Testing Nowcast Monotonicity*

Jack Fosten† Daniel Gutknecht‡

July 28, 2016

Abstract

Nowcasting has become an important tool to many public and private institutions in obtaining timely predictions of low-frequency variables such as Gross Domestic Product (GDP) using current information. Nowcasters often report that a nowcasting method is successful if its predictions improve monotonically as we move towards the publication date of the target variable. In this paper we develop a novel testing approach to formally evaluate the monotonicity of a nowcasting method, in contrast to the existing forecast evaluation literature, which compares the accuracy of different forecast models at a single forecast horizon. In order to highlight the usefulness of this new test, we provide various analytical examples where nowcast monotonicity fails. Formally, our paper extends the methodology of Chernozhukov et al. (2014) for testing many moment inequalities to the case of nowcast monotonicity testing, which allows the number of inequalities to be large relative to the sample size. We show that rolling parameter estimation in the pseudo out-of-sample approach of West (1996) can be accommodated in this setting, and derive new rates on the asymptotic behaviour of the in-sample and out-of-sample splits in relation to the allowed number of nowcast comparisons. We illustrate the performance of our test with a detailed set of Monte Carlo simulations and conclude with an empirical application of nowcasting U.S. real GDP growth and five GDP sub-components. The test reveals that bridge equation methods produce monotonically improving nowcasts of aggregate U.S. GDP growth, whereas for the investment sub-component this is only possible using a forecast averaging scheme.

JEL Classification: C12, C22, C52, C53

Keywords: Nowcasting, Monotonicity, Testing, Moment Inequalities, Bootstrap

---

*We are grateful to Valentina Corradi and Matthias Hartmann for insightful comments and discussions.

†Department of Economics, University of East Anglia, Norwich, NR4 7TJ, UK. E-mail address: j.fosten@uea.ac.uk.

‡Department of Economics, University of Mannheim, 68131 Mannheim, Germany. E-mail address: Daniel.Gutknecht@gmx.de.

1 Introduction

Nowcasting methods have been introduced as a way to provide timely predictions of the current state of the economy, since key variables like GDP are typically published at a low frequency and with a significant publication lag. One of the most desirable features of these methods is that they can be used to revise predictions of GDP multiple times per quarter, as soon as a new piece of information becomes available. Nowcasting methods address the issues of mixed frequency predictors, and also account for the staggered releases of data series by different statistical agencies, known as the 'ragged edge' problem.

However, while there has been a significant body of literature devoted to the development of new nowcasting methods, there has been very little work on how to formally assess their performance. This is particularly surprising as the nowcasting literature often alludes to the trade-off between the timeliness and the precision of available predictor variables when building a nowcasting method, which could lead to fluctuations in nowcast accuracy between different data releases. For example, in nowcasting GDP, Bańbura et al. (2010) note that consumer sentiment indices are more timely but present a worse signal for GDP movements than variables such as industrial production indices. Given this trade-off, we provide analytical examples showing that differing measurement error in macroeconomic series can result in nowcasts which may even worsen as we approach the publication date of GDP.[1] For instance, for a stylized multiple variable bridge equation model often used in nowcasting, it is shown that the mean squared forecast error (MSFE) may not decline when the variable which is released closer to the end of the publication date of the target variable has larger measurement error variance than variables released earlier on. More generally, Patton and Timmermann (2007) demonstrate that (optimal) forecast error variance may not decline if the loss function is asymmetric or the data generating process is nonlinear. It is therefore important to have appropriate methods to detect these changes in nowcast accuracy.

In this paper we propose a novel monotonicity testing procedure to assess whether the accuracy of a nowcasting method improves as new information becomes available. Choosing monotonicity as a performance measure is partially motivated by existing studies in academic papers and policymaking institutions which declare a nowcasting method to be successful if it delivers monotonically improving nowcasts with respect to some loss function each time they are updated throughout the prediction period.[2] However, while past studies have tended to make judgements either based on inspecting graphical evidence of measures such as MSFE, for example Higgins (2014), or using statistical tests

[1] These examples can be found in Appendix A.

[2] Moreover, in the presence of measurement error, monotonicity may also be viewed as an informal measure of the 'marginal gain' of including an additional variable inflicted by measurement error into the nowcast: if the additional noise introduced by that variable is greater than its informational content, nowcasts may not monotonically improve as the publication date is approached.

which are limited to the evaluation of a small number of nowcasts per quarter (Bańbura et al., 2013), our paper aims to provide a formal and robust test for nowcast monotonicity using moment inequalities. In developing this statistical test, we address several non-trivial econometric problems which are new to the literature.

The first issue is that, by tracking the data flow of large economic datasets, nowcasting methods are able to deliver a very large amount of nowcasts per quarter. In our approach, which formulates the null hypothesis of nowcast monotonicity as a set of moment inequalities, this results in the possibility that there is a much larger set of moment inequalities than the quarterly sample size used for model evaluation. In fact, even with a relatively small number of nowcasts per quarter, to fully test for monotonicity involves the use of all possible combinations of pairwise comparisons at different points in the quarter, which may still yield a high-dimensional setting.

Our first contribution addresses this issue by extending the recently developed technique of Chernozhukov et al. (2014) which allows for there to be many more moment inequalities than sample points. We extend their approach, originally geared towards microeconomic applications, to the time series context of nowcast monotonicity testing. We show the validity of Blockwise Multiplier Bootstrap critical values for this test. In doing so, we are able to allow for the number of moment inequalities to grow to infinity with the sample size. Moreover, we formally demonstrate the applicability of a moment selection procedure to eliminate uninformative moment inequalities and to avoid over-conservativeness of the test, and adopt an ad hoc procedure by Zhang and Cheng (2014) to estimate the optimal block length for the BMB. Our method has advantages over the approach followed by Bańbura et al. (2013), who employ a forecast rationality test proposed by Patton and Timmermann (2012), which was not set up to deal with our high dimensional set-up. Other uses of moment inequality-type tests of forecast accuracy testing, such as Corradi and Distaso (2011) who apply the approach of Chernozhukov et al. (2007), are also in the context of a small set of moment inequalities.

The next main issue is that commonly used statistics for measuring nowcast accuracy, such as MSFE, are typically obtained using a sequence of parameter estimates arising from the pseudo out-of-sample approach of West (1996). However, to the best of our knowledge, this type of parameter estimation has not been analysed such a high dimensional context.

The second main contribution is therefore to show the conditions under which rolling parameter estimation error can be ignored in this high-dimensional moment inequality testing context. We derive new rates on the asymptotic behaviour of the in-sample and out-of-sample splits in relation to the number of allowed nowcast comparisons. Under these conditions, testing can proceed as if the true population parameters were known. This result adds a new dimension to previously found conditions where relative rates are derived on the behaviour of only the in-sample and out-of-sample splits, and where the number of model comparisons is fixed and small. In the pairwise case, these results were derived in the seminal paper of West (1996), and in the multiple model case by White (2000) and

Corradi and Swanson (2006).

Our method is strongly complementary to the fast-growing empirical literature of nowcasting, as it can be applied to many different nowcasting methods. An overview of existing empirical nowcasting studies and methodology can be found in the chapter of Bańbura et al. (2013) as well as the recent article of Schumacher (2016). In addition to academic research, nowcasting studies regularly feature in the work of central banks, either in policy documents such as Bell et al. (2014) at the Bank of England, or in Working Papers such as Bańbura et al. (2010) at the European Central Bank. In the United States, some Reserve Banks now use automated systems to publish regular nowcasts of GDP, such as the Federal Reserve Bank of New York's Nowcast Report website[3] and the Federal Reserve Bank of Atlanta's GDPNow model, see Higgins (2014).

This paper is related to two distinct econometric literatures. There is a clear link between this nowcast evaluation method and the existing literature on forecast evaluation. Dating back to Diebold and Mariano (1995) and West (1996), forecast evaluation tests have evolved in many directions: nested model comparisons (Clark and McCracken, 2001, 2005), multiple model comparisons (White, 2000; Romano and Wolf, 2005; Corradi and Swanson, 2006; Hansen et al., 2011), conditional predictive ability testing (Giacomini and White, 2006), and tests involving generated factor regressors (Gonçalves et al., 2015; Fosten, 2016). Our set-up is different to these papers in that our interest is primarily in assessing the accuracy of a single nowcasting method at different points in the quarter, rather than comparing two or models at a fixed point in the quarter. We are also interested in a high-dimensional setting, as mentioned above.

Our approach also bears similarities to monotonicity tests in the applied and theoretical microeconometric literature. Testing for monotonicity in nonparametric regression models has a relatively long history dating back to Hall and Heckman (2000), Ghosal et al. (2000) and, more recently, Ellison and Ellison (2011) and Gutknecht (2016). While all of these are focussed on microeconomic applications, very few studies have looked at monotonicity in the time series context. One example is Patton and Timmermann (2012), which looks from the perspective of testing of forecasters' rationality over a small number of forecast horizons.[4]

We provide detailed Monte Carlo simulations to document how our methodology performs. We then apply our test to nowcasting the aggregate GDP growth rate in the United States, as well as 5 GDP subcomponents, using single-variable bridge equations and also a forecast combination approach for pooling these bridge equation predictions. This builds on several existing studies which we mention in more detail later. As a preview of the results, our test finds that there is no statistical evidence

[3]See: https://www.newyorkfed.org/research/policy/nowcast. [Last Accessed: 15th June 2016]

[4]While the authors examine explicitly only a small number of moment inequalities for adjacent forecast horizons, they point out that many more inequalities could be considered in principle by looking also at horizons spaced further apart. Thus, even in the rather "low dimensional" setting of Patton and Timmermann (2012) there is scope for the testing procedure developed in our paper.

of non-monotonicity in the aggregate GDP growth rate. This is in line with previous studies such as Bańbura et al. (2013) and seems to indicate that one can generate monotonically improving nowcasts of US GDP growth, regardless of the nowcasting method or time period used. On the other hand, the test uncovers evidence of non-monotonicity in the investment sub-component of GDP, whereas monotonicity can be restored by pooling the bridge predictions in a simple equal-weights forecast averaging scheme. This may suggest more generally that nowcasting methods which use forecast averaging are more stable for volatile series like investment.

The rest of the paper is organised as follows. Section 2 describes the set-up of the paper. Section 3 contains an introduction to the the test statistic, defines the bootstrap critical values and presents the asymptotic theory. Section 4 provides the Monte Carlo simulation set-up and results. Section 5 gives the empirical application of the test. Finally, Section 6 concludes the paper. There is also a separate Online Appendix which details some additional results not presented in the main paper or Appendix.

2 Set-up

2.1 Testing Nowcast Monotonicity

The objective is to evaluate nowcasts of a low-frequency target variable, y_t, for which data is available at the time periods $t = 1, ..., T$. Since the vast majority of nowcasting studies use quarterly variables, we will refer to these time periods as quarters in what follows. Data for y_t is published after the end of quarter t. There are S sub-periods per quarter at which the nowcaster observes new information on higher-frequency variables, known as the 'data flow'. We denote this information set $\Omega_{t+i/S}$ for each sub-period $i = 1, ..., S$. At the points $i = 1, ..., S$, the nowcaster uses the current information in $\Omega_{t+i/S}$ to specify a nowcasting model for y_t which uses a set of variables $X_t\,(i)$ and a corresponding set of model parameters θ_i, both of which are of dimension $(n_i \times 1)$.[5] Examples of $X_t(i)$ and θ_i for different nowcasting methods are given below.

The main hypothesis of interest to nowcasters is whether nowcast performance is monotonically improving as we move through the quarter, approaching the publication date of the target variable. In comparing the horizon $i + k$ with some earlier horizon i, we use a loss measure $f_t(\theta_{i+k}, \theta_i)$ which depends on the parameters θ_{i+k} and θ_i and also on the available data. To simplify the notation in the rest of the paper, we will present comparisons of adjacent horizons, $i + 1$ and i, noting that additional pairwise comparisons with $k > 1$ are, of course, available.

This null hypothesis can be written as the set of moment inequalities:

$$H_0 : \mathrm{E}\left[f_t\left(\theta_{i+1}, \theta_i\right)\right] \leq 0 \qquad \text{for all } i = 1, ..., S-1 \tag{1}$$

[5]Typically the variables $X_t\,(i)$ are temporally aggregated to the lower frequency of the target variable from the original higher frequency variables, as described in Examples 1 and 2, below.

versus:

$$H_1 : \mathrm{E}\left[f_t\left(\theta_{i+1}, \theta_i\right)\right] > 0 \qquad \text{for some } i = 1, ..., S-1$$

The null hypothesis is violated when at least one point later in the prediction period has larger loss than some earlier horizon. Since our interest lies in the unconditional mean of a function of unknown parameters, this null hypothesis resembles that in the literature of multiple forecast evaluation, such as White (2000), which are based on the seminal work of West (1996). The main difference is that we explicitly wish to evaluate a single nowcasting method at different horizons, rather than multiple models at a single horizon. We also wish to allow S to be large relative to the sample size.

One of the most commonly used functional forms for $f_t(.)$ in the nowcasting literature is the pairwise comparison of MSFE from linear models. In this case, the function $f_t(.)$ can be written:

$$f_t\left(\theta_{i+1}, \theta_i\right) = u_{i+1,t}^2 - u_{i,t}^2 \tag{2}$$

where $u_{i+1,t} = y_t - X_t\left(i+1\right)' \theta_{i+1}$ and $u_{it} = y_t - X_t\left(i\right)' \theta_i$ are the population forecast errors at time t for nowcast horizons $i+1$ and i.

Since the population parameters θ_{i+1} and θ_i are unknown, we obtain estimates for the set of moment conditions in H_0 using a pseudo out-of-sample nowcasting experiment as in West (1996). We split the sample of time series observations, T, into samples of size R and P. At each of the periods $t = R+1, ..., T$, we make $i = 1, ..., S$ different nowcasts of y_t. For each t and i, we do this by using the latest R observations on the variables $\{X_k\left(i\right)\}_{t=k-R+1}^{t}$ to estimate $\widehat{\theta}_{it}$, noting as above that the specification of $X_t\left(i\right)$ makes use only of the data up to vintage $t+i/S$ in the information set $\Omega_{t+i/S}$. This procedure results in a total of P different nowcasts made at each of the S different nowcast horizons. Therefore, in comparing horizons i and $i+1$, we use sequences of parameter estimates $\{\widehat{\theta}_{it}\}_{t=R+1}^{T}$ and $\{\widehat{\theta}_{i+1,t}\}_{t=R+1}^{T}$. This gives rise to the following estimate of the moment condition:

$$\bar{f}_{i+1,i} = \frac{1}{P}\sum_{t=R+1}^{T} f_t\left(\widehat{\theta}_{i+1,t}, \widehat{\theta}_{it}\right) \qquad \text{for } i = 1, ..., S-1 \tag{3}$$

and so $\bar{f}_{i+1,i}$ is the sample analogue of $\mathrm{E}\left[f_t\left(\theta_{i+1}, \theta_i\right)\right]$. Note that $\{\widehat{\theta}_{it}\}_{t=R+1}^{T}$ and $\{\widehat{\theta}_{i+1,t}\}_{t=R+1}^{T}$ have t subscripts to denote the rolling window used for estimation.

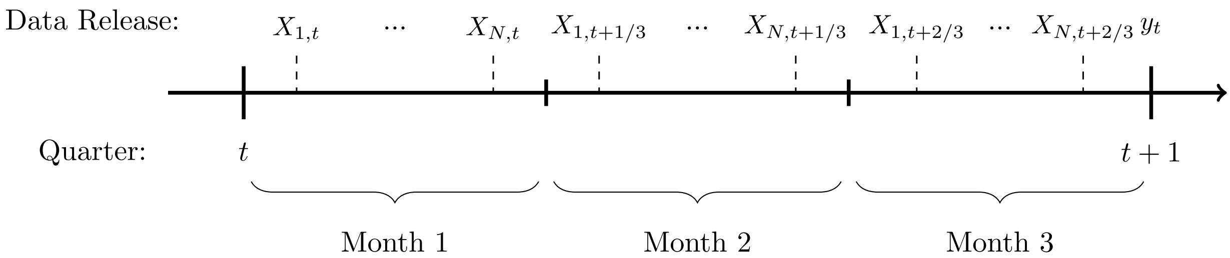
2.2 Nowcasting Set-up and Examples

Before giving some examples of different nowcasting models used in the literature, it is useful to provide an explanation on the data flow which gives rise to the information set $\Omega_{t+i/S}$. In reality, the real-time flow of economic data can be very complex. As described in Bańbura et al. (2013) there could be variables which arrive in the information set at different frequencies (monthly, weekly and daily) and with different publication lags leading to what is known as the 'ragged' or 'jagged' edge. Since the precise form of the data flow is not of paramount importance to this paper, we will present an over-simplified introduction as this does not affect the results which follow.

Suppose there are a set of N potential predictors, X, all of which are measured at the same frequency at F sub-periods within quarter t. We denote a generic observation $X_{i,t+f/F}$ for some variable $i = 1, ..., N$ at a period $t = 1, ..., T$ and sub-period $f = 0, ..., F-1$. Supposing that each of the variables are released at a distinct point throughout the sub-period, then the information set will be incremented a total amount of $S = FN$ times, which corresponds to the number of nowcasts, S, described above. To give a concrete example, in the case where y_t is quarterly and X is monthly then we have $F = 3$, and if there are $N = 10$ variables, the information set expands $S = 30$ times throughout the quarter. Of course, in cases where a group of variables are released at the same time, we would have that $S < FN$.

To elaborate on the timing of the data flow, we assume that the variables are released in the order $i = 1, ..., N$ and, of course, in the sub-period order $f = 1, ..., F$. If these happen at $S = FN$ equal spacings of increments $1/S$, then the information set will expand from $\Omega_t, \Omega_{t+1/S}, ..., \Omega_{t+1-1/S}$ through the addition of the observations $X_{1,t}, X_{2,t}, ..., X_{N,t+(F-1)/F}$ respectively. This is depicted diagrammatically in Figure 1 for the case of monthly variables and a quarterly target, y_t, which is released at the end of the quarter.

