Testing Nowcast Monotonicity with Estimated Factors*

Jack Fosten† Daniel Gutknecht‡

March 15, 2018

Abstract

This paper proposes a test to determine whether 'big data' nowcasting methods, which have become an important tool to many public and private institutions, are monotonically improving as new information becomes available. The test is the first to formalise existing evaluation procedures from the nowcasting literature. We place particular emphasis on models involving estimated factors, since factor-based methods are a leading case in the high-dimensional empirical nowcasting literature, although our test is still applicable to small-dimensional set-ups like bridge equations and MIDAS models. Our approach extends a recent methodology for testing many moment inequalities to the case of nowcast monotonicity testing, which allows the number of inequalities to grow with the sample size. We provide results showing the conditions under which both parameter estimation error and factor estimation error can be accommodated in this high dimensional setting when using the pseudo out-of-sample approach. The finite sample performance of our test is illustrated using a wide range of Monte Carlo simulations, and we conclude with an empirical application of nowcasting U.S. real gross domestic product (GDP) growth and five GDP sub-components. Our test results confirm monotonicity for all but one sub-component (government spending), suggesting that the factor-augmented model may be misspecified for this GDP constituent.

JEL Classification: C12, C22, C52, C53
Keywords: Nowcasting, Factor Models, Moment Inequalities, Bootstrap

*We are grateful to the editor Todd Clark, an Associate Editor, and two anonymous referees, as well as to Valentina Corradi, Rob Sollis, Ana Galvão, and Matthias Hartmann, and seminar participants at Newcastle University Business School, the 2016 CMF-CEF Seville meeting, the SNDE 2017 meeting in Paris, Lund University, Mannheim University and the 70th European Meeting of the Econometric Society in Lisbon 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.

Electronic copy available at: https://ssrn.com/abstract=2815794

1 Introduction

Nowcasting methods have been introduced as a way to provide timely predictions of the current state of the economy. This is important as key policy variables such as GDP are typically published at a low frequency and with a significant publication lag. These methods address important irregularities of macroeconomic datasets, specifically the problems of mixed frequency predictors and the so-called 'ragged edge' problem of the staggered release of data series by different statistical agencies. One of the most desirable features of nowcasting methods is that they can be used to revise predictions of the target variable multiple times per quarter, as soon as any new piece of information becomes available. However, while there has been a significant body of literature devoted to developing new nowcasting methods, there has been very little work on how to formally assess their performance.

Recent empirical nowcasting studies indicate that the evaluative criteria of nowcast methods have moved in a different direction to the traditional forecast evaluation testing literature which dates back to Diebold and Mariano (1995) and West (1996). Since a major advantage of nowcasting methods is to use big data in making very timely predictions of the economy, as noted in Bańbura et al. (2013), it is not as relevant to assess whether a nowcasting method is capable of outperforming naïve benchmarks like autoregressions, or surveys of professional forecasters, which potentially only produce one prediction per quarter. As such, it has become common for nowcasters to check whether nowcasts based on a particular method are monotonically improving as new information is added, using statistics like mean squared forecast error (MSFE). Nowcast monotonicity has been used as an evaluative criterion at policy-making institutions such as the Atlanta Fed and in empirical papers dating back to Giannone et al. (2008) and many more recently (see for example Bańbura et al., 2013, Aastveit et al., 2014, Luciani and Ricci, 2014, Marcellino et al., 2016, Knotek and Zaman, 2017, Bragoli and Fosten, 2018).

Nowcast monotonicity is also an interesting evaluative criterion as it has links to the literature on forecaster rationality. In the context of long-horizon forecasting, Patton and Timmermann (2012) demonstrate that if forecasters are rational and employ a correct model specification, and in the absence of measurement error, then MSFE declines with the forecast horizon. In the nowcasting context, this would imply that MSFE decreases when more information is added and the nowcast horizon shrinks up until the publication date of the target variable. By providing a test for nowcast monotonicity the results may thus be interpreted as a check for correct nowcast model specification. Alternatively, if the test is used to evaluate externally-made predictions taken from institutional nowcasters, the test can take on the interpretation as a test for 'nowcaster rationality'.

Electronic copy available at: https://ssrn.com/abstract=2815794

In this paper, we provide a formal and robust test for nowcast monotonicity based on the many moment inequalities procedure of Chernozhukov et al. (2014). Our test is the first rigorous procedure to assess nowcast performance, and can be used in very general settings. This improves upon the majority of the aforementioned empirical studies, such as Marcellino et al. (2016), which simply use graphical evidence of declining nowcast MSFE and do not formally test for monotonicity. It also improves on some existing papers that do test for monotonicity, such as Bańbura et al. (2013) who apply the test of Patton and Timmermann (2012), which are only set up to deal with a small number of moment inequalities. Our paper, instead, allows for high-dimensional sets of moment inequalities driven by the recent interest in 'big data' nowcasting methods.

