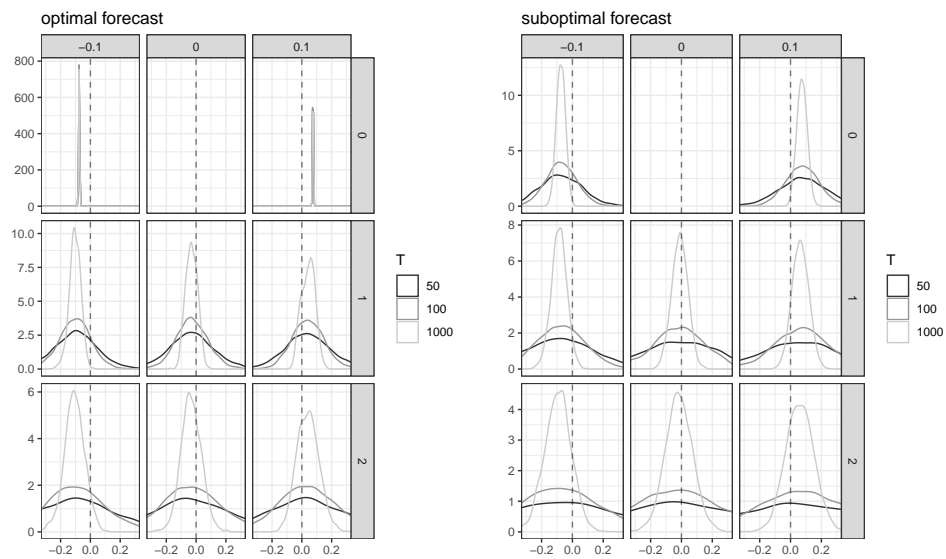
Distribution of $\hat{const}$ in the H_0^{extra} test (h=4, $\alpha = 0.3$) for 1000 Monte Carlo simulations. Columns correspond to the level of mean measurement error ($\mu_\psi \in \{0, -0.1\mu_Y, 0.1\mu_Y\}$), rows to its standard deviation ($\sigma_\psi \in \{0, \sqrt{\mu_Y}\sigma_Y, 2\sqrt{\mu_Y}\sigma_Y\}$). Suboptimal forecast with $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$ is assumed.

Figure A.4: Distribution of $\hat{\beta}$ in the H_0^{extra} test (h=4, $\alpha = 0.3$)

Distribution of $\hat{\beta}$ in the H_0^{extra} test (h=4, $\alpha = 0.3$) for 1000 Monte Carlo simulations. Columns correspond to the level of mean measurement error ($\mu_\psi \in \{0, -0.1\mu_Y, 0.1\mu_Y\}$), rows to its standard deviation ($\sigma_\psi \in \{0, \sqrt{\mu_Y}\sigma_Y, 2\sqrt{\mu_Y}\sigma_Y\}$). Suboptimal forecast with $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$ is assumed.

Figure A.5: Distribution of $\hat{\gamma_1} + \hat{\gamma_2}$ in the H_0^{extra} test (h=4, $\alpha = 0.3$)

Distribution of $\hat{\gamma_1} + \hat{\gamma_2}$ in the H_0^{extra} test (h=4, $\alpha = 0.3$) for 1000 Monte Carlo simulations. Columns correspond to the level of mean measurement error ($\mu_\psi \in \{0, -0.1\mu_Y, 0.1\mu_Y\}$), rows to its standard deviation ($\sigma_\psi \in \{0, \sqrt{\mu_Y}\sigma_Y, 2\sqrt{\mu_Y}\sigma_Y\}$). Suboptimal forecast with $\sigma_{\xi,h} = \sqrt{\mu_Y}\sigma_Y$ is assumed.

Appendix B Set of the Polish macroeconomic forecasts

Table B.1 includes a description of analysed variables. Below a short description of the sources of forecasts used in the analyses is also presented.

Table B.1: List of analysed variables

Variable	Description
GCF nsa yy	Gross capital formation -change (%) in relation to the corresponding period of the previous year (at constant average prices of the previous year)
Export nsa yy	Export -change (%) in relation to the corresponding period of the previous year (at constant average prices of the previous year)
Import nsa yy	Import -change (%) in relation to the corresponding period of the previous year (at constant average prices of the previous year)
Private cons. nsa yy	Private consumption -change (%) in relation to the corresponding period of the previous year (at constant average prices of the previous year)
Public cons. nsa yy	Public consumption -change (%) in relation to the corresponding period of the previous year (at constant average prices of the previous year)
GDP nsa yy	Gross domestic product -change (%) in relation to the corresponding period of the previous year (at constant average prices of the previous year)
CPI yy (avg)	Consumer price index - change (%) in relation to the corresponding period of the previous year
GDP deflator	GDP deflator (%) -based on index measuring the level of prices of all new, domestically produced, final goods and services in an economy in a year.
Harm.unemployment rate (%)	Harmonized unemployment rate (%), monthly data are seasonally adjusted

i) **NBP**

The Economic Institute of the National Bank of Poland prepares forecasts for the purpose of the Monetary Policy Council, assuming the constant NBP interest rate. The projection represents a significant contribution into decisions and communication of the Monetary Policy Council. The projection is prepared three times a year and is published in March, July, and November in the Inflation Report. It contains predictions with a horizon of up to three years. More details can be found on: `https://www.nbp.pl/homen.aspx?f=/en/`

`publikacje/raport_inflacja/projekcja_inflacji.html`. The projection also encompasses probabilistic forecasts presented on radial charts, but they are not directly available. Our set reaches back to 2008.

ii) **European Commission (EC)**
Directorate General of the European Commission prepares economic forecasts for particular countries with a horizon of up to two years and with regard to around 180 variables. The predictions are not derived from a single centralised model but are prepared by teams responsible for particular countries on the basis of models and expert knowledge. They are published three times a year, in line with an internal calendar of the European Commission, known as the *European Semester*. The data have been available since 2003.

iii) **Organisation for Economic Cooperation and Development (OECD)**
Twice a year (in June and December) the Organisation for Economic Cooperation and Development prepares macroeconomic forecasts as part of the *OECD Economic Outlook*. Predictions look up to two years ahead and have been available since 1998. More details concerning the literature utilizing these data can be found in Batchelor (2000) and Timmermann (2007).