Figure 1: Diagrammatic Representation of the Data Flow



The crucial aspect of this data flow which is important to our study is that the number of releases in the quarter may be large. As the number of available economic variables grows large, then so does the number of points at which a nowcaster is able to update her nowcasts throughout the quarter. In our study we will assume an asymptotic set-up where the number of nowcasts, S, is allowed to grow to infinity with the number of predictor variables, N. In other words, $S \longrightarrow \infty$ as $N \longrightarrow \infty$. We think of this as a scenario in line with recent econometric methods of 'big data'.

There have been many parametric model specifications suggested in the nowcasting literature which account for the mixed-frequency and 'ragged edge' problems mentioned above. These are surveyed in Bańbura et al. (2013). We give examples of two of the most commonly used single-model linear nowcasting methods: bridge equations and mixed data sampling (MIDAS) models.

Example 1: Bridge Equations.

Bridge equations are one of the most common nowcasting methods used by central banks (see, for example, Foroni and Marcellino, 2014 and Bulligan et al., 2015). The idea behind bridge equations is to specify a linear model for the low-frequency target variable, y_t, as a function of the high frequency variable aggregated to the lower frequency. For example, suppose y_{T-1} is the last available quarterly observation of GDP growth at quarter $T-1$, and X_T is the last available monthly observation for industrial production growth, which corresponds to the first month of quarter T according to the notation above.[6] Then for all $t = 1, ..., T-1$, the monthly growth series is aggregated to the quarterly frequency using $\overline{X}_t = \sum_{f=0}^{2} X_{t+f/3}$, enabling the specification and estimation of a linear model at the low frequency:

$$y_t = \theta_0 + \theta_1 \overline{X}_t + u_t \text{ for } t = 1, ..., T-1 \tag{4}$$

In order to obtain the nowcast of period T, we require $\overline{X}_T$ but only have X_T available and not $X_{T+1/3}$ and $X_{T+2/3}$. Typically, this 'ragged edge' problem is dealt with using an iterative forecasting model like an $AR(1)$ process, and the monthly forecasts $\widehat{X}_{T+1/3}$ and $\widehat{X}_{T+2/3}$ are obtained. These are aggregated up to get $\overline{\widehat{X}}_T = X_{T+1} + \widehat{X}_{T+1/3} + \widehat{X}_{T+2/3}$ and the nowcast is obtained by:

$$\widehat{y}_T = \widehat{\theta}_0 + \widehat{\theta}_1 \overline{\widehat{X}}_T$$

where typically $\widehat{\theta}_0$ and $\widehat{\theta}_1$ are estimated by OLS. Clearly this requires two steps: a linear model for y_t specified at the lower frequency and estimated using only data up to the end of quarter $T-1$, and an iterative forecast model for the monthly predictor to obtain the prediction $\overline{\widehat{X}}_T$. The idea is that, as more monthly information becomes available, fewer of the monthly observations need to be predicted to obtain the nowcast, and it is hoped that the nowcast becomes more accurate, depending on seasonal effects or measurement error. It is possible to use multiple bridge equations for different variables, or add multiple predictor variables to Equation (4), which may give rise to many different nowcast updates per quarter.

Example 2: MIDAS-type models. Another popular nowcasting methodology is the MIDAS approach of Ghysels et al. (2007) and Clements and Galvão (2008). In contrast to bridge equations, the MIDAS class of models, as noted by Schumacher (2016), aims to keep the predictors at the higher

[6] Examples where monthly information is not available after the last observed quarterly observation are not readily found in the data flow of most countries for typical nowcasting variables and can therefore be ignored. The set-up of Schumacher (2016) makes this clear.

frequency without resorting to time aggregation. This is done using a lag polynomial expressed at the higher frequency, $L^{1/3}$ in the monthly case, where it is understood that $L^{1/3}X_{t+2/3} = X_{t+1/3}$, $L^{2/3}X_{t+2/3} = X_t$, $LX_{t+2/3} = X_{t-1/3}$ and so on. The general MIDAS model is then:

$$y_t = \theta_0 + \theta_1 B(L^{1/3};\beta)X_{t+f/3} + u_t \tag{5}$$

where:

$$B(L^{1/3};\beta) = \sum_{k=0}^{K} b(k;\beta)L^{k/3}$$

In other words, given the latest observed point at the higher frequency, $X_{t+f/3}$ where f could be $0, 1$ or 2, Equation (5) relates y_t directly to the high frequency predictor through the polynomial lag function B, which is parameterised by β. It is therefore possible to specify a different MIDAS model at each of the high frequency (monthly) points $i = 1, 2, 3$. In some sense, the MIDAS approach solves the 'ragged edge' problem by re-aligning the time series so that the variables are shifted forward based on the most recently available data, in contrast to bridge equations where iterative forecasts are first obtained for the monthly variables. As mentioned in Bańbura et al. (2013), this re-alignment gives "a balanced data-set with the most recent information."

Of major importance to MIDAS models is specifying a parsimonious functional form for $B(L^{1/3};\beta)$ so that Equation (5) is not too highly parameterized. Various functional forms have been suggested in the literature. One of the most widely used is the "exponential Almon lag" function (Ghysels et al., 2007) which specifies:

$$b(k;\beta)L^{k/3} = \frac{\exp(\beta_1 k + \beta_2 k^2)}{\sum_{k=0}^{K}\exp(\beta_1 k + \beta_2 k^2)}$$

and can be inserted into Equation (5) and estimated by nonlinear least squares (NLS).

3 Test Statistic and Asymptotic Theory

3.1 Test Statistic

In order to test the null hypothesis H_0, we propose to use the moment inequality testing procedure of Chernozhukov et al. (2014) which allows the number of inequalities to grow with the sample size. When the number of moment inequalities is very large, this may be preferable over the Bonferroni-type procedures used by Bańbura et al. (2013).

The test statistic we propose is a max statistic of the following form:

$$U^* = \max_{1\le i\le \kappa} \sqrt{P}\bar{f}_{i+1,i} \tag{6}$$

where κ is the total number of moment inequalities. As we have written the test for only adjacent horizons, the value of κ is determined by $\kappa = S - 1$. However, clearly it is possible to use a much larger set of up to $\kappa = S(S-1)/2$ inequalities comparing horizon i with $i + k$ for any $k > 0$.[7] We will explore using more than only adjacent horizons in the empirical application.

The null hypothesis of monotonicity, H_0, in Equation (1) is rejected if $U^* > c(\alpha)$ with $c(\alpha)$ denoting a corresponding critical value at significance level α. Rejection of the null implies that there is at least one point later in the quarter which has significantly larger nowcast error loss than at some earlier horizon.

Before describing how to construct these critical values, there is a remark to be made with respect to the existing forecast evaluation literature: in order to apply the approach of Chernozhukov et al. (2014) we need that $v_{i+1,i} > 0$, where $v^2_{i+1,i}$ is the variance of $f_t(\theta_{i+1}, \theta_i)$. Cases where $v_{i+1,i} = 0$ occur in the case of testing equal predictive ability of nested model comparisons as noted by Clark and McCracken (2001, 2005). We rule these out in models such as those in Examples 1 and 2 above, in order to avoid degeneracy of this test statistic. Note that $v_{i+1,i} > 0$ also rules out the use of high-frequency data where the number of sub-periods per quarter is such that $F \to \infty$ (in-fill asymptotics) as the use of these models will yield perfectly correlated forecast errors in the limit. A similar argument holds for factor-based methods where factors are re-estimated after a single data point has been added. The study of nowcast monotonicity using factor models is an important area of future study, but is outside of the scope of the current paper.

3.2 Critical Values

The approach to testing many moment inequalities recently introduced by Chernozhukov et al. (2014) relies on finite sample approximations of the (unknown) asymptotic distribution of the test statistic under H_0 through the use of the BMB procedure. The heuristics of this methodology and of our extension will be outlined after the main theorem, Theorem 1, which establishes the validity of critical values constructed using the BMB.

Before stating the conditions required for Theorem 1, we introduce some further notation and describe the bootstrap algorithm for calculating critical values. Denote $\boldsymbol{\theta}_i = [\theta'_{i+1}, \theta'_i]'$ so that $f_t(\boldsymbol{\theta}_i) = f_t(\theta_{i+1}, \theta_i)$.[8] Likewise, let $f_t(\widehat{\boldsymbol{\theta}}_{it}) = f_t(\widehat{\theta}_{i+1,t}, \widehat{\theta}_{it})$, where $\widehat{\boldsymbol{\theta}}_{it} = [\widehat{\theta}'_{i+1,t}, \widehat{\theta}'_{it}]'$, and let $\boldsymbol{\theta}_i$ be of dimension $(\overline{n}_i \times 1)$ with $\overline{n}_i = n_{i+1} + n_i$, where n_i was defined as the dimension of θ_i and thus typically $\overline{n}_i > 2$. Paralleling Chernozhukov et al. (2014), we employ Bernstein's 'small' and 'large' blocks technique of the BMB.[9] That is, let $q_P > r_P$ denote the 'large' and 'small' blocks respectively and assume

[7] Patton and Timmermann (2012) also set up their test using only adjacent horizons, but do not explore this further.

[8] As mentioned above, the approach can be generalised to allow for more moment inequalities than just the adjacent horizons. In general we can use $f_t(\boldsymbol{\theta}_i) = f_t(\theta_{i+k}, \theta_i)$ for any $k > 0$ and $i = 1, ..., S - k$.

[9] This technique is widely attributed to the paper of Bernstein (1927).

that $q_P + r_P < P/2$. In analogy to their paper, define $I_1 = \{R+1, \ldots, R+q_P\}$, $J_1 = \{R + q_P + 1, \ldots, R + q_P + r_P\}$, ..., $I_{m_P} = \{R + (m_P - 1)(q_P + r_P) + 1, \ldots, R + (m_P - 1)(q_P + r_P) + q_P\}$, $J_{m_P} = \{R+(m_P-1)(q_P+r_P)+q_P+1, \ldots, R+m_P(q_P+r_P)\}$, and $J_{m_P+1} = \{R+m_P(q_P+r_P)+1, \ldots, T\}$. Thus $m \equiv m_P$ defines the number of blocks as the integer part of $m = P/(q_P + r_P)$.[10] Also, define the following two expressions:

$$\overline{\sigma}^2(l) \equiv \max_{1 \le i \le \kappa} \max_I \operatorname{var}\left(l^{-\frac{1}{2}} \sum_{t \in I} f_t(\theta_i) \right)$$

and

$$\underline{\sigma}^2(l) \equiv \max_{1 \le i \le \kappa} \min_I \operatorname{var}\left(l^{-\frac{1}{2}} \sum_{t \in I} f_t(\theta_i) \right),$$

where $\max_I$ and $\min_I$ are taken over all $I \subset \{R+1, \ldots, T\}$ of the form $I = \{t+1, \ldots, t+l\}$. The algorithm of the block multiplier bootstrap is now as follows:

1. Generate standard normal random variables $\varepsilon_1, \ldots, \varepsilon_m$ independent of the data $\{f_t(\widehat{\boldsymbol{\theta}}_i)\}_{t=R+1}^T$.

2. Construct the BMB statistic:

$$W_{BMB} \equiv \max_{1 \le i \le \kappa} \left(\frac{1}{\sqrt{m_P q_P}} \sum_{h=1}^{m_P} \varepsilon_h \sum_{t \in I_h} \left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - \frac{1}{P} \sum_{t=R+1}^{T} f_t(\widehat{\boldsymbol{\theta}}_{it}) \right) \right).$$

3. Calculate $c^{BMB}(\alpha)$ as the conditional $(1-\alpha)$ quantile of W_{BMB} given $\{f_t(\widehat{\boldsymbol{\theta}}_i)\}_{t=R+1}^T$.

The above algorithm has been designed for a fixed large block-small block combination, and size as well as power of the test may be sensitive to the actual choice. As standard selection methods of the block size for dependent data such as Bühlmann and Künsch (1999) or Hall et al. (1995) have not yet been extended to high-dimensional data, we follow and adapt an ad-hoc nested bootstrap procedure suggested by Zhang and Cheng (2014) for the BMB applied to a max statistic to estimate the optimal small-large block combination: first, define a grid (r_P, q_P) of possible small and large block combinations and pick an initial block size for both the small and the large block, say (r_{int}, q_{int}), such that $(r_{int} + q_{int})m_{int} = P$. Conditional on the sample $\{f_t(\widehat{\boldsymbol{\theta}}_i)\}_{t=R+1}^T$, let $s_1, \ldots, s_{m_{int}}$ be i.i.d. uniform random variables on the set $\{0, \ldots, m_{int} - 1\}$ and define $f^*_{(j-1)(r_{int}+q_{int})+i}(\widehat{\boldsymbol{\theta}}_i) = f_{s_j(r_{int}+q_{int})+i}(\widehat{\boldsymbol{\theta}}_i)$ with $1 \le j \le m_{int}$ and $1 \le i \le (q_{int} + r_{int})$. In other words, $\{f^*_t(\widehat{\boldsymbol{\theta}}_i)\}_{t=R+1}^T$ is a block bootstrap sample of the original sample with non-overlapping blocks and initiates the *outer bootstrap* procedure. Since, conditional on the sample, $\overline{f}_{i+1,i}$ is the true mean of this block bootstrap sample, compute the times that the sample mean $\overline{f}_{i+1,i}$ satisfies $P^{\frac{1}{2}} \max_{1 \le i \le \kappa}(\overline{f}^*_{i+1,i} - \overline{f}_{i+1,i}) \le c(\alpha)$ for each grid

[10] That is, $(q_P + r_P)$ characterizes an independent block.

value of (r_P, q_P), where $c(\alpha)$ is computed on the basis of $\{f_t^*(\widehat{\boldsymbol{\theta}}_i)\}_{t=R+1}^T$ repeating steps 1 through 3 from before B_I times (the *inner bootstrap* procedure). Drawing B_O different block bootstrap samples allows to compute the times that $\overline{f}_{i+1,i}$ is contained in the set $\{\overline{f}_{i+1,i} : P^{\frac{1}{2}} \max_{1\leq i\leq\kappa}(\overline{f}^*_{i+1,i} - \overline{f}_{i+1,i}) \leq c(\alpha), i = 1, \ldots, \kappa | \{f_t(\widehat{\boldsymbol{\theta}}_i)\}_{t=R+1}^T\}$ (empirical coverage probability) for each grid choice (r_P, q_P). The pair (r_P, q_P) associated with the empirical coverage probability closest to $1 - \alpha$ can viewed as an estimate of the optimal block choice(s) that captures the dependence structure of the actual data best. This 'optimal choice' can then subsequently be employed to perform the moment inequality test of interest on the original sample.

For the regularity conditions outlined below, let c, c', C, and C' denote generic positive constants that may vary throughout the rest of the paper and that are assumed to depend exclusively on $0 < c_1 \leq C_1 < \infty$ and on $0 < c_2 < 1/4$:

Assumption 1. *For every $i = 1, \ldots, \kappa$, the data $\{f_t(\boldsymbol{\theta}_i)\}_{t=1}^T$ is identically distributed (across t) and β-mixing with the k-th mixing coefficient for $\{f_t(\cdot)\}_{t=1}^T$ defined by:*

$$b_k = b_k(\{f_t(\cdot)\}_{t=1}^T) = \max_{1\leq l\leq T-k} \beta(\sigma(f_1(\cdot), \ldots, f_l(\cdot)), \sigma(f_{l+k}(\cdot), \ldots, f_T(\cdot)), \quad 1 \leq k \leq T-1,$$

where $\sigma(f_t(\cdot), t \in \mathcal{T})$ with $\mathcal{T} \subset \{1, \ldots, T\}$ is the sigma field generated by $f_t(\cdot)$, $t \in \mathcal{T}$.

Assumption 2. *For every $1 \leq i \leq \kappa$, Let $\boldsymbol{\Theta}_i$ be a compact subset of $\mathbb{R}^{\overline{n}}$, where $\overline{n} = \max_{1\leq i\leq\kappa}(\overline{n}_i)$ is finite dimensional. Assume that $f_t(\cdot)$ and $F_{f_t(\boldsymbol{\theta}_i)}(\cdot|\overline{\boldsymbol{\theta}}_i)$, the distribution function of $f_t(\boldsymbol{\theta}_i)$ given $\overline{\boldsymbol{\theta}}_i$, are continuously differentiable for every value $\overline{\boldsymbol{\theta}}_i$ in the interior of $\boldsymbol{\Theta}_i$. Moreover, for every $1 \leq i \leq \kappa$, the following conditions hold:*

$$E\Big[\sup_{\overline{\boldsymbol{\theta}}_i\in\boldsymbol{\Theta}_i} \Big(\nabla_{\boldsymbol{\theta}} F_{f_t(\boldsymbol{\theta}_i)}\Big(f_t(\overline{\boldsymbol{\theta}}_i)\Big|\overline{\boldsymbol{\theta}}_i\Big)\Big)^2\Big] < \infty$$

and

$$E\Big[\sup_{\overline{\boldsymbol{\theta}}_i\in\boldsymbol{\Theta}_i} \Big(\nabla_{\boldsymbol{\theta}} f_t(\overline{\boldsymbol{\theta}}_i)\Big)^2\Big] < \infty.$$

Assumption 3. *Assume that each estimator $\widehat{\boldsymbol{\theta}}_{it}$, $i = 1, \ldots, \kappa$ and $t = R+1, \ldots, T$, admits the following linear representation:*

$$\sqrt{R - q_P - r_P}(\widehat{\boldsymbol{\theta}}_{it} - \boldsymbol{\theta}_i) = \frac{1}{\sqrt{R - q_P - r_P}} \sum_{k=t-R+1}^{t-q_P-r_P} \boldsymbol{\Psi}_{ik} + O_p((R - q_P - r_P)^{-1}),$$

where the influence function $\boldsymbol{\Psi}_{ik}$ is of the same dimension as $\boldsymbol{\theta}_i$ and satisfies $E[\boldsymbol{\Psi}_{ik}] = \mathbf{0}$ as well as $E[\boldsymbol{\Psi}_{ik}\boldsymbol{\Psi}'_{ik}] < \infty$.