Relative to the existing econometric literature of forecast evaluation testing, our paper provides several new contributions: our first and main contribution is that we allow for the possibility that the number of nowcasts per quarter, S, grows to infinity ($S \to \infty$) and therefore so does the number of moment inequalities. That is, while our method is still applicable in finite-dimensional nowcast model approaches such as bridge equations (Schumacher, 2016) or MIDAS models (Clements and Galvão, 2008, 2009) where the number of moment comparisons is typically small (relative to the sample size), it additionally accommodates big data approaches such as the use of factor models which have become widely used for nowcasting (Bańbura et al., 2013, Foroni and Marcellino, 2014, Bańbura and Modugno, 2014 and others). Specifically, we extend the many moment inequalities framework of Chernozhukov et al. (2014), originally geared towards microeconomic applications, to the case of nowcasting. In doing so, we focus on the case of factor-based nowcasting methods as the issue of factor estimation in nowcast evaluation is of separate econometric interest.

A second contribution of the paper is that we consider the issue of parameter as well as factor estimation error. We deal with parameter estimation error resulting from the use of the pseudo out-of-sample approach of West (1996), and derive rate conditions under which it does not affect the validity of critical values from the high-dimensional moment inequality testing procedure of Chernozhukov et al. (2014). More specifically, we find that, in allowing for $S \to \infty$, there is a cost to pay through a stronger rate condition relative to papers such as White (2000). Likewise, we provide the conditions required for using factor-augmented models of Stock and Watson (2002a,b) and Bai and Ng (2006) in this nowcast monotonicity test, with the factors estimated by Principal Components Analysis (PCA), after solving the issue of the 'ragged edge'. This is important as existing papers on the estimation of dynamic factor models with missing observations, such as the Kalman filtering approach of Doz et al. (2011), have focused on the properties of the factor estimates themselves and give no guidance

in our nowcast evaluation context. In a similar way to the case of parameter estimation error, we find that the rates of Bai and Ng (2006) have to be tightened in order for factor estimation error not to contribute asymptotically to the critical values of the test.

Finally, we investigate the performance of our test in finite samples through an extensive range of Monte Carlo simulations. These allow for non-monotonicity to occur from different forms of model misspecification and investigate sensitivity to parameter and factor estimation error, as well as varying the number of moment inequalities and out-of-sample splits. Our findings suggest that our test performs well across different scenarios, even in the presence of estimation error. We then apply our test to nowcasting the aggregate GDP growth rate in the United States, as well as five GDP subcomponents, using a data specification similar to that of Bańbura and Modugno (2014). As a preview of the results, our test confirms that there is no statistical evidence of non-monotonicity in the aggregate GDP growth rate. This is in line with previous studies such as Bańbura et al. (2013), which use a different factor estimation procedure and time period, and seems to indicate that the finding of monotonicity is robust. On the other hand, in the government spending sub-component of GDP our test finds significant violations of monotonicity. This suggests that a different model should be employed for nowcasting government spending. The results are very robust to different data spans, dataset configurations and estimation schemes.

The rest of the paper is organised as follows. Section 2 describes the nowcast monotonicity set-up of the paper, and provides details of the factor-augmented model set-up we consider in the theoretical results. Section 3 contains an introduction to the the test statistic, defines the bootstrap critical values and states their formal validity. Section 4 provides the Monte Carlo simulation settings and results. Section 5 gives the empirical application of the test. Finally, Section 6 concludes the paper. There is also a separate document of Supplementary Material which details the proofs of the results of the paper and provides additional Monte Carlo and empirical results not presented in the main text.

2 Set-up

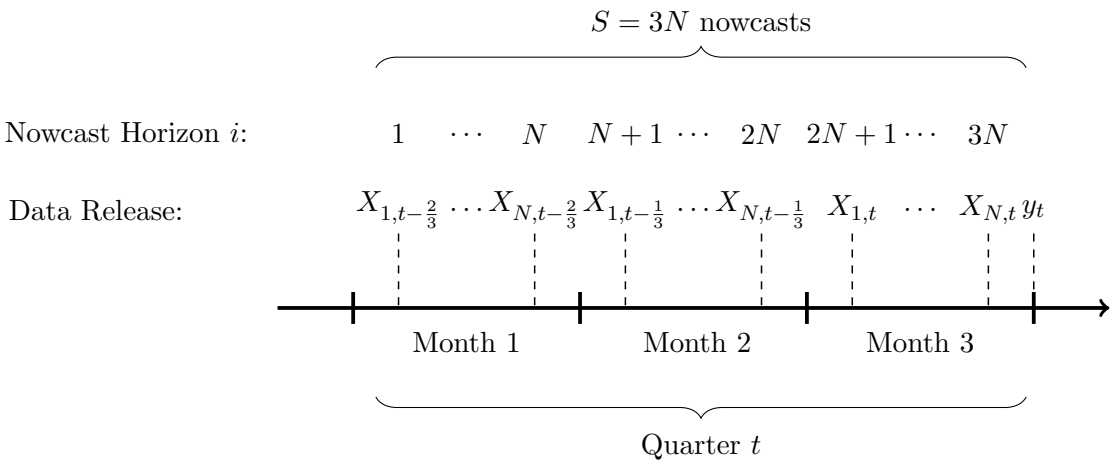
2.1 Testing Nowcast Monotonicity

The objective is to evaluate nowcasts of a low-frequency target variable, y_t, with T observations at quarterly periods $t = 1, ..., T$. Since data for y_t is not published until after the end of quarter t, we make a set of S nowcasts at intervals $i = 1, ..., S$, starting at the beginning of quarter t and ending when y_t is published. This is achieved using a large set of N candidate predictors X_{jt}, for $j = 1, ..., N$,

which are observed at a higher frequency than y_t. As an example, X_{jt} are taken to be monthly predictors which are observed at $3T$ monthly periods $t = 1/3, 2/3, 1, ..., T - 1/3, T$. The extension to higher frequencies is straightforward.