Assumption 4. *Assume that:*

$$E\left[\max_{1\le t\le T}\max_{1\le i\le\kappa}\left|f_t(\boldsymbol{\theta}_i)-E\Big[f_t(\boldsymbol{\theta}_i)\Big]\right|^k\right]<C_1 \qquad 1\le k\le 4.$$

Assumption 5. *Let $R, P \longrightarrow \infty$ as $T \longrightarrow \infty$ and suppose that $\kappa \longrightarrow \infty$.*[11] *Moreover, assume that $q_P = o(P)$, $r_P = o(q_P)$ as $P \longrightarrow \infty$ and that there exist a sequence $\zeta_{P1} \longrightarrow 0$ as $P \longrightarrow 0$. We impose the following rate conditions: for c_2 defined above, it holds that $P^{\frac{2c_2+1}{2}}\kappa R^{-1} \longrightarrow 0$, $P^{c_2}\kappa R^{-\frac{1}{2}} \longrightarrow 0$, and:*

$$\max\left\{\frac{C'P^{\frac{(2c_2+1)}{2}}\kappa}{R}, \frac{C'P^{c_2}\kappa}{R^{\frac{1}{2}}}\right\}\le\zeta_{p1}\le\frac{C_1}{P^{c_2}\log\kappa}.$$

Also, $\max\{m_P b_r, (r_P/q_P)\log^2\kappa\}\le C_1P^{-c_2}$ *and* $q_P C\log^{\frac{5}{2}}(\kappa P)\le C_1P^{\frac{1}{2}-c_2}$.

Assumption 6. *Assume that $v^2_{i+1,i}$, the variance of $f_t(\boldsymbol{\theta}_i)$, satisfies $v^2_{i+1,i}>0$ for all i, $1\le i\le\kappa$.*

Assumption 1 characterizes the time dependence in the data, while Assumptions 2 and 3 are required to address parameter estimation error in the proof of Theorem 1. In particular, Assumption 3 states that $\sqrt{R-q_P-r_P}(\widehat{\boldsymbol{\theta}}_{it}-\boldsymbol{\theta}_i)$ is an asymptotically linear statistic with the representation as outlined above (cf. Newey, 1990), which holds true for many von Mises statistics and M-estimators of time series models. Assumption 4 requires the existence and finiteness of certain moment expressions. Note that there are some additional technical conditions associated with Assumption 4, which are required to relax a bounded support assumption made by Chernozhukov et al. (2014) for the case of dependent data for illustrative purposes. These have been relegated to the supplementary material due to their technical nature and to their otherwise minor importance. We will refer to these conditions as Assumption 4* in the following. The first part of Assumption 5 is, again, standard and allows that we have asymptotics both for parameter estimation (R) and nowcast evaluation (P) as in West (1996). The second asymptotic assumption on κ, however, is new to the nowcasting set-up. Using big-data approaches it is possible to generate many nowcasts per quarter and therefore the assumption that $\kappa\to\infty$ allows us to consider cases in which the number of moment inequalities are greater than the sample size. As discussed in Sections 2.2 and 3.1, we think of this growth in the number of nowcasts as a result of a growing number of predictor variables, N, rather than a growth in the number of sub-periods, F. Notice, however, that, in case of $N\longrightarrow\infty$, the variance constraint $v^2_{i+1,i}>0$ for all $1\le i\le\kappa$ from Assumption 6 may only hold true when additional information is contained in every new predictor variable included in the nowcast. Assumption 6 therefore rules out settings where

[11] As stated in the text before, we assume that the number of nowcasts per quarter $S=FN$, which is related to κ in the case of adjacent-only comparisons through $\kappa=S-1$, or in the case where we use the full set of pairwise combinations as $\kappa=S(S-1)/2$. This rate is interpreted such that $S\to\infty$ as $N\to\infty$, but F remains finite, to rule out in-fill asymptotics.

predictor variables are perfectly correlated as N grows large or, similarly, where irrelevant predictors are included. The last part of Assumption 5 governs the interplay of R, P, and the number of moment conditions κ, as well as of the increase of the small and large block size. For instance, for $c_2 = \frac{1}{20}$, Assumption 5 is satisfied when e.g. $P = R$ and $\kappa = R^{\frac{1}{4}}$. Alternatively, if $P = R^{\frac{1}{2}}$, κ could be chosen to be of order $R^{\frac{1}{2}-\epsilon}$ for some $\epsilon > \frac{1}{20}$.[12]

Theorem 1. *Recall the definition of the constants from above and suppose that* $c_1 \leq \overline{\sigma}^2(q_P) \leq \max\{\underline{\sigma}^2(r_P), \overline{\sigma}^2(q_P)\} \leq C_1$. *Moreover, assume that Assumptions 1, 2, 3, 4, 4*, 5, and 6 hold. Then, there exist positive constants c, C such that under* H_0*:*

$$Pr\left(U^* > c^{BMB}(\alpha)\right) \leq \alpha + CP^{-c}, \tag{7}$$

where $c^{BMB}(\alpha)$ *is the corresponding critical value at level* α *from the Block Multiplier Bootstrap procedure described before. If* $E\left[f_t\left(\boldsymbol{\theta}_i\right)\right] = 0$ *for all* $1 \leq i \leq \kappa$*, then:*

$$\left|Pr\left(U^* > c^{BMB}(\alpha)\right) - \alpha\right| \leq CP^{-c}. \tag{8}$$

Theorem 1 yields valid critical values for a test of the moment inequalities described by (1). That is, the theorem states that the BMB procedure yields critical values in such a way that the test has approximately size $\alpha \in (0, 1)$. This is accomplished without referring to the asymptotic distribution of the test statistic, but by basing the proof on finite sample approximations, which hold even when the number of moment inequalities κ is much larger than the sample size P. More specifically, Chernozhukov et al. (2014) show that, under certain regularity conditions, the distribution of $\max_{1\leq i\leq\kappa} \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left(f_t(\boldsymbol{\theta}_i) - \mathrm{E}\left[f_t(\boldsymbol{\theta}_i)\right]\right)$, the statistic evaluated at the true $\boldsymbol{\theta}_i$, can be approximated directly by that of $\max_{1\leq i\leq\kappa} Y_i$ with $Y = (Y_1, \ldots, Y_\kappa)'$ a centered normal random vector with covariance matrix $\mathrm{E}\left[YY'\right] = (1/m_P q_P) \sum_{h=1}^{m_P} \mathrm{E}\left[\left(\sum_{t\in I_h} \boldsymbol{f}_t - \mathrm{E}\left[\boldsymbol{f}_t\right]\right)\left(\sum_{t\in I_h} \boldsymbol{f}_t - \mathrm{E}\left[\boldsymbol{f}_t\right]\right)'\right]$ in the sense of Kolmogorov distance.[13] Since the covariance structure of $\max_{1\leq i\leq\kappa} Y_i$ is unknown, a similar approximation with $\max_{1\leq i\leq\kappa} Y_i$ is also established for the bootstrap statistic described before. The main contribution of the proof of Theorem 1 therefore consists in showing that the parameter estimation

[12] Note that the rates of Assumption 5 may be improved as the proof of Theorem 1 essentially uses the inequality:

$$\max_{1\leq i\leq\kappa} \left|\frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right| \leq \sum_{i=1}^{\kappa} \left|\frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right|.$$

[13] The Kolmogorov distance between the distributions of two random variables X and Y is defined by $\sup_{t\in\mathbb{R}}|\Pr(X \leq t) - \Pr(Y \leq t)|$.

error in the test as well as in the bootstrap statistic is of smaller order than the approximation error derived in Chernozhukov et al. (2014).

The second part of Theorem 1, namely Equation (8), shows that when all moment inequalities are binding, the asymptotic size of the test coincides with the nominal size α. Moreover, as pointed out by Chernozhukov et al. (2014), Equation (8) can be understood as a high dimensional Bootstrap CLT for maxima of (non-studentized) sample averages, which can be approximated by the BMB.

3.3 Critical Values with Moment Inequality Selection

When the number of non-binding and thus uninformative moment inequalities with $\mathrm{E}\left[f_t\left(\boldsymbol{\theta}_i\right)\right] < 0$ is large, one-step critical values described in the previous section may become too conservative.[14] A possible solution to this problem is the application of moment selection procedures which do not consider inequalities in the calculation of the critical values that are unlikely to be binding, i.e. inequalities whose estimated counterpart lies below a certain threshold value. We therefore adopt a two-step method proposed by Chernozhukov et al. (2014) for the Multiplier Bootstrap to the BMB used in this paper. Let $0 < \beta_P < \alpha/2$ be some constant which may depend on the out-of-sample size P. Then let $c^{BMB}(\beta_P)$ analogously denote the one-step critical values calculated using the procedure above, for size β_P rather than α. Finally denote the moment inequality selection set $\widehat{J}_{BMB}$ as:

$$\widehat{J}_{BMB} = \left\{ i \in \{1, ..., \kappa\} : \sqrt{P}\bar{f}_{i+1,i} > -2c^{BMB}(\beta_P) \right\}$$

Then the two-step critical values with moment inequality selection, which we denote $c^{BMB,2S}(\alpha)$, can be calculated using the following procedure:

1. Generate standard normal random variables $\varepsilon_1, \ldots, \varepsilon_m$ independent of the data $\{f_t(\widehat{\boldsymbol{\theta}}_i)\}_{t=R+1}^T$.
2. Construct the BMB statistic:

$$W_{\widehat{J}_{BMB}} \equiv \begin{cases} \max_{i\in\widehat{J}_{BMB}} \left(\frac{1}{\sqrt{m_P q_P}} \sum_{h=1}^{m_P} \varepsilon_h \sum_{t\in I_h} \left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - \frac{1}{P} \sum\limits_{t=R+1}^{T} f_t(\widehat{\boldsymbol{\theta}}_{it}) \right) \right) & \text{if } \widehat{J}_{BMB} \text{ is not empty} \\ 0 & \text{if } \widehat{J}_{BMB} \text{ is empty} \end{cases}$$

3. Calculate $c^{BMB,2S}(\alpha)$ as the conditional $(1 - \alpha + 2\beta_P)$ quantile of $W_{\widehat{J}_{BMB}}$ given $\{f_t(\widehat{\boldsymbol{\theta}}_i)\}_{t=R+1}^T$.

Given this two-step procedure for the computation of critical values, the following result establishes their theoretical validity:

[14]The problem of many non-binding inequalities is potentially aggravated if a researcher considers as many moment inequalities as possible in practice (and not only adjacent ones) to detect different (potential) violations of monotonicity.

Theorem 2. *Let $c_1 \leq \overline{\sigma}^2(q_P) \leq \max\{\underline{\sigma}^2(r_P), \overline{\sigma}^2(q_P)\} \leq C_1$ and suppose that the assumptions of Theorem 1 hold. Moreover, suppose that $\sup_{P \geq 1} \beta_P < \alpha/2$ and $\log(1/\beta_P) \leq C_1 \log P$, where β_P was defined above. Then, there exist positive constants c and C depending only on c_1 and C_1 such that under H_0:*

$$Pr\Big(U^* > c^{BMB,2S}(\alpha)\Big) \leq \alpha + CP^{-c}, \tag{9}$$

where $c^{BMB,2S}(\alpha)$ is the corresponding critical value at level α from the two-step Block Multiplier Bootstrap procedure with moment selection described before. If $\mathrm{E}[f_t(\boldsymbol{\theta}_i)] = 0$ for all $1 \leq i \leq \kappa$, then:

$$Pr\Big(U^* > c^{BMB,2S}(\alpha)\Big) \leq \alpha - 3\beta_P - CP^{-c}, \tag{10}$$

so that, if in addition $\beta_P \leq C_1 P^{-c_1}$, then $\Big|Pr\Big(U^* > c^{BMB,2S}(\alpha)\Big) - \alpha\Big| \leq CP^{-c}$.

The proof of Theorem 2 has been relegated to the supplementary material. Similar to Theorem 1, results are sub-divided into two parts: the first part of Theorem 2 addresses the case where only some inequalities are binding, while the second part considers the case when all moments are zero, in other words $\mathrm{E}[f_t(\boldsymbol{\theta}_i)] = 0$ for all $1 \leq i \leq \kappa$. As expected, the possibility of dropping some inequalities from the set of moment restrictions used for the computation of critical values when they are actually informative, leads to an additional error of order $3\beta_P$. This error, however, becomes negligible when $\beta_P \leq C_1 P^{-c_1}$ and the same convergence rate as in Theorem 1 is obtained.

4 Monte Carlo Simulation

4.1 Set-up

In this section we investigate the finite sample properties of the test statistic in Equation (6) using the 'small' block 'large' block BMB employed by Chernozhukov et al. (2014). We will use the MSFE loss function for f to measure nowcast accuracy as described in Equation (2).

As mentioned above, it may be advisable to use more than just the adjacent horizons $i+1$ and i in order to have more moment inequalities which could improve the power of the test. Since using the full set of $S(S-1)/2$ inequalities is not computationally feasible in this Monte Carlo simulation even for reasonable levels of S, we will use inequalities comparing $i+k$ to i for all $k = \{1, S/10, S/5\}$. By setting the number of nowcast models to be $S = \{30, 80, 130\}$, this implies a total of $\kappa = \{80, 215, 350\}$ moment inequalities. In considering sample sizes $P = \{100, 200\}$ we allow for scenarios in which the number of moment inequalities is larger than the number of observations. For the bootstrap, we consider a grid of values for the small and large blocks equal to $r_P = \{0, 1, 2\}$ and $q_P = \{4, 5, 6, 7, 8\}$.

We will also provide some results on the optimal block length selection procedure, but are unable to explore this in great detail for reasons of computational complexity. Similarly, we only present results without moment inequality selection. Finally, we use $B = 399$ bootstrap replications and $M = 1000$ Monte Carlo replications.

Since Theorem 1 shows that parameter estimation error is irrelevant under our outlined assumptions, for each of the out-of-sample periods $t = R + 1, ..., T$ we can directly simulate the nowcast errors $u_{i,t}$ for $i = 1, ..., S$. The general form of the data generating process (DGP) for each of the forecast errors $\varepsilon_{i,t}$ closely follows that of Zhang and Cheng (2014) and Jin et al. (2016) allowing for cross-sectional and time series dependence is:

$$u_{i,t} = \rho u_{i,t-1} + (1-\rho)\left(\sqrt{1-\phi}v_{i,t} + \sqrt{\phi}v_{0,t}\right) \tag{11}$$

where ρ and ϕ are scalars such that $|\rho| < 1$ and $0 \leq \phi < 1$ and $v_{i,t} \sim i.i.d.\mathcal{N}(0, \sigma_i^2)$ for all $i = 0, ..., S$. The term $v_{0,t}$ is introduced only as a device to produce cross-sectional dependence. Note that in this set-up when $\sigma_i^2 = 1$ for all i, it follows that for some errors $u_{i,t}$ and $u_{j,v}$, we have $\text{Cor}(u_{i,t}, u_{j,v}) = \rho^{|t-v|}\phi$. Therefore we can choose ρ and ϕ to vary the degree of time-series and cross-section dependence. By varying σ_i^2 for different i we can generate scenarios of non-monotonicity to assess the power of the test.

In order to assess the performance of the test when the null hypothesis is true, we use a parameterisation of the DGP with constant forecast error variance across all i:

DGP-N: $\sigma_i^2 = 1$ for all $i = 0, ..., S$.

This 'least favourable case' is often chosen to assess the performance of the test when the null is satisfied as it is for this case that rejection rates should be close to the nominal level α.

In formulating DGPs under the alternative, there are many different ways in which monotonicity can be violated. We follow the literature on monotonicity testing in formulating three different types of violation of monotonicity, differing in the degree and persistence of violation (e.g., Ghosal et al., 2000). All three cases are depicted graphically in Figure 2.

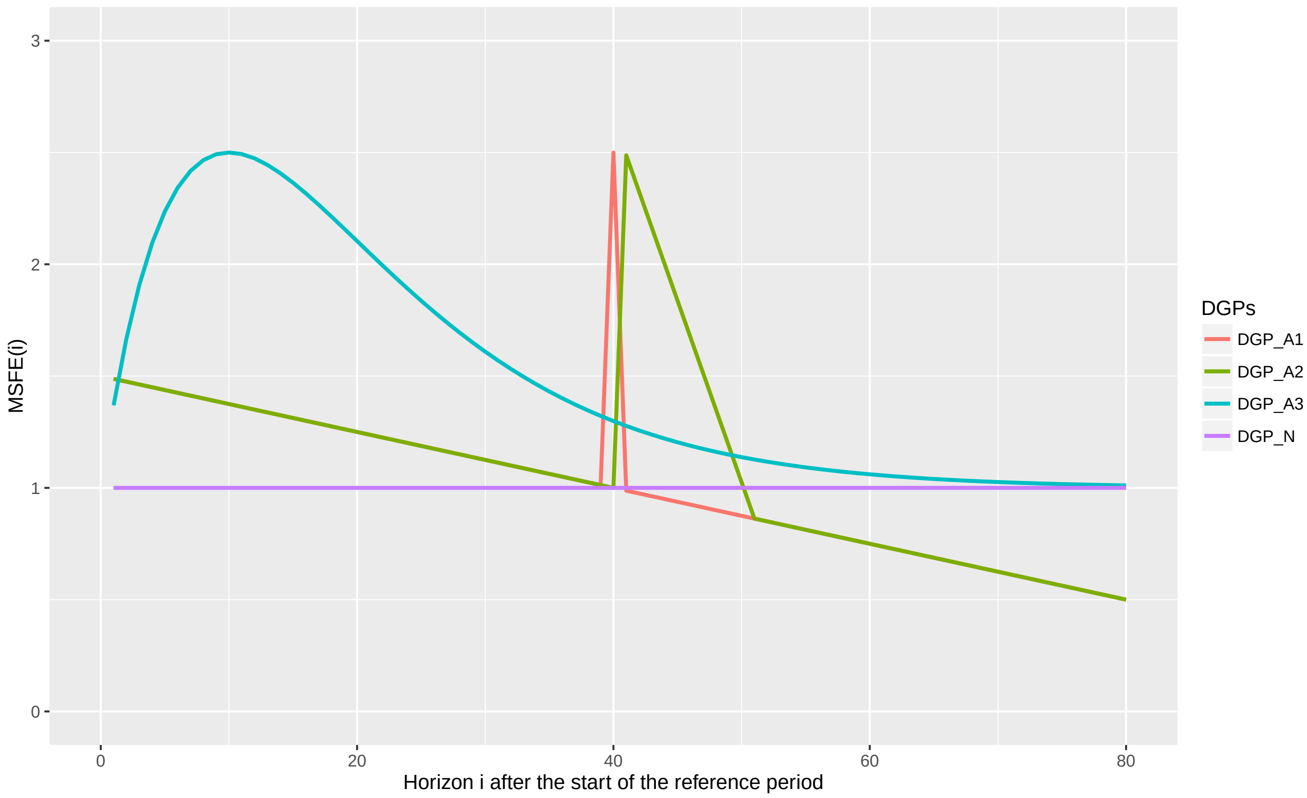
DGP-A1: For $i = \frac{S+1}{2}$, $\sigma_i^2 = \frac{3}{2} - \frac{i}{S} + K$ with $K > 0$, else for all $i \neq \frac{S+1}{2}$, $\sigma_i^2 = \frac{3}{2} - \frac{i}{S}$.

DGP-A2: For all $i = \frac{S+1}{2} + j$ and $j = 1, ..., J$, $\sigma_i^2 = \frac{3}{2} - \frac{i}{S} + K\frac{(J-j+1)}{J}$ with $K > 0$, else $\sigma_i^2 = \frac{3}{2} - \frac{i}{S}$.

DGP-A3: For all i, $\sigma_i^2 = 1 + \frac{8K}{Se^{-1}} i e^{-\frac{8i}{S}}$, with $K > 0$.

In each of the above cases, we can set $K = \widetilde{K}/(1-\phi)$ so that the largest violation of monotonicity is of magnitude $\widetilde{K}$. The rescaling by $(1-\phi)^{-1}$ is required because violations in monotonicity due

Figure 2: MSFE Profiles Under the Null (DGP-N) and Alternatives (DGP-A1, DGP-A2, DGP-A3)



Notes: This example has been created for $S = 80$ with parameter values $\widetilde{K} = 1.5$ and $J = 10$.

to changes in σ_i^2 are dampened when ϕ is large according to the DGP in Equation (11).[15] We can therefore use this parameter, which we set to $\widetilde{K} = 1.5$, to analyse the power properties of the test. Case A1 is a violation of monotonicity at a single point (spike) occurring at the point $i = S/2$, which should help to detect the test's sensitivity to small periods of violations of the null. In case A2, there are a finite number of J violations of monotonicity starting at $i = S/2$ and decreasing in size. This case is designed to mimic possible violations of monotonicity due to measurement error in a specific variable, whose effect slowly tapers off as the variable moves through the model. The final case is a standard non-monotonic function often used in simulation studies on tests for monotonicity. Here, the $MSFE$ follows a humped shape, with violations of monotonicity at all $i \leq S/8$.

[15] It can be seen that the long-run variance of $\varepsilon_{i,t}$ is $(1+\rho)^{-1}(1-\rho)[(1-\phi)\sigma_i^2 + \phi\sigma_0^2]$. Therefore supposing $\sigma_1^2 = 1$ and $\sigma_2^2 = 1 + K$, then the ratio $\text{Var}[\varepsilon_{2,t}]/\text{Var}[\varepsilon_{1,t}] = 1 + (1-\phi)K$. Only by setting $K = \widetilde{K}/(1-\phi)$ do we fix this ratio to be equal to $1 + \widetilde{K}$, independently of ϕ.

4.2 Results

Tables 1, 2 and 3 present the rejection rates for test using the test statistic in Equation (6), for a nominal size of 5%. The tables of results are for the combinations of time series and cross section dependence (ρ, ϕ) equal to $(0.2, 0.2)$, $(0.5, 0.2)$ and $(0.5, 0.5)$ respectively.

We can see from the results that our test delivers reasonably good size control across the various specifications we consider. In general, the results are best where the 'small' block size is $r = 0$, and the 'large' block length is small, such as $q = 4$ and $q = 5$. For example, in Table 3 for block lengths $(r, q) = (0, 5)$ and for the most realistic dependence parameters for macroeconomic data, $(\rho, \phi) = (0.5, 0.5)$, we see that the empirical sizes for $P = 200$ are very close to nominal size with the rejection rate 4.9% for both $k = 80, 215$ and 4.2% for $\kappa = 350$. For the smaller sample size $P = 100$ there is some slight under-sizing in all specifications, particularly for the lower levels of dependence. This under-sizing and the magnitude of under-rejections seen in our results is similar to Zhang and Cheng (2014)'s test results of different hypotheses. The results are relatively insensitive to the block lengths, although there is a small decline as the small block length increases to $r = 2$, which could be due to the large amount of observations being dropped by the large-block small-block procedure. The optimal choice of block length may be determined by the procedure described above.

Turning to the power results, we see some qualitative differences between the first two DGPs, DGP-A1 and DGP-A2, and the final DGP-A3. The power results are very good for DGP-A1 and DGP-A2, which have short spiky violations of the null, with rejection rates above 98% in every specification for the larger sample size $P = 200$, and higher than 80% in most cases for $P = 100$. For DGP-A3, where there is smooth violation of the null, as seen in Figure 2, the power properties of the test are somewhat worse. Particularly for the smaller sample size $P = 100$, rejection frequencies tend to be below 40%, although as the sample increases to $P = 200$, the rejection rate moves above 90% in Table 1, and above 70% in the other two cases. We feel that the performance under this DGP could be improved by considering a larger set of moment inequalities, which we did not consider for computational reasons.[16]

Finally, Figure 4 below presents a small set of results for the optimal block length procedure outlined in Section 3.2 for the data generating process DGP-A1 with sample size $P = 100$ and dependence combination $\rho = 0.2$ and $\phi = 0.2$. We have set the number of inner and outer bootstrap replications described in that section to $B_I = 299$ and $B_O = 400$, respectively, while the number of Monte Carlo and of bootstrap replications to calculate the critical values are as before.

The results suggest that the optimal block length procedure does a reasonable job in picking a suitable (r, q) combination from the grid as rejection rates lie amid those from Table 1 for the

[16] When assessing the robustness of our Monte Carlo experiment, we tried a version where we took the set of moment inequalities using only the adjacent horizons where $\kappa = S - 1$, and the performance under DGP-A3 was even poorer. This is because, with such smooth deviations from the null, horizons very close together are almost indistinguishable from the null case.

Table 1: Rejection Frequencies - Nominal size 5%

		$P = 100$, $\rho = 0.2$, $\phi = 0.2$											
		$\kappa = 80$				$\kappa = 215$				$\kappa = 350$			
		DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3
$r = 0$	$q = 4$	0.032	0.966	0.969	0.129	0.018	0.943	0.964	0.284	0.023	0.937	0.972	0.401
$r = 0$	$q = 5$	0.033	0.963	0.963	0.113	0.016	0.938	0.964	0.286	0.016	0.934	0.975	0.396
$r = 0$	$q = 6$	0.029	0.961	0.965	0.118	0.015	0.936	0.962	0.272	0.018	0.926	0.970	0.377
$r = 0$	$q = 7$	0.030	0.960	0.965	0.133	0.016	0.934	0.952	0.266	0.012	0.934	0.963	0.373
$r = 0$	$q = 8$	0.031	0.954	0.962	0.133	0.017	0.919	0.958	0.263	0.014	0.922	0.969	0.365
$r = 1$	$q = 4$	0.032	0.960	0.967	0.113	0.017	0.936	0.958	0.287	0.018	0.932	0.964	0.363
$r = 1$	$q = 5$	0.027	0.960	0.961	0.119	0.012	0.926	0.957	0.265	0.012	0.926	0.968	0.358
$r = 1$	$q = 6$	0.027	0.963	0.969	0.123	0.018	0.928	0.950	0.273	0.012	0.924	0.960	0.367
$r = 1$	$q = 7$	0.024	0.954	0.961	0.112	0.012	0.922	0.944	0.251	0.016	0.917	0.961	0.352
$r = 1$	$q = 8$	0.025	0.956	0.962	0.116	0.014	0.925	0.960	0.260	0.016	0.919	0.962	0.346
$r = 2$	$q = 4$	0.027	0.954	0.963	0.115	0.009	0.922	0.953	0.248	0.013	0.914	0.957	0.337
$r = 2$	$q = 5$	0.027	0.958	0.963	0.119	0.015	0.926	0.948	0.252	0.010	0.924	0.950	0.335
$r = 2$	$q = 6$	0.022	0.952	0.958	0.108	0.010	0.913	0.946	0.237	0.014	0.914	0.954	0.341
$r = 2$	$q = 7$	0.025	0.955	0.958	0.115	0.012	0.915	0.948	0.254	0.007	0.920	0.955	0.322
$r = 2$	$q = 8$	0.026	0.958	0.956	0.126	0.012	0.918	0.945	0.249	0.015	0.911	0.950	0.319
		$P = 200$, $\rho = 0.2$, $\phi = 0.2$											
		$\kappa = 80$				$\kappa = 215$				$\kappa = 350$			
		DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3
$r = 0$	$q = 4$	0.035	1.000	1.000	0.271	0.026	1.000	1.000	0.795	0.039	0.999	1.000	0.940
$r = 0$	$q = 5$	0.033	0.999	1.000	0.286	0.022	1.000	1.000	0.791	0.036	0.999	1.000	0.929
$r = 0$	$q = 6$	0.028	0.999	1.000	0.284	0.020	1.000	1.000	0.780	0.032	0.998	1.000	0.920
$r = 0$	$q = 7$	0.032	1.000	1.000	0.278	0.022	1.000	1.000	0.765	0.026	0.999	1.000	0.930
$r = 0$	$q = 8$	0.029	0.999	1.000	0.275	0.020	1.000	1.000	0.775	0.030	0.999	1.000	0.917
$r = 1$	$q = 4$	0.037	1.000	1.000	0.280	0.019	1.000	1.000	0.790	0.033	0.999	1.000	0.923
$r = 1$	$q = 5$	0.024	0.999	1.000	0.273	0.021	1.000	1.000	0.776	0.027	0.999	1.000	0.917
$r = 1$	$q = 6$	0.031	0.999	1.000	0.270	0.019	1.000	1.000	0.773	0.025	0.997	1.000	0.918
$r = 1$	$q = 7$	0.028	0.999	1.000	0.264	0.016	1.000	1.000	0.761	0.030	0.999	1.000	0.919
$r = 1$	$q = 8$	0.029	0.999	1.000	0.270	0.017	1.000	1.000	0.755	0.021	0.998	1.000	0.911
$r = 2$	$q = 4$	0.027	1.000	1.000	0.268	0.021	0.999	1.000	0.761	0.024	0.999	1.000	0.908
$r = 2$	$q = 5$	0.034	1.000	1.000	0.275	0.020	1.000	1.000	0.766	0.027	0.997	1.000	0.906
$r = 2$	$q = 6$	0.028	1.000	1.000	0.265	0.018	1.000	1.000	0.758	0.022	0.999	1.000	0.915
$r = 2$	$q = 7$	0.028	1.000	1.000	0.271	0.019	1.000	1.000	0.757	0.026	0.998	1.000	0.912
$r = 2$	$q = 8$	0.023	0.999	1.000	0.262	0.015	1.000	1.000	0.749	0.020	0.998	1.000	0.901

Table 2: Rejection Frequencies - Nominal size 5%

		$P = 100$, $\rho = 0.5$, $\phi = 0.2$											
		$\kappa = 80$				$\kappa = 215$				$\kappa = 350$			
		DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3
$r = 0$	$q = 4$	0.050	0.863	0.881	0.107	0.038	0.806	0.857	0.190	0.033	0.765	0.869	0.222
$r = 0$	$q = 5$	0.041	0.850	0.855	0.106	0.033	0.785	0.825	0.181	0.023	0.742	0.845	0.198
$r = 0$	$q = 6$	0.042	0.848	0.850	0.098	0.021	0.763	0.808	0.176	0.019	0.724	0.817	0.179
$r = 0$	$q = 7$	0.034	0.838	0.839	0.099	0.024	0.759	0.814	0.154	0.021	0.720	0.786	0.171
$r = 0$	$q = 8$	0.031	0.829	0.838	0.094	0.026	0.752	0.781	0.150	0.015	0.706	0.795	0.149
$r = 1$	$q = 4$	0.047	0.853	0.869	0.110	0.031	0.791	0.837	0.194	0.030	0.747	0.841	0.211
$r = 1$	$q = 5$	0.042	0.843	0.851	0.102	0.025	0.759	0.815	0.172	0.020	0.719	0.805	0.181
$r = 1$	$q = 6$	0.034	0.832	0.847	0.098	0.026	0.747	0.803	0.158	0.019	0.698	0.782	0.163
$r = 1$	$q = 7$	0.027	0.817	0.842	0.089	0.022	0.736	0.785	0.146	0.015	0.708	0.778	0.155
$r = 1$	$q = 8$	0.031	0.816	0.834	0.088	0.021	0.751	0.776	0.151	0.017	0.699	0.768	0.150
$r = 2$	$q = 4$	0.045	0.851	0.864	0.109	0.026	0.769	0.823	0.180	0.025	0.725	0.807	0.182
$r = 2$	$q = 5$	0.038	0.830	0.855	0.102	0.022	0.745	0.802	0.171	0.018	0.714	0.784	0.170
$r = 2$	$q = 6$	0.029	0.803	0.839	0.093	0.020	0.735	0.782	0.150	0.018	0.688	0.774	0.156
$r = 2$	$q = 7$	0.034	0.822	0.843	0.089	0.018	0.736	0.771	0.144	0.014	0.691	0.761	0.147
$r = 2$	$q = 8$	0.036	0.830	0.825	0.083	0.017	0.738	0.773	0.133	0.015	0.681	0.749	0.137
		$P = 200$, $\rho = 0.5$, $\phi = 0.2$											
		$\kappa = 80$				$\kappa = 215$				$\kappa = 350$			
		DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3
$r = 0$	$q = 4$	0.062	0.995	0.998	0.210	0.052	0.993	0.997	0.582	0.053	0.991	1.000	0.735
$r = 0$	$q = 5$	0.051	0.996	0.996	0.194	0.039	0.992	0.997	0.539	0.035	0.989	1.000	0.698
$r = 0$	$q = 6$	0.049	0.997	0.997	0.191	0.034	0.990	0.997	0.514	0.033	0.990	0.999	0.668
$r = 0$	$q = 7$	0.040	0.996	0.998	0.182	0.027	0.990	0.995	0.488	0.031	0.988	0.999	0.652
$r = 0$	$q = 8$	0.037	0.997	0.998	0.174	0.030	0.991	0.996	0.472	0.029	0.990	0.999	0.630
$r = 1$	$q = 4$	0.060	0.996	0.999	0.213	0.050	0.993	0.997	0.555	0.045	0.990	0.999	0.719
$r = 1$	$q = 5$	0.047	0.997	0.995	0.207	0.033	0.989	0.996	0.514	0.037	0.990	0.999	0.666
$r = 1$	$q = 6$	0.041	0.996	0.998	0.193	0.029	0.988	0.995	0.510	0.028	0.990	0.999	0.646
$r = 1$	$q = 7$	0.040	0.994	0.995	0.184	0.025	0.990	0.995	0.477	0.023	0.986	0.999	0.624
$r = 1$	$q = 8$	0.036	0.994	0.996	0.173	0.025	0.988	0.996	0.468	0.024	0.988	0.999	0.630
$r = 2$	$q = 4$	0.053	0.995	0.997	0.212	0.042	0.992	0.997	0.542	0.037	0.990	1.000	0.689
$r = 2$	$q = 5$	0.049	0.996	0.997	0.187	0.029	0.990	0.995	0.494	0.026	0.989	0.998	0.648
$r = 2$	$q = 6$	0.042	0.995	0.997	0.173	0.026	0.988	0.994	0.475	0.025	0.987	0.998	0.639
$r = 2$	$q = 7$	0.042	0.994	0.994	0.177	0.026	0.986	0.995	0.465	0.021	0.985	0.999	0.623
$r = 2$	$q = 8$	0.035	0.992	0.996	0.176	0.022	0.986	0.997	0.455	0.021	0.986	0.996	0.595