The S different nowcasting horizons are brought about by tracking the arrival of new information in real-time, or the 'data flow' as termed by Bańbura et al. (2013). Assuming that the observations X_{jt} are released sequentially in the order of their j variable index and then in the order of their t time index, the timeline of data releases appears as in Figure 1.

Figure 1: Graphical Illustration of the Data Flow



Several comments are useful here. Firstly, this example simplifies the set-up without loss of generality so that y_t is released at the end of quarter t, whereas in the empirical application we see, for instance, that U.S. GDP data for quarter t are, in practice, released some way into quarter $t + 1$. Secondly, as we move from $i = 1$ through $i = S$, the nowcast horizon shrinks as we approach the fixed publication date of y_t. This is a slightly different set-up to forecasting studies where the forecast origin is held fixed and the horizon is increased into the future. Finally, it can be seen that the so-called "ragged edge" occurs in the middle of this timeline, where the data are available for some months of the first few variables but are not available for the remainder.

In the above example, the monthly set-up gives rise to a total number of $S = 3N$ different nowcasts per quarter. This number of nowcasts can be made arbitrarily large by increasing the number of variables, N. This is particularly relevant for recent "big data" approaches to nowcasting which often allow $N \to \infty$. The value of S can also be very large when daily predictors are used, or even higher frequency financial series. On the other hand, S could be relatively small if the nowcaster chooses to use one monthly predictor and make $S = 3$ nowcasts per quarter. Additionally, some empirical

nowcasting papers choose a small number of cut-off dates (such as the end of each month) and make only $S = 3$ nowcasts based on models which aggregate all of the available data for the previous month. In this paper, we wish to allow for any of these scenarios, and will therefore allow a framework where $S \to \infty$.

We define $\Omega_{t-1+i/S}$ for $i = 1, ..., S$ as the collection of data releases at a point i during quarter t, which culminates with Ω_t at the end of quarter t where $i = S$ and all of the quarter's data points have been released. This information set corresponds to all data releases on the left of a given point in the timeline in Figure 1. Then at each point i the nowcaster makes a nowcast based on the available data at point i and an $(m_i \times 1)$ vector of parameters θ_i. We denote this nowcast, which is usually the conditional mean $\mathrm{E}\left[y_t|\Omega_{t-1+i/S}; \theta_i\right]$, by $y_{i,t}\left(\theta_i\right)$ in the following.

Given the motivation in the introduction, the central hypothesis of interest is whether nowcast performance is monotonically improving as we move through the quarter, approaching the publication date of the target variable. Given some loss function $L(.)$, whose properties will be discussed in Section 3.2 below, we are interested in knowing whether the nowcast error loss at a point $i + k$ is lower than some earlier point i. The null hypothesis is formed of $S(S-1)/2$ moment inequalities for each pairwise comparison of nowcast points $i + k$ and i:

$$H_0 : \mathrm{E}\left[L(y_t - y_{i+k,t}(\theta_{i+k})) - L(y_t - y_{i,t}(\theta_i))\right] \leq 0 \qquad \text{for all } i = 1, ..., S-1; \quad k = 1, \ldots, S-i \quad (1)$$

versus:

$$H_1 : \mathrm{E}\left[L(y_t - y_{i+k,t}(\theta_{i+k})) - L(y_t - y_{i,t}(\theta_i))\right] > 0 \qquad \text{for some } i = 1, ..., S-1; \quad k = 1, \ldots, S-i$$

The null hypothesis is violated when at least one point later in the prediction period has larger loss than some earlier horizon.

Relative to papers such as Patton and Timmermann (2012) for testing forecast monotonicity, we consider all possible $S(S-1)/2$ pairwise moment inequalities for the test, rather than just adjacent inequalities, in order to detect any violation of H_0. Note that, by doing so, even in empirical applications where the number of nowcast points is relatively small, $S(S-1)/2$ can be very large which further motivates a high-dimensional set-up. We leave the functions $L(.)$ and $y_{i,t}\left(\theta_i\right)$ to be very general in order to incorporate the vast majority of commonly used nowcasting methods, though it is very typical that the MSFE loss function $L(x) = x^2$ is used for nowcast evaluation. For $y_{i,t}\left(\theta_i\right)$, as mentioned in the introduction, we can easily incorporate different methods operating with a 'small'

or a 'large' number of regressors.

Since the population parameters θ_{i+k} 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 P quarterly out-of-sample periods $t = R+1, ..., T$, we make $i = 1, ..., S$ different nowcasts of y_t. We start out estimating the parameters with the first R observations, and then for each t and i, we use an expanding window of observations using the information set $\{\Omega_{s-1+i/S}\}_{s=1}^{t}$, known as the recursive scheme. We focus on the recursive rather than the rolling scheme (where the estimation window is fixed at length R), though the extension of our theoretical results to the latter is straightforward. The focus on the recursive scheme is partly as it is widely used in empirical studies and partly as we use factor models which are known to be quite robust to structural instabilities, with the latter being one of the main reasons for using the rolling scheme (Clark and McCracken, 2009).