Table 3: Rejection Frequencies - Nominal size 5%

		$P = 100$, $\rho = 0.5$, $\phi = 0.5$											
		$\kappa = 80$				$\kappa = 215$				$\kappa = 350$			
		DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3
$r = 0$	$q = 4$	0.050	0.893	0.906	0.123	0.041	0.855	0.896	0.214	0.028	0.844	0.906	0.246
$r = 0$	$q = 5$	0.046	0.886	0.888	0.110	0.028	0.847	0.882	0.197	0.020	0.819	0.891	0.227
$r = 0$	$q = 6$	0.038	0.880	0.883	0.107	0.020	0.839	0.867	0.181	0.020	0.805	0.871	0.196
$r = 0$	$q = 7$	0.036	0.869	0.876	0.103	0.022	0.832	0.867	0.169	0.016	0.802	0.852	0.192
$r = 0$	$q = 8$	0.031	0.868	0.869	0.102	0.024	0.828	0.842	0.149	0.014	0.782	0.847	0.182
$r = 1$	$q = 4$	0.054	0.885	0.897	0.122	0.036	0.852	0.886	0.200	0.026	0.823	0.875	0.226
$r = 1$	$q = 5$	0.045	0.880	0.892	0.118	0.027	0.834	0.854	0.178	0.022	0.806	0.871	0.205
$r = 1$	$q = 6$	0.040	0.874	0.883	0.106	0.023	0.829	0.856	0.182	0.014	0.796	0.845	0.196
$r = 1$	$q = 7$	0.032	0.861	0.873	0.106	0.023	0.820	0.846	0.155	0.018	0.781	0.843	0.179
$r = 1$	$q = 8$	0.031	0.861	0.868	0.105	0.017	0.822	0.833	0.164	0.019	0.792	0.838	0.164
$r = 2$	$q = 4$	0.056	0.880	0.890	0.130	0.026	0.837	0.859	0.190	0.025	0.808	0.866	0.199
$r = 2$	$q = 5$	0.051	0.866	0.886	0.113	0.023	0.817	0.852	0.176	0.017	0.793	0.854	0.192
$r = 2$	$q = 6$	0.033	0.860	0.866	0.109	0.018	0.815	0.846	0.155	0.022	0.770	0.836	0.175
$r = 2$	$q = 7$	0.034	0.855	0.864	0.104	0.016	0.815	0.836	0.158	0.013	0.779	0.833	0.170
$r = 2$	$q = 8$	0.033	0.864	0.866	0.102	0.017	0.813	0.823	0.145	0.015	0.767	0.829	0.167
		$P = 200$, $\rho = 0.5$, $\phi = 0.5$											
		$\kappa = 80$				$\kappa = 215$				$\kappa = 350$			
		DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3	DGP-N	DGP-A1	DGP-A2	DGP-A3
$r = 0$	$q = 4$	0.055	0.995	0.998	0.229	0.058	0.998	0.998	0.617	0.062	0.994	0.999	0.764
$r = 0$	$q = 5$	0.049	0.995	0.997	0.209	0.049	0.998	0.997	0.588	0.042	0.994	1.000	0.733
$r = 0$	$q = 6$	0.048	0.996	0.996	0.207	0.039	0.998	0.997	0.563	0.038	0.994	0.998	0.705
$r = 0$	$q = 7$	0.039	0.996	0.998	0.199	0.028	0.998	0.996	0.537	0.031	0.993	0.999	0.694
$r = 0$	$q = 8$	0.038	0.996	0.997	0.198	0.030	0.997	0.996	0.531	0.031	0.991	0.999	0.675
$r = 1$	$q = 4$	0.057	0.996	0.997	0.222	0.057	0.998	0.998	0.589	0.054	0.994	1.000	0.748
$r = 1$	$q = 5$	0.050	0.996	0.998	0.204	0.044	0.997	0.997	0.552	0.044	0.993	0.998	0.721
$r = 1$	$q = 6$	0.043	0.996	0.997	0.206	0.032	0.997	0.997	0.545	0.028	0.992	0.999	0.683
$r = 1$	$q = 7$	0.039	0.995	0.997	0.187	0.035	0.997	0.995	0.526	0.030	0.990	1.000	0.662
$r = 1$	$q = 8$	0.038	0.994	0.996	0.190	0.026	0.996	0.997	0.513	0.021	0.992	0.999	0.666
$r = 2$	$q = 4$	0.052	0.995	0.996	0.223	0.050	0.997	0.998	0.566	0.053	0.994	0.999	0.729
$r = 2$	$q = 5$	0.048	0.997	0.997	0.208	0.039	0.999	0.996	0.532	0.037	0.993	1.000	0.695
$r = 2$	$q = 6$	0.034	0.995	0.997	0.191	0.031	0.996	0.994	0.522	0.029	0.992	0.999	0.674
$r = 2$	$q = 7$	0.040	0.995	0.994	0.178	0.026	0.998	0.997	0.517	0.025	0.991	0.999	0.662
$r = 2$	$q = 8$	0.043	0.994	0.995	0.184	0.027	0.997	0.996	0.498	0.024	0.991	0.998	0.637

Table 4: Optimal Block Choice Procedure - Rejection Frequencies

$\kappa = 80$	$\kappa = 215$	$\kappa = 350$
0.959	0.927	0.898

corresponding setup, at least for $\kappa = 80$ and $\kappa = 215$. For $\kappa = 350$, we observe rejection rates which are slightly below the ones from Table 1, but the difference is only marginal. We leave it as a topic for future research whether additional iterations of the optimal block length procedure may improve results further.

5 Empirical Application

In this section we will present the results of our monotonicity test in nowcasting the United States aggregate real GDP growth rate, and the growth rate of 5 GDP subcomponents, using single-variable bridge equation methods and forecast combinations involving these bridge equations. These methods are commonly used by central banks. We improve upon existing empirical studies of bridge equation methods by assessing how they perform sequentially as new information becomes available throughout the nowcast period. This is in contrast to studies such as Schumacher (2016) which compares different bridge equation models by looking at their relative performance at a single point in the quarter. We also aim to provide new evidence on the performance of forecast combinations of bridge models, thereby complementing the results of papers such as Rünstler et al. (2009).

We also add to the literature of nowcasting the individual components of GDP, which are less common than the large volume of empirical studies on nowcasting aggregate real GDP. Some examples include Baffigi et al. (2004) who take a bottom-up approach to nowcasting aggregate GDP, and Antolín Díaz et al. (2016) who nowcast real consumption separately to aggregate GDP. We find this to be an important exercise as nowcasting the GDP sub-components to some extent mimics the way that the Bureau of Economic Analysis (BEA) constructs the first estimates of aggregate GDP, as mentioned in the Working Paper for the Atlanta Fed's GDPNow model, see Higgins (2014).

5.1 Description of Data and Nowcasting Methods

The quarterly real GDP variable and its 5 sub-components (consumption, investment, government expenditure, exports and imports), along with 13 monthly bridge equation predictors, are displayed in Table 5. These are accessed from the Haver Analytics database USECON.[17] The monthly variables we use are similar to existing nowcasting studies and include survey data, employment, prices and housing series.

[17]We are grateful to Now-Casting Economics Ltd. for access to this data.

Table 5: Monthly Bridge Equation Variables and Quarterly GDP Component Target Variables

Variable	Publication Lag	Source	Frequency
University of Michigan: Consumer Sentiment	27	UMICH	Monthly
Conference Board: Consumer Confidence	31	CB	Monthly
ISM Manufacturing: PMI Composite Index	32	ISM	Monthly
Domestic Retail Auto Sales	33	BEA	Monthly
Unemployment Rate: 16 years +	34	BLS	Monthly
Retail Sales & Food Services	43	CENSUS	Monthly
PPI: Finished Goods	43	BLS	Monthly
Housing Starts	47	CENSUS	Monthly
CPI-U: All Items	47	BLS	Monthly
Industrial Production Index	47	FRB	Monthly
Building Permits	55	CENSUS	Monthly
Exports: Goods	64	CENSUS	Monthly
Imports: Goods	64	CENSUS	Monthly
Real Gross Domestic Product	119	BEA	Quarterly
Real Personal Consumption Expenditures	119	BEA	Quarterly
Real Gross Private Domestic Investment	119	BEA	Quarterly
Real Gov Consumption Exp & Gross Inv	119	BEA	Quarterly
Real Exports of Goods & Services	119	BEA	Quarterly
Real Imports of Goods & Services	119	BEA	Quarterly

Notes: Publication Lag: approximate time in days between the start of the reference month/quarter and the first data release. Source: UMICH (University of Michigan), CB (The Conference Board), ISM (Institute for Supply Management), BEA (Bureau of Economic Analysis), BLS (Bureau of Labor Statistics), CENSUS (Census Bureau) and FRB (Federal Reserve Board). All variables are seasonally adjusted, where appropriate.

Table 5 also includes some information regarding the approximate publication lag of each variable, defined as the difference in days between the start of the reference month/quarter and the first release of the series.[18] The first-release of GDP occurs around a month after the end of the reference quarter, for example Q1 is released near the end of April. We can see that the survey data is the most timely of the variables we consider, which is why they are frequently used in nowcasting studies, as mentioned by Bańbura et al. (2013). The timeliest is the University of Michigan Consumer Sentiment Index which is actually published before the end of the month to which it pertains. The longer publication lags occur in the industrial production, housing and trade variables.

This information is then used to generate the flow of data which determines the sequence of nowcasting models we use to predict each quarter's GDP growth. Starting from the first day of

[18] These publication lags were defined based on the release date of the last available data in the vintage accessed on 8th June 2016.

the quarter and finishing around day 119 when GDP is first released, we create a pseudo-calendar containing all of the variables released in that period.[19] Therefore we have 3 months' releases of the 13 monthly variables in the nowcast period, and an additional 11 releases which occur in the backcast period before GDP is released. This yields a total of $S = 50$ different nowcast updates per quarter.

We use data on all months and quarters between 1978Q2 and 2016Q1, which gives a total of $T = 152$ quarterly observations and $3T = 456$ monthly observations. We start making nowcasts of 1990Q1 using a quarterly rolling window length of $R = 47$ and continue to nowcast up to 2016Q1; a total of $P = 105$ predictions at each point in the nowcast period. We measure nowcast accuracy using the MSFE statistic described in Equation (2), which is therefore the average of 105 squared nowcast errors for each $i = 1, ..., 50$. All variables are transformed to stationary growth rates by either differencing or log-differencing. We choose the set of moment inequalities to be the same as that used in the Monte Carlo section, comparing horizon $i + k$ to i for all $k = \{1, S/10, S/5\}$. This yields a total of $\kappa = 134$ moment inequalities; a high-dimensional setting on which to apply our test. We will apply the moment inequality selection procedure outlined above, with $\beta_P = 0.01$ as in Chernozhukov et al. (2014).

We consider two different nowcasting methods. The first is a bridge equation method where we use the single predictor variable which has just been released at that point in the data flow. We augment the model described earlier with an AR(1) term to calculate nowcasts in a similar way to Equation (4):

$$\widehat{y}_{it} = \widehat{\theta}_{0it} + \widehat{\theta}_{1it}\widehat{\overline{X}}_{it} + \widehat{\theta}_{2it}y_{t-1} \text{ for } i = 1, ..., S \text{ and } t = R + 1, ..., T$$

There are a few things to note. Firstly, the parameters $\widehat{\theta}_{0it}$, $\widehat{\theta}_{1it}$ and $\widehat{\theta}_{2it}$ are subscripted both by i and t to denote both the variable used, and the rolling window of data used to estimate them. We estimate these parameters by OLS. Secondly, note that $\widehat{y}_{it}$ is also subscripted by i to be clear that there are $S = 50$ distinct nowcasts made of the same quarter for all quarters $t = R + 1, ..., T$. Thirdly, as mentioned above, we deal with the ragged edge problem for the predictor X_{it} by forecasting the most recent periods using a monthly AR(1) model. At different points in the data flow from $i = 1, ..., 50$, this implies that nowcasts generated from the predictor $\widehat{\overline{X}}_t$ may either require 3, 2, 1 or no monthly forecasts to be made of the underlying monthly series.[20] Finally, since there are some periods at the beginning of the nowcast period for y_t for which y_{t-1} is also not available, a backcast of y_{t-1} is generated using an AR(1) model.

[19] This is a pseudo-calendar rather than an actual release calendar because we assume for simplicity that the same publication lag for each variable applies to every data point we observe. This is necessary as it gives us a fixed number of nowcasts per quarter. The pseudo-calendar would deviate from the actual release calendar in instances where a certain variable was released just before GDP in some quarters and just after GDP in other quarters, but we expect this simplification to have little or no impact on the results.

[20] For example, by the time GDP is first released, the University of Michigan CSI is available in all three months of the nowcast period and therefore requires no monthly forecasts to be made in order to obtain nowcasts.

The second nowcasting method we use is a simple 'equal-weights' forecast combination based on these individual bridge equation models. At each point $i = 1, ..., 50$, the simple average is taken of all of the nowcasts made up to point i:

$$\widehat{y}_{it}^{FC} = \frac{1}{i} \sum_{k=1}^{i} \widehat{y}_{kt} \text{ for } i = 1, ..., S \text{ and } t = R+1, ..., T$$

Therefore the second nowcast at $i = 2$ is made by averaging bridge equation predictions $i = 1$ and $i = 2$ and so on. This is different to Rünstler et al. (2009) who take a few fixed points per quarter at which they average bridge forecasts across all variables.

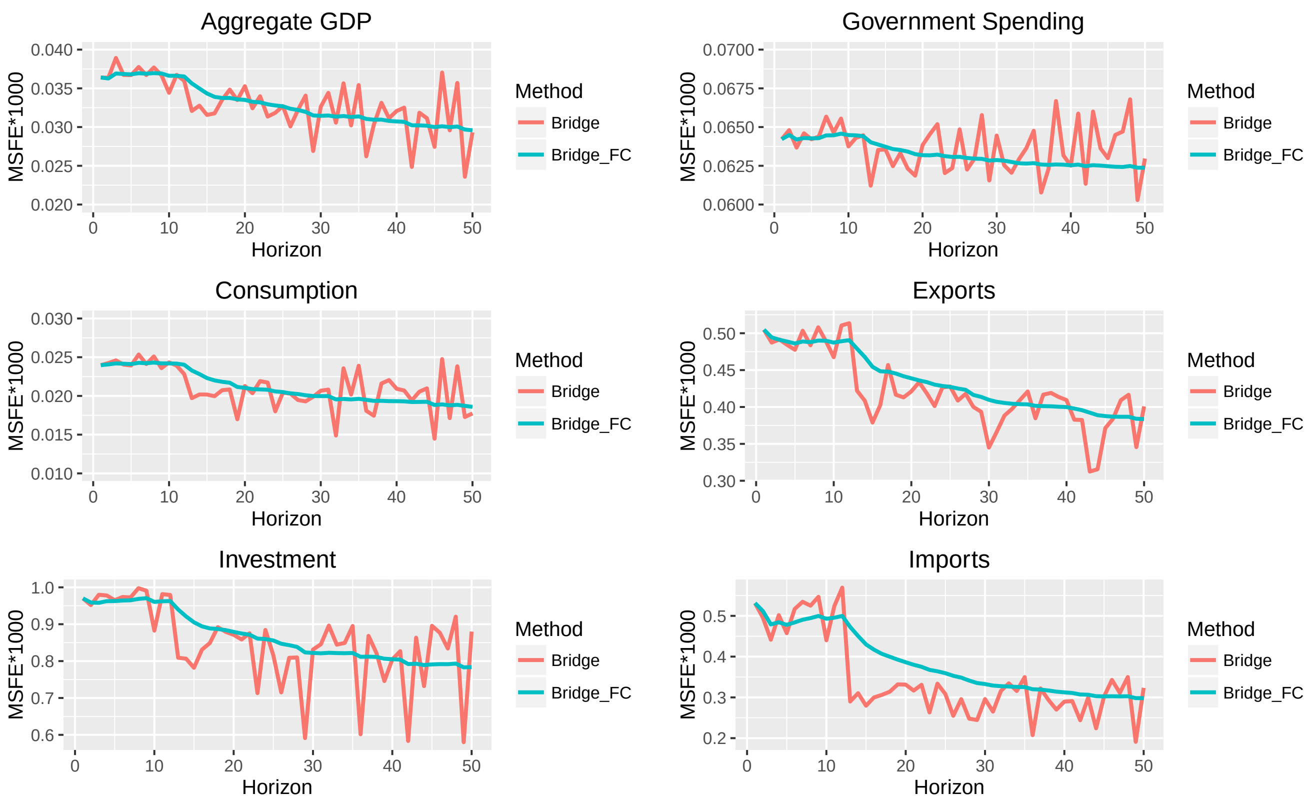
5.2 Results

The graphs in Figure 3 depict the evolution of MSFE from the earliest point in the quarter up until the release of GDP, for the aggregate real GDP growth target, and the 5 GDP sub-components. In the top left panel, the results indicate that the bridge equation method (red line) produces a clearly downwards sloping MSFE for the aggregate GDP growth rate, indicating that nowcasts appear to improve as we approach the publication date of GDP. There is also a predominantly downwards-sloping pattern for the 5 GDP subcomponents, perhaps with the exception of Government Spending which appears to be flat. In all of the series there are spikes up and down, but given the scale of the graphs, these are only particularly large in the case of investment; a result which is to be expected given the volatility of investment series. Note that the jump-down in the MSFE common to most series just after point 10 is due to the previous quarter's GDP being released.

The same results for the forecast combination of the bridge equation methods shows a much more stable declining MSFE pattern. Except for small increases at the beginning of the nowcast period, there are very smooth improvements in the nowcasts as we approach the end of the prediction period. This is most notable in the case of the volatile investment series, where pooling substantially flattens the spikes in the MSFE profile.

The key innovation of our paper is that, rather than relying only on the graphical evidence in Figure 3, we now have a formal statistical test to assess whether these MSFE profiles are monotonically declining. The results of the nowcast monotonicity test are displayed in Table 6. The upper panel reports the results for the pure bridge equation method. For the headline aggregate GDP series, the test clearly finds no evidence that monotonicity is violated with a p-value of $p = 0.27$. This mirrors the result of Bańbura et al. (2013) but for a different nowcasting method and observation period. This suggests that it is possible to generate monotonically improving nowcasts of U.S. real GDP growth irrespective of the method or data span used. The results for government spending, imports and exports are similar to aggregate GDP, and there is no significant evidence of non-monotonicity

Figure 3: MSFE by Nowcast Horizon for Real GDP and 5 GDP Sub-components



Notes: The lines correspond to the MSFE for the Bridge equation method and the Bridge equation forecast combination ("Bridge_FC") method. Horizon refers to the number of the model, starting from the beginning of the reference quarter.

in consumption although the p-value is close to 0.1. On the other hand, turning to the results for investment, with a p-value of 0.068, we find that there is evidence against monotonicity. This implies that, while the bridge equation method produces constantly improving nowcasts of the other GDP sub-components, for investment we should consider a different nowcasting method if monotonicity is the objective. It is also worth noting that the moment inequality selection procedure does not eliminate any of the inequalities for any of the nowcast variables, meaning that non-binding moment inequalities are not extreme enough to pass the threshold for elimination.[21]

Turning to the results of the forecast combination approach to nowcasting in the lower panel of Table 6, a somewhat different picture emerges. As expected from the graphical evidence in Figure 3, we see that the p-values of the test significantly increase in all cases except government spending. The

[21] This may not come as a big surprise given the rather smooth and moderate decline of most MSFE profiles.

Table 6: Monotonicity Test Results for Bridge and Forecast Combination Methods

	Bridge Equation				
	$1000 \times U^*$	50%	90%	95%	p-value
Aggregate GDP	0.1109	0.0854	0.1475	0.1796	0.2682
Consumption	0.1054	0.0604	0.1087	0.1370	0.1228
Investment	3.0726	1.7415	2.8567	3.3497	0.0677
Government Spending	0.0522	0.0669	0.1066	0.1206	0.7895
Exports	1.0665	0.7250	1.4953	1.8648	0.2732
Imports	1.3875	1.0652	2.6812	3.4284	0.3233
	Bridge Equation Forecast Combination				
	$1000 \times U^*$	50%	90%	95%	p-value
Aggregate GDP	0.0065	0.0082	0.0373	0.0496	0.5840
Consumption	0.0032	0.0054	0.0322	0.0458	0.6917
Investment	0.1092	0.2038	0.7472	0.9386	0.7694
Government Spending	0.0029	0.0034	0.0136	0.0186	0.5739
Exports	0.0272	0.1049	0.5034	0.6046	0.8972
Imports	0.1594	0.2418	1.2528	1.5085	0.6516

Notes: The test statistic U^* is re-scaled by 1000 for presentation purposes. The 50%, 90% and 95% percentiles of the BMB bootstrap distribution are presented and also re-scaled. The p-value is the one-sided rejection probability based on the empirical bootstrap distribution.

most notable result is that, in the case of investment, whereas we reject monotonicity for the pure bridge equation method, the reverse is true for forecast averaging, and we conclude that this method is able to provide monotonically improving nowcasts throughout the quarter. Similarly, looking at the results for the consumption sub-component, we see a much larger p-value which makes the evidence for monotonicity more compelling when the forecast averaging approach is used.

6 Conclusion

This paper proposes a novel test to formally assess the performance of various nowcasting procedures. The test is based on evaluating the monotonicity of nowcast predicitions, which may be violated in the presence of measurement error in one (or more) of the nowcasting variables or in case of model misspecification, through moment inequalities. As a special feature of the nowcasting setup we allow explicitly for a large number of moment inequalities relative to the sample size. This is motivated by the recent nowcasting literature which has developed methods capable of producing a large number of nowcasts of variables like GDP each quarter.

As a theoretical contribution, we extend the recent test for many moment inequalities of Chernozhukov et al. (2014) to the time series case of nowcast monotonicity testing. In doing so, we also show that a moment selection procedure to rule out non-binding inequalities proposed in their paper is applicable in our context as well. Moreover, we derive new asymptotic conditions required to prove the negligibility of rolling parameter estimation error used in generating nowcast errors in the pseudo out-of-sample framework of West (1996). This rate links the rolling window length, the out-of-sample evaluation size and the number of moment inequalities, the last of which is new to the literature. Our method differs from previous forecast accuracy tests not only in being able to accommodate this high-dimensional nowcasting setting, but also because our interest is in evaluating the nowcasts of a single method at various points in a quarter, rather than comparing two or more models at a single fixed point.

In addition to Monte Carlo simulations, we provide an in-depth empirical application to nowcasting the growth rate of U.S. aggregate real GDP, and the growth rate of 5 GDP subcomponents using single-variable bridge equations and a pooled bridge equation approach. In line with existing studies, our test reveals that nowcasts of aggregate GDP growth are monotonically improving. However, the test finds evidence of non-monotonicity in the investment subcomponent using single-variable bridge equations, whereas forecast combinations provide a stable path of monotonically improving nowcasts of investment throughout the quarter. Far from being an exhaustive study of the use of bridge equations in nowcasting GDP, future empirical work might apply our test in more detail, using different bridge variables such as those used in bottom-up approaches like Higgins (2014), or to assess the performance of bridge equations in nowcasting Euro Area GDP growth to shed more light on the results of Schumacher (2016).

7 Appendix A: Analytical Example of Nowcast (Non)monotonicity

In this section we provide a simple example which demonstrates the possibility of violations of monotonicity in realistic scenarios, using the bridge equation method described in the text. A quarterly variable y_t is related to an unobserved latent monthly variable y_t^* using the following identity:

$$y_t = y_t^* + y_{t+1/3}^* + y_{t+2/3}^* \tag{12}$$

which holds for any stock variable such as quarterly GDP (see Bańbura and Modugno, 2014), which is the sum of 3 unobserved monthly GDP components. The data generating process (DGP) for the latent monthly variable y_t^* is the AR(1) model:

$$y_t^* = \rho y_{t-1/3}^* + \epsilon_t \tag{13}$$

where $\epsilon_t \overset{iid}{\sim} (0, \sigma_\epsilon^2)$. There are also N candidate monthly predictor variables X_{it} which proxy the latent y_t^* up to a measurement error term:

$$X_{it} = y_t^* + \eta_{it} \tag{14}$$

where $\eta_{it} \overset{iid}{\sim} (0, \sigma_{\eta_i}^2)$, noting that the variance of this measurement error, σ_i^2 varies across i. By rearranging Equations (13) and (14) we have the implied AR(1) process for X_{it} as a function of both the error term ϵ_t and measurement error η_{it}:

$$X_{it} = \rho X_{i,t-1/3} + \eta_{it} - \rho\eta_{i,t-1/3} + \epsilon_t \tag{15}$$

Similarly, using Equations (12) and (14), for each $i = 1, ..., N$ we can rewrite y_t as a function of the observable X_{it} plus measurement error:

$$y_t = X_{it} + X_{i,t+1/3} + X_{i,t+2/3} + \sum_{k=0}^{2} \eta_{i,t+k/3} \tag{16}$$

Nowcasting proceeds as follows. For a given variable i, three monthly releases are observed per quarter: $X_{it}, X_{i,t+1/3}$ and $X_{i,t+2/3}$. Using the bridge equation method described in the text, when the first month, X_{it}, is released, the nowcaster first makes forecasts of $X_{i,t+1/3}$ and $X_{i,t+2/3}$ in order to nowcast y_t. With the AR(1) model, the optimal forecasts are $\widehat{X}_{i,t+1/3} = \rho X_{it}$ and $\widehat{X}_{i,t+2/3} = \rho^2 X_{it}$. This gives rise to the first nowcast, denoted $\widehat{y}_{1t}$:

$$\begin{aligned}\widehat{y}_{1t} &= X_{it} + \widehat{X}_{i,t+1/3} + \widehat{X}_{i,t+2/3} \\ &= (1 + \rho + \rho^2)X_{it}\end{aligned}$$

In the second point in the quarter, after $X_{i,t+1/3}$ is released, the nowcaster only has to predict $X_{i,t+2/3}$, which is $\widehat{X}_{i,t+2/3} = \rho X_{i,t+1/3}$. Therefore the second nowcast of y_t, denoted $\widehat{y}_{2t}$ is:

$$\begin{aligned}\widehat{y}_{2t} &= X_{it} + X_{i,t+1/3} + \widehat{X}_{i,t+2/3} \\ &= X_{it} + (1 + \rho)X_{i,t+1/3}\end{aligned}$$

Finally, once $X_{i,t+2/3}$ is released, no forecasts need to be made of the X's in order to nowcast y_t, and so the third nowcast is:

$$\widehat{y}_{3t} = X_{it} + X_{i,t+1/3} + X_{i,t+2/3}$$

Now, using the expressions for the three nowcasts $\widehat{y}_{1t}$, $\widehat{y}_{2t}$ and $\widehat{y}_{3t}$ and using Equations (15) and

(16), after some simple algebra we arrive at the nowcast errors:

$$\begin{aligned}
\widehat{y}_{1t} - y_t &= (1 + \rho + \rho^2)\eta_{it} - (1 + \rho)\epsilon_{t+1/3} - \epsilon_{t+2/3} \\
\widehat{y}_{2t} - y_t &= \eta_{it} + (1 + \rho)\eta_{i,t+1/3} - \epsilon_{t+2/3} \\
\widehat{y}_{3t} - y_t &= \eta_{it} + \eta_{i,t+1/3} + \eta_{i,t+2/3}
\end{aligned}$$

Intuitively, as we start the quarter our nowcast is only based on the observation X_{it}, and so we have measurement error of the first month, η_{it}, projected into the nowcast period, plus the aggregation of forecast errors from the AR(1) model. However, as we move further into the quarter, we lose the AR forecast errors but also gain the measurement error which comes from observing $X_{i,t+1/3}$ and $X_{i,t+2/3}$.

7.1 Case I: Single Variable Bridge Equation Method

When there is only one predictor variable, $N = 1$, we can use the above nowcast errors to obtain analytical expressions for nowcast losses. Using the MSFE loss function at the three nowcast horizons, and using the assumption that ϵ_t and η_{it} are *iid*, we get:

$$\begin{aligned}
MSFE_1 &= [1 + \rho + \rho^2]^2\sigma^2_{\eta_1} + [1 + (1 + \rho)^2]\sigma^2_\epsilon \\
MSFE_2 &= [1 + (1 + \rho)^2]\sigma^2_{\eta_1} + \sigma^2_\epsilon \\
MSFE_3 &= 3\sigma^2_{\eta_1}
\end{aligned}$$

Now a simple inspection of these $MSFE$s reveals that, when the AR(1) persistence parameter is low, nowcast monotonicity is very easily violated. For example, when $\rho = 0$, $MSFE_1 = \sigma^2_{\eta_1} + 2\sigma^2_\epsilon$, $MSFE_2 = 2\sigma^2_{\eta_1} + \sigma^2_\epsilon$ and $MSFE_3 = 3\sigma^2_{\eta_1}$. In this case, monotonicity is violated when $\sigma^2_{\eta_1} < \sigma^2_\epsilon$. In other words, when the variance of measurement error is larger than the variance of the AR(1) forecast errors, and y_t has no persistence, it is preferable to make nowcasts using only X_{it} and not $X_{i,t+1/3}$ and $X_{i,t+2/3}$ as the increased measurement error from observing these terms causes a violation in monotonicity.

On the other hand, when persistence is fairly large, as in the case of most macroeconomic series, it becomes less likely that monotonicity is violated. For example, when $\rho = 0.5$ we have $MSFE_1 = 3.0625\sigma^2_{\eta_1} + 3.25\sigma^2_\epsilon$, $MSFE_2 = 3.25\sigma^2_{\eta_1} + \sigma^2_\epsilon$ and $MSFE_3 = 3\sigma^2_{\eta_1}$. In this case, $MSFE_2 > MSFE_1$ only if $\sigma^2_{\eta_1} > 11.01\sigma^2_\epsilon$, which requires more extreme measurement error to violate monotonicity.

7.2 Case II: Multiple Variable Bridge Equation Method

Taking the simplest multiple variable case where $N = 2$, we have 6 data releases throughout the quarter in the order $X_{1t}, X_{2t}, X_{1,t+1/3}, X_{2,t+1/3}, X_{1,t+2/3}, X_{2,t+2/3}$. Now, if we use the single equation bridge method to make nowcasts, with the predictor being the one which has most recently been released, then in a similar way to above, we have the following sequence of 6 $MSFE$s:

$$\begin{aligned}
MSFE_1 &= [1+\rho+\rho^2]^2\sigma^2_{\eta_1} + [1+(1+\rho)^2]\sigma^2_\epsilon \\
MSFE_2 &= [1+\rho+\rho^2]^2\sigma^2_{\eta_2} + [1+(1+\rho)^2]\sigma^2_\epsilon \\
MSFE_3 &= [1+(1+\rho)^2]\sigma^2_{\eta_1} + \sigma^2_\epsilon \\
MSFE_4 &= [1+(1+\rho)^2]\sigma^2_{\eta_2} + \sigma^2_\epsilon \\
MSFE_5 &= 3\sigma^2_{\eta_1} \\
MSFE_6 &= 3\sigma^2_{\eta_2}
\end{aligned}$$

From this example it is very simple to see that, regardless of the persistence parameter ρ, monotonicity is violated when $\sigma^2_{\eta_2} > \sigma^2_{\eta_1}$. In other words, this occurs when the variable which is released closer to the end of the quarter has a larger measurement error variance with respect to the unobserved monthly component y^*_t.

This motivating analytical example can be matched rather closely with real-world nowcasting examples. Bańbura et al. (2010) note that in nowcasting GDP there is a "tradeoff between timeliness and precision" since variables such as industrial production are observed with a larger publication lag, but present a better signal for GDP developments. On the other hand, survey series such as consumer sentiment indices are produced in a more timely fashion, but represent weaker signals for economic movements. It can often be very difficult to predict *a priori* whether one variable has more measurement error than another. Our test is able to detect when the inclusion of such variables significantly violates monotonicity. If monotonicity is detected in this case, it may be a good idea to use model averaging or other combination techniques which might reduce the effect of non-monotonicity while also allowing the nowcaster to exploit the timeliness of the data.

8 Appendix B: Proofs of Main Results

Before stating the proof of Theorem 1, we define various quantities and outline three preliminary Lemmas, which will be used subsequently. The proofs of these Lemmas can be found after the proof of Theorem 1. Also, recall that c, c', C, and C' denote generic positive constants whose values may vary from line to line. Define:

$$\overline{U} \equiv \max_{1\le i\le \kappa} \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - \mathrm{E}\Big[f_t(\theta_i) \Big] \right)$$

and

$$U_0 \equiv \max_{1\le i\le\kappa} \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left(f_t(\boldsymbol{\theta}_i) - \mathrm{E}\Big[f_t(\theta_i)\Big]\right).$$

Moreover, let:

$$W_{BMB} \equiv \max_{1\le i\le\kappa} \left(\frac{1}{\sqrt{m_P q_P}} \sum_{h=1}^{m_P} \varepsilon_h \sum_{t\in I_h} \left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - \frac{1}{P} \sum_{t=R+1}^{T} f_t(\widehat{\boldsymbol{\theta}}_{it}) \right)\right),$$

and

$$\overline{W}_{BMB} \equiv \max_{1\le i\le\kappa} \left(\frac{1}{\sqrt{m_P q_P}} \sum_{h=1}^{m_P} \varepsilon_h \sum_{t\in I_h} \left(f_t(\boldsymbol{\theta}_i) - \frac{1}{P} \sum_{t=R+1}^{T} f_t(\boldsymbol{\theta}_i) \right)\right),$$

and

$$W^0_{BMB} \equiv \max_{1\le i\le\kappa} \left(\frac{1}{\sqrt{m_P q_P}} \sum_{h=1}^{m_P} \varepsilon_h \sum_{t\in I_h} \left(f_t(\boldsymbol{\theta}_i) - \mathrm{E}\Big[f_t(\boldsymbol{\theta}_i)\Big] \right)\right),$$

where $\varepsilon_1, \ldots, \varepsilon_{m_P}$ are standard normal random variables independent of the data $\{f_t(\cdot)\}_{t=1}^{T}$. In Section S1 of the supplementary material, we show that under the assumptions of Theorem 1, the results of Theorem 7.1 of Chernozhukov et al. (2014) continue to hold, namely:

$$\rho_P^U \equiv \sup_{t\in\mathbb{R}} \left| \Pr\Big(U_0 \le t\Big) - \Pr\Big(\max_{1\le i\le\kappa} Y_j \le t\Big) \right| \le CP^{-c}, \tag{17}$$

and

$$\rho_P^W \equiv \sup_{t\in\mathbb{R}} \left| \Pr\Big(W^0_{BMB} \le t \Big| \{f_t\}_{t=1}^{T}\Big) - \Pr\Big(\max_{1\le i\le\kappa} Y_j \le t\Big) \right| \le C'P^{-c'}, \tag{18}$$

where $Y = (Y_1, \ldots, Y_\kappa)'$ is a centered normal random vector with covariance matrix:

$$\mathrm{E}\Big[YY'\Big] = (1/m_P q_P) \sum_{h=1}^{m_P} \mathrm{E}\Big[\Big(\sum_{t\in I_h} \boldsymbol{f}_t - \mathrm{E}\Big[\boldsymbol{f}_t\Big]\Big)\Big(\sum_{t\in I_h} \boldsymbol{f}_t - \mathrm{E}\Big[\boldsymbol{f}_t\Big]\Big)'\Big],$$

and $\boldsymbol{f}_t$ denotes a stacked vector in $\mathbb{R}^\kappa$.