We take the ragged edge at point i into account, and use only the information available at point i to estimate $\widehat{\theta}_{it}$. Therefore, in comparing horizons i and $i+k$, we use sequences of parameter estimates $\{\widehat{\theta}_{it}\}_{t=R+1}^{T}$ and $\{\widehat{\theta}_{i+k,t}\}_{t=R+1}^{T}$. This gives rise to a sample analogue of the moment conditions considered in H_0:

$$\frac{1}{P}\sum_{t=R+1}^{T}\Big[L(y_t - y_{i+k,t}(\widehat{\theta}_{i+k,t})) - L(y_t - y_{i,t}(\widehat{\theta}_{it}))\Big] \qquad \text{for } i = 1, ..., S-1; \quad k = 1, \ldots, S-i, \tag{2}$$

which we will use in Section 3 below for our test statistic.

2.2 Nowcasting with Factor-Augmented Models

In forming the nowcast loss differentials defined above, of particular interest is when we make nowcasts of the target variable y_t using the factor augmented model of Stock and Watson (2002a,b) and Bai and Ng (2006), adapted to the case of nowcasting. The simplest version is the unrestricted factor-MIDAS model used by Kim and Swanson (2017) amongst others:

$$y_t = \gamma' W_t + \beta_0' F_t + \beta_1' F_{t-1/3} + \beta_2' F_{t-2/3} + \varepsilon_t \qquad t = 1, 2, ..., T \tag{3}$$

This is a quarterly regression model of y_t on W_t, which is is a set of 'must-have regressors' such as a constant and lags of y_t, and monthly lags of F_t, the r unobserved factors based on the monthly static factor model:

$$X_t = \Lambda F_t + u_t \qquad t = 1/3, 2/3, 1, ..., T-2/3, T-1/3, T \tag{4}$$

where X_t is an $N \times 1$ vector of observed variables, Λ is an $N \times r$ matrix of factor loadings, F_t is an $r \times 1$ vector of unknown factors and u_t is an $N \times 1$ vector of idiosyncratic disturbances. Note that for the theoretical results of the paper we treat the number of factors, r, as known whereas in practice this can be consistently estimated by the information criteria of Bai and Ng (2002) and others. We also abstract from structural breaks in these models, as is common in the forecast evaluation literature. In the empirical application we explore the results using data spans guided by factor loading break dates, which can be determined using methods such as Breitung and Eickmeier (2011) and Corradi and Swanson (2014).

We make out-of-sample nowcasts for quarters $t = R+1, ..., T$ and nowcast points $i = 1, ..., S$ by repeatedly estimating the factors by PCA in expanding monthly windows from observation $1/3$ until the end of the quarter t, at the various nowcast points i. This proceeds in three stages.

In the first stage, we take account of the "ragged-edge" which occurs at the end of the recursive window estimation sample. Specifically, at points $t = R+1, ..., T$, we use a $3t \times N$ window of monthly observations in the matrix $X^{(t)} = [X_{1/3}, X_{2/3}, ..., X_t]'$, where the ragged edge only occurs in the last 3 monthly observations of each window. For example, if $i \leq N$, then only information is available for month 1 of quarter t up to variable i, and no other information is available for the rest of the quarter. In that case, the panel is unbalanced and we cannot directly apply PCA to estimate the factors. We therefore make predictions of the missing observations up to month t and variable N which gives a recursive data matrix:

$$\underset{3t \times N}{\widehat{X}^{(i,t)}} = \begin{bmatrix} X_{1,1/3} & \dots & \dots & \dots & \dots & X_{N,1/3} \\ \vdots & & & & & \vdots \\ X_{1,t-2/3} & \dots & X_{i,t-2/3} & \widehat{X}_{i+1,t-2/3} & \dots & \widehat{X}_{N,t-2/3} \\ \widehat{X}_{1,t-1/3} & \dots & \widehat{X}_{i,t-1/3} & \widehat{X}_{i+1,t-1/3} & \dots & \widehat{X}_{N,t-1/3} \\ \widehat{X}_{1t} & \dots & \widehat{X}_{it} & \widehat{X}_{i+1,t} & \dots & \widehat{X}_{Nt} \end{bmatrix} \text{ for } t = R+1, ..., T; i \leq N \qquad (5)$$

with typical element $\{\widehat{X}_j^{(i,t)} : t = R+1, ..., T; j = 1/3, 2/3, 1, ..., t-2/3, t-1/3, t; i = 1, ..., S\}$. Analogously, when $N < i \leq 2N$, the predicted observations only appear in the second-to-last and last row of $\widehat{X}^{(i,t)}$, and when $i > 2N$ they only appear in the last row. This data matrix differs from the out-of-sample factor estimation approaches of Gonçalves et al. (2017) and Fosten (2016) due to these predicted observations in the last three rows to solve the ragged edge. There have been different suggestions of how to balance the panel in the literature. Stock and Watson (2002b) suggest to use an

EM algorithm to fill in missing observations. A state space approach was suggested by Bańbura and Modugno (2014) using the Kalman filtering method of Doz et al. (2011). Other alternatives include using interpolations to predict missing observations before using PCA (Kim and Swanson, 2017) and vertical alignment (Altissimo et al., 2010).