Lemma 1. *For some ζ_{P1} satisfying $\zeta_{P1} \log\kappa \le C_1 P^{-c_2}$, where all constants have been defined in Section 3.2, it holds that:*

$$Pr\left(\max_{1\le i\le\kappa} \sqrt{P} \left| \frac{1}{P} \sum_{t=R+1}^{T} f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i) \right| > \zeta_{P1} \right) \le CP^{-c_2}$$

Lemma 2. *For some* ζ_{P1} *as in Lemma 1 and all constants as in Section 3.2, it holds that:*

$$Pr\left(\max_{1\leq i\leq\kappa}\left(\frac{1}{m_P q_P}\sum_{h=1}^{m_P}\sum_{t\in I_h}\left(f_t(\widehat{\boldsymbol{\theta}}_{it})-f_t(\boldsymbol{\theta}_i)\right)^2\right)^{\frac{1}{2}}>\zeta_{P1}\right)\leq CP^{-c_2}.$$

Lemma 3. *For some* φ_P *with* $0<\varphi_P\leq CP^{-c}$*, it holds that:*

$$Pr\left(c^{BMB}(\alpha)\geq c_0(\alpha+\varphi_P)\right)\geq 1-CP^{-c}$$

and

$$Pr\left(c^{BMB}(\alpha)\leq c_0(\alpha-\varphi_P)\right)\geq 1-CP^{-c},$$

where $c_0(\gamma)$*,* $\gamma\in(0,1)$*, denotes the* $(1-\gamma)$ *quantile of the distribution of* $\max_{1\leq j\leq\kappa}Y_j$*.*

Proof of Theorem 1. First, note that under H_0, it holds that:

$$U^*=\max_{1\leq i\leq\kappa}\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}f_t(\widehat{\boldsymbol{\theta}}_{it})\leq\max_{1\leq i\leq\kappa}\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}\left(f_t(\widehat{\boldsymbol{\theta}}_{it})-\mathrm{E}\Big[f_t(\theta_i)\Big]\right)=\overline{U},$$

where the equality holds when $\mathrm{E}\Big[f_t(\theta_i)\Big]=0$ for all i, $1\leq i\leq\kappa$. Moreover, invoking Assumption 5 and in particular $(r_P/q_P)\log^2\kappa\leq C_1P^{-c_2}$ and

$$q_P\,\mathrm{E}\Big[\max_{1\leq t\leq T}\max_{1\leq i\leq\kappa}\Big|f_t(\boldsymbol{\theta}_i)-\mathrm{E}\Big[f_t(\boldsymbol{\theta}_i)\Big]\Big|^4\Big]\log^{\frac{5}{2}}(\kappa P)\leq C_1P^{\frac{1}{2}-c_2}$$

thereof, and using the same steps as in the proof of Theorem 7.1 in Chernozhukov et al. (2014), the following result can be verified conditional on the data:

$$\Pr\left(\left|\overline{W}_{BMB}-W^0_{BMB}\right|>\zeta'_{P1}\middle|\{f_t\}_{t=1}^T\right)\leq CP^{-c'},$$

where $\zeta'_{P1}=C'P^{-\frac{1}{4}+\frac{3c_2}{4}}\log\kappa$. By the arguments of the proof of Theorem 7.1 and using ρ_P^W from Equation (18), it thus follows that:

$$\rho_P\equiv\sup_{t\in\mathbb{R}}\left|\Pr\left(\overline{W}_{BMB}\leq t\middle|\{f_t\}_{t=1}^T\right)-\Pr\left(\max_{1\leq i\leq\kappa}Y_j\leq t\right)\right|\leq C'P^{-c'}\tag{19}$$

with probability larger than $1-CP^{-c}$. Now, to prove Theorem 1, we proceed in two steps: first, we establish that:

$$\Pr\left(\left|\overline{U} - U_0\right| > \zeta_{P1}\right) \leq CP^{-c} \tag{20}$$

and

$$\Pr\left(\Pr\left(\left|W^{BMB} - \overline{W}^{BMB}\right| > \zeta''_{P1}\middle|\{f_t(\cdot)\}_{t=1}^{T}\right) > C'P^{-c'}\right) \leq C'P^{-c'}, \tag{21}$$

which deal with the estimation error in the test and the bootstrap statistic, respectively. Then, in a second step, we verify the first (only some moment inequalities are binding) and the second (all moment inequalities are binding) claim of Theorem 1.

We start with the expression in (20). Using Lemma 1, it follows that:

$$\left|\overline{U} - U_0\right| \leq \max_{1\leq i\leq \kappa}\left|\sqrt{P}\left(\frac{1}{P}\sum_{t=R+1}^{T} f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right)\right| \leq \zeta_{P1}$$

with probability $1 - CP^{-c}$. Next, we turn to Equation (21). Start again with:

$$\left|W^{BMB} - \overline{W}^{BMB}\right| \leq \max_{1\leq i\leq \kappa}\left|\frac{1}{\sqrt{m_P q_P}}\sum_{h=1}^{m_P}\varepsilon_h\sum_{t\in I_h}\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right) - \frac{1}{P}\sum_{t=R+1}^{T}\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right)\right|.$$

Conditional on $\{f_t(\cdot)\}_{t=1}^{T}$, the vector:

$$\frac{1}{\sqrt{m_P q_P}}\sum_{h=1}^{m_P}\varepsilon_h\sum_{t\in I_h}\left(\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right) - \frac{1}{P}\sum_{t=R+1}^{T}\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right)\right),$$

$1 \leq i \leq \kappa$ is normal with mean zero and all diagonal elements of the covariance matrix bounded by:

$$\max_{1\leq i\leq \kappa}\frac{1}{\sqrt{m_P q_P}}\sum_{h=1}^{m_P}\sum_{t\in I_h}\left(\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right) - \frac{1}{P}\sum_{t=R+1}^{T}\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right)\right)^2, \tag{22}$$

For the last expression, observe that:

$$\begin{aligned}
&\left(\frac{1}{\sqrt{m_P q_P}}\sum_{h=1}^{m_P}\sum_{t\in I_h}\left(\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right) - \frac{1}{P}\sum_{t=R+1}^{T}\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right)\right)^2\right)^{\frac{1}{2}} \\
\leq\ &\left(\frac{1}{\sqrt{m_P q_P}}\sum_{h=1}^{m_P}\sum_{t\in I_h}\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right)^2\right)^{\frac{1}{2}} + \sqrt{m_P q_P}\left|\frac{1}{P}\sum_{t=R+1}^{T}\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right)\right|
\end{aligned} \tag{23}$$

where the inequality follows by an application of Minkowski's inequality and the fact that:

$$\frac{1}{\sqrt{m_P q_P}} \sum_{h=1}^{m_P} \sum_{t \in I_h} \left(\frac{1}{P} \sum_{t=R+1}^{T} \Big(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i) \Big) \right)^2 = \sqrt{m_P q_P} \left(\frac{1}{P} \sum_{t=R+1}^{T} \Big(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i) \Big) \right)^2 .$$

The first term on the right hand side (RHS) of (23) is bounded by ζ_{P1} with probability larger than $1 - CP^{-c}$ by Lemma 2. The second term on the RHS of (23) is also bounded by ζ_{P1} with probability larger than $1 - CP^{-c}$ by Lemma 1 and the fact that $\sqrt{m_P q_P}/\sqrt{P} \longrightarrow 0$. Hence Equation (22) is bounded by $C\zeta_{P1}^2$ with probability $1 - CP^{-c}$. Then, by Markov's inequality (conditional on the data):

$$\Pr\left(\left| W^{BMB} - \overline{W}^{BMB} \right| > \zeta_{P1}^{\frac{1}{2}} \log^{\frac{1}{2}} \kappa \,\middle|\, \{f_t(\cdot)\}_{t=1}^{T} \right) \leq \frac{C\zeta_{P1} \log^{\frac{1}{2}} \kappa}{\zeta_{P1}^{\frac{1}{2}} \log^{\frac{1}{2}} \kappa} \leq CP^{-c},$$

which follows from Proposition 1.1.3 in Talagrand (2003) and the rate condition in Assumption 5.[22] The claim from Equation (21) follows by setting $\zeta''_{P1} = \zeta_{P1}^{\frac{1}{2}} \log^{\frac{1}{2}} \kappa$.

We now prove the claims of Theorem 1 starting with the case where $\mathrm{E}\Big[f_t(\boldsymbol{\theta}_i)\Big] < 0$ for at least some $1 \leq i \leq \kappa$. Under H_0 it holds that:

$$\begin{aligned}
\Pr\Big(U^* > c^{BMB}(\alpha) \Big) &\leq \Pr\Big(\overline{U} > c^{BMB}(\alpha) \Big) \\
&\leq \Pr\Big(U_0 > c^{BMB}(\alpha) - \zeta_{P1} \Big) + \Pr\Big(\Big| \overline{U} - U_0 \Big| > \zeta_{P1} \Big) \\
&\leq \Pr\Big(U_0 > c_0(\alpha + CP^{-c} + 8\zeta''_{P1} \log^{\frac{1}{2}} \kappa) - \zeta_{P1} \Big) + C'P^{-c'} \\
&\leq \Pr\Big(U_0 > c_0(\alpha + CP^{-c} + 16\zeta''_{P1} \log^{\frac{1}{2}} \kappa) \Big) + C'P^{-c'} \\
&\leq \Pr\Big(\max_{1 \leq j \leq \kappa} Y_j > c_0(\alpha + CP^{-c} + 16\zeta''_{P1} \log^{\frac{1}{2}} \kappa) \Big) + \rho_P^U + C'P^{-c'} \\
&= \alpha + CP^{-c} + 16\zeta''_{P1} \log^{\frac{1}{2}} \kappa + \rho_P^U + C'P^{-c'} \\
&\leq \alpha + CP^{-c},
\end{aligned}$$

where the second inequality holds since $0 < \zeta_{P1} \leq CP^{-c}$ and $\Pr\Big(\Big\{ U_0 + \Big(\overline{U} - U_0 \Big) > c^{BMB}(\alpha) \Big\} \cap$

[22] Proposition 1.1.3 in Talagrand (2003): For centered normal random variables ξ_i, $i = 1, \ldots, \kappa$ with $\max_{1 \leq i \leq \kappa} \mathrm{E}[\xi_i^2] < \infty$, it holds that:

$$\mathrm{E}\Big[\max_{1 \leq i \leq \kappa} \xi_i \Big] \leq \sqrt{2 \max_{1 \leq i \leq \kappa} \mathrm{E}[\xi_i^2] \log \kappa}.$$

$\left\{\overline{U} - U_0 > \zeta_{P1}\right\}\Big) \leq \Pr\Big(\Big\{\left|\overline{U} - U_0\right| > \zeta_{P1}\Big\}\Big)$, the third inequality follows from Equation (20) and the first claim of Lemma 3, the fourth inequality from Equation (33) in the proof of Lemma 3 below, the fifth inequality from Equation (17), while the equality follows from the fact that $\max_{1\leq j\leq \kappa}$ has no point masses. This establishes the first claim of Theorem 1. For the second claim, suppose that $\mathrm{E}\Big[f_t(\boldsymbol{\theta}_i)\Big)\Big] = 0$ for all $1 \leq i \leq \kappa$. Then, under H_0:

$$\begin{aligned}
\Pr\left(U^* > c^{BMB}(\alpha)\right) &= \Pr\left(\overline{U} > c^{BMB}(\alpha)\right) \\
&\geq \Pr\left(U_0 > c^{BMB}(\alpha) + \zeta_{P1}\right) - \Pr\left(\left|\overline{U} - U_0\right| > \zeta_{P1}\right) \\
&\geq \Pr\left(U_0 > c_0(\alpha - CP^{-c} - 8\zeta''_{P1}\log^{\frac{1}{2}}\kappa) + \zeta_{P1}\right) - C'P^{-c'} \\
&\geq \Pr\left(U_0 > c_0(\alpha - CP^{-c} - 16\zeta''_{P1}\log^{\frac{1}{2}}\kappa)\right) - C'P^{-c'} \\
&\geq \Pr\left(\max_{1\leq j\leq\kappa} Y_j > c_0(\alpha - CP^{-c} - 16\zeta''_{P1}\log^{\frac{1}{2}}\kappa)\right) - \rho_P^U - C'P^{-c'} \\
&= \alpha - CP^{-c} - 16\zeta''_{P1}\log^{\frac{1}{2}}\kappa - \rho_P^U - C'P^{-c'} \\
&\geq \alpha - CP^{-c},
\end{aligned}$$

where the first inequality holds again since $0 < \zeta_{P1} \leq CP^{-c}$ and $\Pr\Big(\Big\{U_0 + \Big(\overline{U} - U_0\Big) > c^{BMB}(\alpha)\Big\} \cap \left\{\overline{U} - U_0 > \zeta_{P1}\right\}\Big) \leq \Pr\Big(\Big\{\left|\overline{U} - U_0\right| > \zeta_{P1}\Big\}\Big)$, the second inequality follows from Equation (20) and the second claim of Lemma 3, the third inequality from the reverse of Equation (33) in the proof of Lemma 3 below, and the fourth inequality from Equation (17). Together with the previous set of inequalities, this establishes the second claim of Theorem 1. ■

Remark 1. *In Section S1 of the supplementary material, we show formally that the results of Theorem 7.1 (more specifically, Theorems B.1 and B.2 therein) of Chernozhukov et al. (2014) remain valid for unbounded random variables, which satisfy the assumptions stated in the main text and Assumption 4* in the supplement.*

Proof of Lemma 1. Applying Markov's inequality, the probability is bounded by:

$$\frac{\mathrm{E}\left[\max_{1\leq i\leq\kappa}\left|\frac{\sqrt{P}}{P}\sum_{t=R+1}^{T} f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right|\right]}{\zeta_{P1}}. \tag{24}$$

The numerator can be further bounded by:

$$\begin{aligned}
&\mathrm{E}\left[\sum_{i=1}^{\kappa}\left|\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T} f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right|\right] \\
\leq\ & \kappa \max_{1\leq i\leq\kappa} \mathrm{E}\left[\left|\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T} f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right|\right] \\
=\ & \kappa P^{\frac{1}{2}} \max_{1\leq i\leq\kappa} \mathrm{E}\left[\left| f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right|\right],
\end{aligned}$$

where the last equality follows from identical distribution. Note that the final expectation can written as:

$$\mathrm{E}\left[\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right)1_{\geq 0} + \left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right)1_{<0}\right], \tag{25}$$

where $1_{\geq 0} \equiv 1\left\{\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right) > 0\right\}$ is an indicator equal to one if the event $\left\{\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right) > 0\right\}$ is true and zero otherwise ($1_{<0}$ is defined analogously). We will examine only the first term of (25), the second one follows by identical arguments. Using the Cauchy Schwarz inequality and the fact that $1^2_{<0} = 1_{<0}$, the first term can be bounded by:

$$\left(\mathrm{E}\left[\left(f_t(\widehat{\boldsymbol{\theta}}_{it}) - f_t(\boldsymbol{\theta}_i)\right)^2\right]\right)^{\frac{1}{2}} \times \left(\mathrm{E}\left[1_{\geq 0}\right]\right)^{\frac{1}{2}} \equiv (I_P)^{\frac{1}{2}} \times (II_P)^{\frac{1}{2}}. \tag{26}$$

For I_P, a first order mean value expansion around $\boldsymbol{\theta}_i$ yields:

$$\mathrm{E}\left[(\widehat{\boldsymbol{\theta}}_{it} - \boldsymbol{\theta}_i)'\nabla_{\boldsymbol{\theta}_i} f_t(\boldsymbol{\theta}_i^*)'\nabla_{\boldsymbol{\theta}_i} f_t(\boldsymbol{\theta}_i^*)(\widehat{\boldsymbol{\theta}}_{it} - \boldsymbol{\theta}_i)\right], \tag{27}$$

where $\boldsymbol{\theta}_i^*$ is a vector of values on the line segment between $\widehat{\boldsymbol{\theta}}_{it}$ and $\boldsymbol{\theta}_i$. To analyze this term, we apply the linear representation from Assumption 3 in combination with a leave-one-block-out procedure ("Jackknife"), which omits the last $q_P + r_P$ observations prior to t for the rolling estimation of $\widehat{\boldsymbol{\theta}}_{it}$. That is:

$$\sqrt{R - q_P - r_P}(\widehat{\boldsymbol{\theta}}_{it} - \boldsymbol{\theta}_i) = \frac{1}{\sqrt{R - q_P - r_P}} \sum_{k=t-R+1}^{t-q_P-r_P} \boldsymbol{\Psi}_{ik} + O_p((R - q_P - r_P)^{-1}),$$

where $\boldsymbol{\Psi}_{ik}$ is of dimension $(m \times 1)$ satisfying $\mathrm{E}[\boldsymbol{\Psi}_{ik}] = \mathbf{0}$ and $\mathrm{E}[\boldsymbol{\Psi}_{ik}\boldsymbol{\Psi}_{ik}'] < \infty$. Multiplying the terms of the expansion by $\frac{(R-q_P-r_P)}{(R-q_P-r_P)}$, we can insert the linear representation from above to obtain for the expectation expression:

$$\frac{1}{(R-q_P-r_P)}\,\mathrm{E}\Bigg[\Bigg(\frac{1}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\boldsymbol{\Psi}_{ik}+O_p((R-q_P-r_P)^{-1})\Bigg)'$$
$$\times\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)'\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)\Bigg(\frac{1}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\boldsymbol{\Psi}_{ik}+O_p((R-q_P-r_P)^{-1})\Bigg)\Bigg],\quad(28)$$

which can be re-written as:

$$\frac{1}{R-q_P-r_P}\,\mathrm{E}\Bigg[\Bigg(\frac{1}{\sqrt{(R-q_P-r_P)}}\sum_{k=t-R+1}^{t-q_P-r_P}\boldsymbol{\Psi}_{ik}\Bigg)'\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)'$$
$$\times\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)\Bigg(\frac{1}{\sqrt{(R-q_P-r_P)}}\sum_{k=t-R+1}^{t-q_P-r_P}\boldsymbol{\Psi}_{ik}\Bigg)\Bigg]+O\Bigg(\frac{1}{(R-q_P-r_P)}\Bigg).$$

Concentrating on the lead expression, we only examine the first element of the sum:

$$\Bigg(\frac{1}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\boldsymbol{\Psi}_{ik}\Bigg)'\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)'\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)\Bigg(\frac{1}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\boldsymbol{\Psi}_{ik}\Bigg)$$

explicitly, the remaining terms (and in particular, the cross-products) will follow by analogous arguments. Thus:

$$\frac{1}{(R-q_P-r_P)}\,\mathrm{E}\Bigg[\Bigg(\frac{\nabla_{\boldsymbol{\theta}_i}f_t^{(1)}(\boldsymbol{\theta}_i^*)}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\Psi_{ik,1}\Bigg)\times\Bigg(\frac{\nabla_{\boldsymbol{\theta}_i}f_t^{(1)}(\boldsymbol{\theta}_i^*)}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\Psi_{ik,1}\Bigg)\Bigg],$$

where $\Psi_{ik,1}$ denotes again the first element of $\boldsymbol{\Psi}_{ik}$, and $\nabla_{\boldsymbol{\theta}_i}f_t^{(1)}(\cdot)$ the first element of the $(2m\times1)$ derivative vector. By independence of observations that lie at least q_P+r_P periods apart, we can re-write the expectation as:

$$\mathrm{E}\Bigg[\frac{1}{(R-q_P-r_P)}\Bigg\{\sum_{k=t-R+1}^{t-q_P-r_P}\Psi_{ik,1}^2+2\sum_{k=t-R+1}^{t-q_P-r_P}\sum_{s>k}\Psi_{ik,1}\Psi_{is,1}\Bigg\}\Bigg]\times\mathrm{E}\Bigg[\Bigg(\nabla_{\boldsymbol{\theta}_i}f_t^{(1)}(\boldsymbol{\theta}_i^*)\Bigg)^2\Bigg].\qquad(29)$$

Moreover, since the data is β-mixing (with coefficient b_{q+r}) and by the identical distribution, it holds that:

$$\Bigg|\mathrm{E}\Bigg[\Psi_{ik,1}\Psi_{is,1}\Bigg]\Bigg|\le b_{q+r}(k-s)\,\mathrm{E}\Bigg[\Psi_{ik,1}^2\Bigg]\qquad\forall s\neq k.$$

Then, using the fact that:

$$\sum_{k=t-R+1}^{t-q_P-r_P} \sum_{s>k} b_{q+r}(k-s) = (R-q_P-r_P) \sum_{j=1}^{\infty} b_{q+r}(j)$$

with $j = k - s$, we obtain for Equation (29):

$$\mathrm{E}\left[\frac{1}{(R-q_P-r_P)}\left\{(R-q_P-r_P)\,\mathrm{E}\left[\Psi_{i1,1}^2\right] + 2(R-q_P-r_P)\,\mathrm{E}\left[\Psi_{i1,1}^2\right]\sum_{j=1}^{\infty} b_{q+r}(j)\right\}\right] \times \mathrm{E}\left[\left(\nabla_{\boldsymbol{\theta}_i} f_t^{(1)}(\boldsymbol{\theta}_i^*)\right)^2\right].$$

Since $\sum_{j=1}^{\infty} b_{q+r}(j) < \infty$, it follows that this expression is of order $O(1)$. As a result, $(I_P)^{\frac{1}{2}}$ is of order $O((R-q_P-r_P)^{-\frac{1}{2}})$.