In the second stage, we estimate the factors by PCA on the data matrix in Equation (5). For every out-of-sample window $t = R+1, ..., T$ and every nowcast point $i = 1, ..., S$ the PCA optimisation problem is:

$$\left(\underset{3t\times r}{\widehat{F}^{(i,t)}}, \underset{N\times r}{\widehat{\Lambda}^{(i,t)}}\right) = \arg\min_{\Lambda,F}\left\{\frac{1}{3tN}\text{tr}\left(\widehat{X}^{(i,t)} - F^{(t)}\Lambda'\right)\left(\widehat{X}^{(i,t)} - F^{(t)}\Lambda'\right)'\right\} \tag{6}$$

subject to the normalizations that $F^{(t)\prime}F^{(t)}/3t = I_r$ for all t and that $\Lambda'\Lambda/N$ is diagonal, where $F^{(t)} = [F_{1/3}, ..., F_t]'$ is a $3t \times r$ recursive matrix of monthly factors in each quarter t, in an analogous way to $X^{(t)}$. The solution is to set $\widehat{F}^{(i,t)}$ as the r eigenvectors corresponding to the r largest eigenvalues of the $3t \times 3t$ covariance matrix $\widehat{X}^{(i,t)}\widehat{X}^{(i,t)\prime}/3tN$. These eigenvalues are denoted by the $r \times r$ diagonal matrix $\widehat{V}^{(i,t)}$ for each i and t. Then since $\widehat{F}^{(i,t)\prime}\widehat{F}^{(i,t)}/3t = I_r$ under the normalisation for the factors, we obtain a simple expression for the estimated loadings, $\widehat{\Lambda}^{(i,t)} = \widehat{X}^{(i,t)\prime}\widehat{F}^{(i,t)}/3t$. The difference to existing papers on factor estimation is the additional dependence of the estimated factors and loadings on i, the nowcast point.

The final stage is to make the nowcast. For this we obtain OLS estimates using recursive regressions of y_t onto W_t and the estimated factors $\{\widehat{F}_j^{(i,t)} : t = R+1, ..., T; j = 1/3, 2/3, ..., t; i = 1, ..., S\}$. For unrestricted factor-MIDAS model as in Equation (3), the estimated factors are skip-sampled from the monthly frequency and inserted into the regression in a quarterly fashion. In other words, $\widehat{F}_t^{(i,t)}$ contains periods $j = 1, 2, \ldots, t$, while $\widehat{F}_{t-1/3}^{(i,t)}$ contains observations on the second month of every quarter (periods $j = 2/3, 5/3, \ldots, t-1/3$) and $\widehat{F}_{t-2/3}^{(i,t)}$ on the first month (periods $j = 1/3, 4/3, \ldots, t-2/3$). Then, collecting $\widehat{\boldsymbol{F}}_t^{(i,t)} = (\widehat{F}_t^{(i,t)}, \widehat{F}_{t-1/3}^{(i,t)}, \widehat{F}_{t-2/3}^{(i,t)})'$ and $\widehat{\boldsymbol{\beta}}_{it} = (\widehat{\beta}_{0it}', \widehat{\beta}_{1it}', \widehat{\beta}_{2it}')'$, the nowcasts are written:

$$y_{i,t}\left(\widehat{\boldsymbol{F}}_t^{(i,t)}, \widehat{\boldsymbol{\beta}}_{it}, \widehat{\gamma}_{it}\right) = \widehat{\gamma}_{it}'W_t + \widehat{\beta}_{0it}'\widehat{F}_t^{(i,t)} + \widehat{\beta}_{1it}'\widehat{F}_{t-1/3}^{(i,t)} + \widehat{\beta}_{2it}'\widehat{F}_{t-2/3}^{(i,t)} \tag{7}$$

for $t = R+1, ..., T$ and $i = 1, ..., S$. These nowcasts are then used directly in Equation (2) to calculate the loss differential series.

Note that, in the absence of measurement error, if we have correctly specified the model for the data generating process, we expect the null hypothesis of monotonicity to hold. Heuristically this is so because, as i grows towards S and we add more and more information, the predictions used to

solve the ragged edge in $\widehat{X}^{(i,t)}$ are increasingly replaced with actual observations. This implies that the effect of the factors being contaminated with ragged edge predictions vanishes throughout the prediction period, meaning that measures like MSFE are expected to decline as we move towards the end of the nowcast period.

3 Test Statistic and Critical Values

3.1 Test Statistic

To test the null hypothesis H_0 in Equation (1) we use a statistic which is based on the empirical moment inequalities introduced in Equation (2). That is, the test statistic is chosen to be a max statistic of the following form:

$$U^* = \max_{i,k \in \mathcal{C}_S} \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left[L(y_t - y_{i+k,t}(\widehat{\theta}_{i+k,t})) - L(y_t - y_{i,t}(\widehat{\theta}_{it})) \right], \tag{8}$$

where the max could be taken over different sets of moment inequality comparisons, for example the set of all possible $S(S-1)/2$ pairwise comparisons $\mathcal{C}_S = \{i, k : i = 1, \ldots, S-1, k = 1, \ldots, S-i\}$, or the set of only adjacent comparisons $\mathcal{C}_S = \{i, k : i = 1, \ldots, S-1, k = i+1\}$. The null hypothesis of monotonicity, H_0, in Equation (1) is then 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 the critical values, there is a remark to be made with respect to the set of moment inequalities to be used. Note that in order to apply the results of Chernozhukov et al. (2014) formally we need that for every given i and k combination, the variance of $L(y_t - y_{i+k,t}(\theta_{i+k})) - L(y_t - y_{i,t}(\theta_i))$, i.e. $v^2_{i+k,i} \equiv \text{Var}\left(L(y_t - y_{i+k,t}(\theta_{i+k})) - L(y_t - y_{i,t}(\theta_i))\right)$, is bounded away from zero (a consequence of Assumption SM2 in the Supplementary Material). If we use the factor-based approach of the previous section and re-estimate on every data release, the factors estimated using adjacent releases differ only by the new data point on a single variable out of the total set of N variables. Therefore in the limit as $N \to \infty$, adjacent horizons' factors become perfectly correlated and $L(y_t - y_{i+1,t}(\theta_{i+1})) - L(y_t - y_{i,t}(\theta_i)) = 0$ almost surely for every t, which implies that $v^2_{i+k,i} = 0$.

If N is very large, a possible solution to this problem is to drop a subset of moment inequalities which are close together, by bounding the release horizons sufficiently far apart by some fraction of N. This can be done in such a way that $v^2_{i+k,i}$ is not equal to zero. We use this general approach to prevent $v^2_{i+k,i} = 0$ in comparing i to $i+k$ by setting k to be at least equal to some deterministic

sequence which we call k_S^c. That is, for some $S > 1$, let:

$$k_S^c = \min\Big\{k \in \{1, \ldots, S-i\} : v_{i+k,i}^2 > c > 0 \quad \forall i = 1, ..., S-1\Big\} \tag{9}$$

Intuitively, k_S^c is the smallest integer k for which the variance $v_{i+k,i}^2$ is larger than some constant c for any i. Note that this still implies that we make use of every data release in the out-of-sample nowcasting experiment, but merely drop some of the pairwise comparisons of these quantities in forming the test statistic. A regularity condition (see below) then rules out the occurrence of pathological cases of non-monotonicity, which may arise by ignoring the horizons between $i + k_S^c$ and i.

Finally, we point out that, while $v_{i+k,i}^2 = 0$ might be a theoretical concern in the limit, our simulation results as well as our empirical application support the view that this spacing is of minor practical relevance in finite sample settings. Indeed, even in the limit, $v_{i+k,i}^2 = 0$ is typically also not an issue at all if the factors are re-estimated only on a fixed number of calendar dates such as the end of each month, as described above.

3.2 Critical Values

Before we introduce the critical values for our test and establish their formal validity, we introduce the shorthands $\Delta L_t(\theta_{i+k}, \theta_i) = L(y_t - y_{i+k,t}(\theta_{i+k})) - L(y_t - y_{i,t}(\theta_i))$ and $\Delta L_t(\widehat{\theta}_{i+k,t}, \widehat{\theta}_{it}) = L(y_t - y_{i+k,t}(\widehat{\theta}_{i+k,t})) - L(y_t - y_{i,t}(\widehat{\theta}_{it}))$ for any $i = 1, \ldots, S-1$ and $k = 1, \ldots, S-i$ combination. Moreover, let κ denote the cardinality of $\mathcal{C}_S$, i.e. $\kappa \equiv |\mathcal{C}_S|$.

In order to establish critical values for the statistic in (8) such that the test has asymptotic size $\alpha \in (0, 1)$, we adopt the approach to testing many moment inequalities recently introduced by Chernozhukov et al. (2014). This procedure does not refer to the asymptotic distribution of the test statistic, but relies on finite sample approximations of the (unknown) asymptotic distribution of the test statistic under H_0 through the use of the Block Multiplier Bootstrap (BMB) procedure. That is, Chernozhukov et al. (2014) show that, under certain regularity conditions, the distribution of $\max_{i,k \in \mathcal{C}_S} \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \Big(\Delta L_t(\theta_{i+k}, \theta_i) - \mathrm{E}\Big[\Delta L_t(\theta_{i+k}, \theta_i)\Big]\Big)$, the statistic evaluated at the true $(\theta_i', \theta_{i+k}')'$ and re-centered by $\mathrm{E}\Big[\Delta L_t(\theta_{i+k}, \theta_i)\Big]$, can be approximated directly by that of $\max_{1 \leq l \leq \kappa} Y_l$ with $Y = (Y_1, \ldots, Y_\kappa)'$, a centered normal random vector of dimension κ (recall that κ denotes the cardinality of the set $\mathcal{C}_S$) with covariance matrix $\mathrm{E}\Big[YY'\Big]$, in Kolmogorov distance. This approximation holds for a given sample size T (and other fixed parameter values) up to some bound $C'P^{-c'}$, where c' and C' are constants which exclusively depend on some unique constants c_1, C_1, and c_2 (see below for a definition). Since the covariance structure of $\max_{1 \leq l \leq \kappa} Y_l$ is unknown, a similar approximation with