Next, we turn to II_P in Equation (26), which by iterated expectations (recall that $\widehat{\boldsymbol{\theta}}_{it}$ is estimated using the first $R - q_P - r_P$ observations) can be written as:

$$\mathrm{E}\left[F_{f_t(\cdot)}\left(f_t(\widehat{\boldsymbol{\theta}}_{it})\middle|\widehat{\boldsymbol{\theta}}_{it}\right)\right]. \tag{30}$$

Given Assumption 2, a first order mean value expansion of $F_{f_t(\cdot)}\left(f_t(\widehat{\boldsymbol{\theta}}_{it})\middle|\widehat{\boldsymbol{\theta}}_{it}\right)$ around $\boldsymbol{\theta}_i$ yields:

$$0 + \nabla_{\boldsymbol{\theta}_i} F_{f_t(\cdot)}\left(f_t(\boldsymbol{\theta}_i^*)\middle|\boldsymbol{\theta}_i^*\right)\nabla_{\boldsymbol{\theta}_i} f_t(\boldsymbol{\theta}_i^*)(\widehat{\boldsymbol{\theta}}_{it} - \boldsymbol{\theta}_i). \tag{31}$$

Using the linear representation from before, this expression can be written as:

$$\frac{1}{\sqrt{R-q_P-r_P}}\nabla_{\boldsymbol{\theta}_i} F_{f_t(\cdot)}\left(f_t(\boldsymbol{\theta}_i^*)\middle|\boldsymbol{\theta}_i^*\right)\nabla_{\boldsymbol{\theta}_i} f_t(\boldsymbol{\theta}_i^*)\left(\frac{1}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P} \boldsymbol{\Psi}_{ik} + O_p((R-q_P-r_P)^{-1})\right).$$

Re-inserting this term into Equation (30), invoking the identical distribution and Assumption 3 together with the fact that observations that lie at least $q_P + r_P$ periods apart are independent, we obtain

$$\begin{aligned} & \mathrm{E}\left[\nabla_{\boldsymbol{\theta}_i} F_{f_t(\cdot)}\left(f_t(\boldsymbol{\theta}_i^*)\middle|\boldsymbol{\theta}_i^*\right)\nabla_{\boldsymbol{\theta}_i} f_t(\boldsymbol{\theta}_i^*)\right] \times \mathrm{E}\left[\boldsymbol{\Psi}_{i(t-q_P-r_P)}\right] + O((R-q_P-r_P)^{-1}) \\ = \; & 0 + O((R-q_P-r_P)^{-1}). \end{aligned}$$

It thus follows that $(II_P)^{\frac{1}{2}}$ is of order $O((R-q_P-r_P)^{-\frac{1}{2}})$. Hence, the claim of the Lemma follows by the rate condition in Assumption 5. ■

Proof of Lemma 2. By Markov's inequality, we can bound the probability by:

$$\Pr\left(\max_{1\le i\le\kappa}\left(\frac{1}{m_P q_P}\sum_{h=1}^{m_P}\sum_{h\in I_h}\left(f_t(\widehat{\boldsymbol{\theta}}_{it})-f_t(\boldsymbol{\theta}_i)\right)^2\right)^{\frac{1}{2}}>\zeta_{P1}\right)$$

$$\le \frac{\mathrm{E}\left[\max_{1\le i\le\kappa}\left(\frac{1}{m_P q_P}\sum_{h=1}^{m_P}\sum_{t\in I_h}\left(f_t(\widehat{\boldsymbol{\theta}}_{it})-f_t(\boldsymbol{\theta}_i)\right)^2\right)^{\frac{1}{2}}\right]}{\zeta_{P1}}.$$

By the same arguments as in Lemma 1 we can bound the numerator on the RHS by:

$$\kappa\max_{1\le i\le\kappa}\mathrm{E}\left[\left(\frac{1}{m_P q_P}\sum_{h=1}^{m_P}\sum_{t\in I_h}\left(f_t(\widehat{\boldsymbol{\theta}}_{it})-f_t(\boldsymbol{\theta}_i)\right)^2\right)^{\frac{1}{2}}\right].$$

A first order mean value expansion around $\boldsymbol{\theta}_i$ yields:

$$\kappa\max_{1\le i\le\kappa}\mathrm{E}\left[\left(\frac{1}{m_P q_P}\sum_{h=1}^{m_P}\sum_{t\in I_h}(\widehat{\boldsymbol{\theta}}_{it}-\boldsymbol{\theta}_i)'\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)'\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)(\widehat{\boldsymbol{\theta}}_{it}-\boldsymbol{\theta}_i)\right)^{\frac{1}{2}}\right].$$

Using the linear representation from before, this equation yields:

$$\frac{\kappa}{\sqrt{(R-q_P-r_P)}}\max_{1\le i\le\kappa}\mathrm{E}\left[\left(\frac{1}{m_P q_P}\sum_{h=1}^{m_P}\sum_{t\in I_h}\left(\frac{1}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\boldsymbol{\Psi}_{ik}\right)'\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)'\right.\right.$$
$$\left.\left.\times\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)\left(\frac{1}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\boldsymbol{\Psi}_{ik}\right)\right)^{\frac{1}{2}}\right]+o(1).$$

We only examine the first element of the sum:

$$\left(\frac{1}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\boldsymbol{\Psi}_{ik}\right)'\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)'\nabla_{\boldsymbol{\theta}_i}f_t(\boldsymbol{\theta}_i^*)\left(\frac{1}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\boldsymbol{\Psi}_{ik}\right)$$

explicitly, the remaining terms (and in particular, the cross-products) will follow by analogous arguments. Thus:

$$\frac{\kappa}{\sqrt{(R-q_P-r_P)}}\max_{1\le i\le\kappa}\mathrm{E}\left[\left(\frac{1}{m_P q_P}\sum_{h=1}^{m_P}\sum_{t\in I_h}\left(\frac{\nabla_{\boldsymbol{\theta}_i}f_t^{(1)}(\boldsymbol{\theta}_i^*)}{\sqrt{R-q_P-r_P}}\sum_{k=t-R+1}^{t-q_P-r_P}\Psi_{ik,1}\right)\right.\right.$$

$$\times \left(\frac{\nabla_{\boldsymbol{\theta}_i} f_t^{(1)}(\boldsymbol{\theta}_i^*)}{\sqrt{R - q_P - r_P}} \sum_{k=t-R+1}^{t-q_P-r_P} \Psi_{ik,1} \right) \Bigg)^{\frac{1}{2}} \Bigg],$$

where $\Psi_{ik,1}$ denotes again the first element of $\boldsymbol{\Psi}_{ik}$, and $\nabla_{\boldsymbol{\theta}_i} f_t^{(1)}(\cdot)$ the first element of the $(2m \times 1)$ derivative vector. Applying Jensen's inequality and repeating the arguments of Lemma 1, we obtain:

$$\frac{\kappa}{\sqrt{(R - q_P - r_P)}} \max_{1 \le i \le \kappa} \mathrm{E}\Bigg[\Bigg(\frac{1}{m_P q_P} \sum_{h=1}^{m_P} \sum_{t \in I_h} \left(\frac{\nabla_{\boldsymbol{\theta}_i} f_t^{(1)}(\boldsymbol{\theta}_i^*)}{\sqrt{R - q_P - r_P}} \sum_{k=t-R+1}^{t-q_P-r_P} \Psi_{ik,1} \right)$$
$$\times \left(\frac{\nabla_{\boldsymbol{\theta}_i} f_t^{(1)}(\boldsymbol{\theta}_i^*)}{\sqrt{R - q_P - r_P}} \sum_{k=t-R+1}^{t-q_P-r_P} \Psi_{ik,1} \right) \Bigg)^{\frac{1}{2}} \Bigg] = O\left(\frac{\kappa}{(R - q_P - r_P)^{\frac{1}{2}}} \right).$$

By Assumption 5, this is of smaller order than $C_1 P^{-c_2}$.

■

Proof of Lemma 3. The proof of this Lemma follows by similar arguments as in the proof of Theorem 4.3 of Chernozhukov et al. (2014). That is, to establish the first claim, note that for any $x \in \mathbb{R}$:

$$\Pr\left(W^{BMB} \le x \,\middle|\, \{f_t(\cdot)\}_{t=1}^T \right) \le \Pr\left(\overline{W}^{BMB} \le x + \zeta''_{P1} \,\middle|\, \{f_t(\cdot)\}_{t=1}^T \right) + \Pr\left(\left| W^{BMB} - \overline{W}^{BMB} \right| > \zeta''_{P1} \,\middle|\, \{f_t(\cdot)\}_{t=1}^T \right), \tag{32}$$

where ζ''_{P1} was defined as $\zeta''_{P1} = \zeta_{P1}^{\frac{1}{2}} \log^{\frac{1}{2}} \kappa$ and the second part of the RHS follows since:

$$\Pr\left(\left\{ \overline{W}^{BMB} + \left(W^{BMB} - \overline{W}^{BMB} \right) \le x \right\} \cap \left\{ \left| W^{BMB} - \overline{W}^{BMB} \right| > \zeta''_{P1} \right\} \,\middle|\, \{f_t(\cdot)\}_{t=1}^T \right)$$
$$\le \Pr\left(\left| W^{BMB} - \overline{W}^{BMB} \right| > \zeta''_{P1} \,\middle|\, \{f_t(\cdot)\}_{t=1}^T \right).$$

The RHS of Equation (32) can be further bounded by:

$$\Pr\left(\max_{1 \le j \le \kappa} Y_j \le x + \zeta''_{P1} \right) + \rho_P + \Pr\left(\left| W^{BMB} - \overline{W}^{BMB} \right| > \zeta''_{P1} \,\middle|\, \{f_t(\cdot)\}_{t=1}^T \right),$$

where ρ_P was defined in Equation (19). Moreover, by the same arguments as in the proof of Theorem

4.3, for any $\gamma \in (0, 1 - 8\zeta''_{P1} \log^{\frac{1}{2}} \kappa)$ and sufficiently large P one can show that:

$$\Pr\left(\max_{1 \leq j \leq \kappa} Y_j \leq c_0(\gamma + 8\zeta''_{P1} \log^{\frac{1}{2}} \kappa) + \zeta''_{P1} \right) \leq 1 - \gamma$$

and thus:

$$c_0(\gamma + 8\zeta''_{P1} \log^{\frac{1}{2}} \kappa) + \zeta''_{P1} \leq c_0(\gamma), \tag{33}$$

which follows since $\max_{1 \leq j \leq \kappa} Y_j$ has no point masses. Now, setting $x = c_0(\alpha + CP^{-c} + 8\zeta''_{P1} \log^{\frac{1}{2}} \kappa)$ in Equation (32), it holds that:

$$\begin{aligned} & \Pr\left(W^{BMB} \leq c_0(\alpha + CP^{-c} + 8\zeta''_{P1} \log^{\frac{1}{2}} \kappa) \middle| \{f_t(\cdot)\}_{t=1}^T \right) \\ \leq\ & 1 - \alpha - CP^{-c} + \rho_P + \Pr\left(\left| W^{BMB} - \overline{W}^{BMB} \right| > \zeta''_{P1} \middle| \{f_t(\cdot)\}_{t=1}^T \right) \\ \leq\ & 1 - \alpha \end{aligned}$$

on the events $\rho_P \leq CP^{-c}$ and $\Pr\left(\left| W^{BMB} - \overline{W}^{BMB} \right| > \zeta''_{P1} \middle| \{f_t(\cdot)\}_{t=1}^T \right) \leq CP^{-c}$, which hold with probability $1 - CP^{-c}$ by Equations (19) and (21). Since $CP^{-c} + 8\zeta''_{P1} \log^{\frac{1}{2}} \kappa \leq C'P^{-c'}$, this implies the first claim. The second claim follows by analogous arguments starting with:

$$\begin{aligned} \Pr\left(W^{BMB} \leq x \middle| \{f_t(\cdot)\}_{t=1}^T \right) \ \geq\ & \Pr\left(\overline{W}^{BMB} \leq x - \zeta''_{P1} \middle| \{f_t(\cdot)\}_{t=1}^T \right) \\ & - \Pr\left(\left| W^{BMB} - \overline{W}^{BMB} \right| > \zeta''_{P1} \middle| \{f_t(\cdot)\}_{t=1}^T \right), \end{aligned}$$

and mirroring the steps from before.

■

References

Antolín Díaz, J., T. Drechsel, and I. Petrella (2016). Tracking the Slowdown in Long-Run GDP Growth. *Available at SSRN: http://ssrn.com/abstract=2762383*.

Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting 20*(3), 447–460.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-Casting and the Real-Time

Data Flow. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2A. North-Holland.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Nowcasting. *European Central Bank Working Paper 1275*.

Bańbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*(1), 133–160.

Bell, V., L. W. Co, S. Stone, and G. Wallis (2014). Nowcasting UK GDP growth. *Bank of England Quarterly Bulletin 2014 Q4*, 58–68.

Bernstein, S. (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen 97*(1), 1–59.

Bühlmann, P. and H. Künsch (1999). Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis 31*, 295–310.

Bulligan, G., M. Marcellino, and F. Venditti (2015). Forecasting economic activity with targeted predictors. *International Journal of Forecasting 31*(1), 188–206.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Testing many moment inequalities. cemmap Working Paper CWP65/13, Massachusetts Institute of Technology.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*(5), 1243–1284.

Clark, T. E. and M. W. McCracken (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics 105*(1), 85–110.

Clark, T. E. and M. W. McCracken (2005). Evaluating direct multistep forecasts. *Econometric Reviews 24*(4), 369–404.

Clements, M. P. and A. B. Galvão (2008). Macroeconomic Forecasting With Mixed-Frequency Data. *Journal of Business & Economic Statistics 26*(4), 546–554.

Corradi, V. and W. Distaso (2011). Multiple Forecast Model Evaluation. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting*, pp. 391–413. New York: Oxford University Press.

Corradi, V. and N. R. Swanson (2006). Predictive Density and Conditional Confidence Interval Accuracy Tests. *Journal of Econometrics 135*(1-2), 187–228.

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

Ellison, G. and S. F. Ellison (2011). Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics 3*, 1–36.

Foroni, C. and M. Marcellino (2014). A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates. *International Journal of Forecasting 30*(3), 554–568.

Fosten, J. (2016). Forecast Evaluation with Factor-Augmented Models. *UEA School of Economics Working Paper 2016-05*.

Ghosal, S., A. Sen, and A. W. van der Vaart (2000). Testing monotonicity of regression. *Annals of Statistics 28*(4), 1054–1092.

Ghysels, E., A. Sinko, and R. Valkanov (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews 26*(1), 53–90.

Giacomini, R. and H. White (2006). Tests of Conditional Predictive Ability. *Econometrica 74*(6), 1545–1578.

Gonçalves, S., M. W. McCracken, and B. Perron (2015). Tests of Equal Accuracy for Nested Models with Estimated Factors. *Federal Reserve Bank of St. Louis Working Paper 2015-025A*.

Gutknecht, D. (2016). Testing monotonicity under endogeneity. *Journal of Econometrics 190*, 100–114.

Hall, P. and J. Heckman (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Annals of Statistics 28*, 20–39.

Hall, P., J. Horowitz, and B. Jing (1995). On blocking rules for the bootstrap with dependent data. *Biometrika 82*, 561–574.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The Model Confidence Set. *Econometrica 79*(2), 453–497.

Higgins, P. C. (2014). GDPNow: A Model for GDP 'Nowcasting'. *FRB Atlanta Working Paper 2014-7*.

Jin, S., V. Corradi, and N. R. Swanson (2016). Robust forecast comparison. *Available at SSRN 2605927*.

Newey, W. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics 5*(2), 99–135.

Patton, A. J. and A. Timmermann (2007). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics 140*(2), 884–918.

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

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Rünstler, G., K. Barhoum, S. Benk, R. Cristadoro, a. D. E. N. Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. V. a. N. Nieuwenhuyze (2009). Short-term forecasting of GDP using large datasets: A pseudo real-time forecast evaluation exercise. *Journal of Forecasting 28*(7), 595–611.

Schumacher, C. (2016). A comparison of MIDAS and bridge equations. *International Journal of Forecasting 32*(2), 257–270.

Talagrand, M. (2003). *Spin Glasses: A Challenge for Mathematicians*, Volume 46 of *A Series of Modern Surveys in Mathematics*. Springer Press.

West, K. D. (1996). Asymptotic Inference about Predictive Ability. *Econometrica 64*(5), 1067–1084.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica 68*(5), 1097–1126.

Zhang, X. and G. Cheng (2014). Bootstrapping high dimensional time series. Working paper, University of Missouri-Columbia.