$\max_{1\leq l\leq\kappa} Y_l$ can also be established for the BMB bootstrap statistic described in the following. Our theoretical results in Theorem 1 below then suggest that the actual test statistic, which under H_0 satisfies:

$$U^* = \max_{i,k\in\mathcal{C}_S} \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \Delta L_t(\widehat{\theta}_{i+k,t}, \widehat{\theta}_{it}) \leq \max_{i,k\in\mathcal{C}_S} \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \Big(\Delta L_t(\widehat{\theta}_{i+k,t}, \widehat{\theta}_{it}) - \mathrm{E}\Big[\Delta L_t(\theta_{i+k}, \theta_i)\Big]\Big)$$

can indeed be approximated by $\max_{i,k\in\mathcal{C}_S} \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \Big(\Delta L_t(\theta_{i+k}, \theta_i) - \mathrm{E}\Big[\Delta L_t(\theta_{i+k}, \theta_i)\Big]\Big)$. In other words, Theorem 1 below establishes that parameter and factor estimation error in the test as well as in the bootstrap statistic can be bounded by $C_1 P^{-c_2}$, where the bound holds again for fixed values of T, κ, and the other parameters. This suggests that both, the approximation error with the max of a Gaussian process as well as the factor and parameter estimation error, become negligible asymptotically as $P \longrightarrow \infty$.

Paralleling Chernozhukov et al. (2014), we apply the 'small' and 'large' blocks technique widely attributed to the paper of Bernstein (1927). That is, let $q_P > r_P$ denote the 'large' and 'small' blocks respectively and assume 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 $m = P/(q_P + r_P)$ and so $(q_P + r_P)$ characterizes an independent block.

The algorithm of the block multiplier bootstrap is now as follows:

1. Generate standard normal random variables $\epsilon_1, \ldots, \epsilon_m$ independent of the data $\{\Delta L_t(\widehat{\theta}_{i+k,t}, \widehat{\theta}_{it})\}_{t=R+1}^{T}$.

2. Retaining only the large blocks I_h, $h = 1, \ldots, m_P$, construct the BMB statistic:

$$W_{BMB} \equiv \max_{i,k\in\mathcal{C}_S} \left(\frac{1}{\sqrt{m_P q_P}} \sum_{h=1}^{m_P} \epsilon_h \sum_{j\in I_h} \left(\Delta L_j(\widehat{\theta}_{i+k,j}, \widehat{\theta}_{i,j}) - \frac{1}{P} \sum_{t=R+1}^{T} \Delta L_t(\widehat{\theta}_{i+k,t}, \widehat{\theta}_{i,t}) \right) \right).$$

3. Calculate $c^{BMB}(\alpha)$ as the conditional $(1-\alpha)$ quantile of W_{BMB} given $\{\Delta L_t(\widehat{\theta}_{i+k,t}, \widehat{\theta}_{it})\}_{t=R+1}^{T}$.

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, in the Supplementary Material we propose an ad-hoc nested bootstrap procedure suggested by Zhang and Cheng (2014) to estimate the optimal small-large

block combination. Without claiming optimality of any kind, this choice can then subsequently be employed to perform the moment inequality test of interest on the original sample. The proposed method appears to perform well in Monte Carlo and empirical settings.

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 $0 < c_2 < 1/4$. Also, define the k-th mixing coefficient for β-mixing data $\{\Delta L_t(\cdot)\}_{t=1}^T$ as:

$$b_\tau = b_\tau(\{\Delta L_t(\cdot)\}_{t=1}^T) = \max_{1\leq d\leq T-k} \beta(\sigma(\Delta L_1(\cdot),\ldots,\Delta L_d(\cdot)), \sigma(\Delta L_{d+\tau}(\cdot),\ldots,\Delta L_T(\cdot)), \quad 1 \leq \tau \leq T-1,$$

where $\sigma(\Delta L_t(\cdot), t \in \mathcal{T})$ with $\mathcal{T} \subset \{1,\ldots,T\}$ is the sigma field generated by $\Delta L_t(\cdot)$, $t \in \mathcal{T}$. We impose the following regularity conditions:

Assumption 1. *For every $i,k \in \mathcal{C}_S$, the data $\{\Delta L_t(\theta_{i+k},\theta_i)\}_{t=1}^T$ is strictly stationary (across t) and β-mixing where the size of the $\tau-th$ mixing coefficient is determined in Assumption 4 below.*

Assumption 2. *Assume that:*

$$E\left[\max_{i,k\in\mathcal{C}_S}\left|\Delta L_t(\theta_{i+k},\theta_i) - E\Big[\Delta L_t(\theta_{i+k},\theta_i)\Big]\right|^s\right] < C_1 \qquad 1 \leq s \leq 4.$$

as well as:

$$E\left[\|\nabla_F \Delta L_t(\theta_{i+k},\theta_i)\|^{4d_{r_P}}\right] < C \quad \textit{and} \quad E\left[\|\nabla_\beta \Delta L_t(\theta_{i+k},\theta_i)\|^{4d_{r_P}}\right] < C,$$

for all $t = 1,\ldots,T$ and $i,k \in \mathcal{C}_S$, where d_{r_P} is again defined in Assumption 4 below.

Assumption 3. *For every $t = 1,\ldots,T$, and $i,k \in \mathcal{C}_S$, in an open neighbourhood of θ_{i+k} and θ_i, the function $\Delta L_t(\theta_{i+k},\theta_i)$ is measurable and continuously differentiable with respect to all elements of θ_i and θ_{i+k} with probability one.*

Assumption 4. *Let $R,P,N \longrightarrow \infty$ as $T \longrightarrow \infty$ and assume that there exists a sequence $\zeta_{P1} \longrightarrow 0$ as $P \longrightarrow \infty$. Recalling c_2 defined above, we impose the following conditions: let $P/R^{\frac{1}{2}} \longrightarrow \infty$ and $P/N = O(1)$, and assume that*

$$C' \max\left\{\frac{P^{1+c_2}}{R}, \frac{P^{\frac{1}{2}+c_2}}{N}\right\} \leq \zeta_{p1} \leq \frac{C_1}{P^{c_2}}.$$

Moreover, assume that the block sizes satisfy $q_P = o(P)$, $r_P = o(q_P)$ as $P \longrightarrow \infty$. The mix-

ing coefficient b_{r_P} *and the block sizes* r_P *and* q_P *together with the number of blocks* m_P *satisfy* $\max\{m_P b_{r_P}, (r_P/q_P)\log^2\kappa\} \leq C_1 P^{-c_2}$ *and* $q_P C \log^{\frac{5}{2}}(\kappa P) \leq C_1 P^{\frac{1}{2}-c_2}$*, where* $b_{r_P} = o\left(r_P^{-2d_{r_P}/(d_{r_P}-1)}\right)$ *and* $d_{r_P} > 1$*.*

Assumption 5. *For some* c *defined before, the sequence* $k_S^{c_1}$ *defined in (9) exists for every* $S = 1, 2, \ldots,$ *and for every* i*,*

$$Pr\left(\left\{\max_{1\leq t\leq T, k<k_S^{c_1}} \Delta L_t(\theta_{i+k}, \theta_i) > 0\right\} \cap \left\{\max_{1\leq t\leq T, k\geq k_S^{c_1}} \Delta L_t(\theta_{i+k}, \theta_i) \leq 0\right\}\right) \longrightarrow 0$$

as $S, T \longrightarrow \infty$*.*

Assumption 1 characterizes the time dependence in the data, while Assumption 2 requires the existence and finiteness of certain moment expressions. Note that the moment conditions in Assumption 2 involve the loss function, and thus the set of admissible nowcast error distributions which satisfy Assumption 2 will in fact depend on the actual loss function employed. We point out, however, that when nowcast errors are for instance normally distributed with finite mean and variance for every $i = 1, \ldots, S$, then the moment conditions in the first part of Assumption 2 are satisfied for commonly-used loss functions such as the symmetric squared error loss or the asymmetric linear exponential (Linex) loss. Assumption 3 on the other hand requires differentiability of $\Delta L_t(\theta_{i+k}, \theta_i)$, which rules out loss functions such as mean absolute deviation loss, but could, in principle, be weakened along the lines of McCracken (2000). Assumption 4 modifies the asymptotics as in West (1996) with new rates due to our nowcasting set-up. By using big-data approaches allowing the number of moment inequalities, $\kappa = \kappa(N)$, to grow to infinity with the number of variables N and with the sample size split into $T = R + P$, this assumption governs the interplay of these three sequences. Also noting that with calendar release dates linked to the number of variables, meaning κ will at most be of order $O(N^2)$, the first part of Assumption 4 requires R to grow slower than P^2, but faster than P^{1+2c_2}, where $0 < c_2 < 1/4$. This is slightly stricter than $P/R \to 0$, the requirement imposed by West (1996) to ignore parameter estimation error asymptotically, and, when $c_2 = \frac{1}{8}$ for instance, allows for $R \in (cP^{\frac{5}{4}}, CP^2)$. Similarly, we allow the number of variables N to grow as fast as P, but require that $(P^{\frac{1}{2}+2c_2}/N) \longrightarrow 0$, a condition which is again slightly stricter than the out-of-sample analogue of Bai and Ng (2006) who require $\sqrt{P}/N \to 0$ to eliminate factor estimation error asymptotically. This tightening of the growth rates may be viewed as the 'price to pay' to let the number of moment inequalities grow with the sample size. For instance, if again $c_2 = \frac{1}{8}$, we can allow for $N \in (cP^{\frac{3}{4}}, CP)$, which in turn implies that $N/T \longrightarrow 0$. The second part of Assumption 4 determines the rate of the small (r_P)

and large block (q_P) sizes, respectively, and sets the mixing size to be of order $-2d_{r_P}/(d_{r_P}-1)$. It postulates that the rate of the large block size q_P is at most of order $P^{\frac{1}{2}-c_2}$ (if $c_2 = \frac{1}{8}$: $q_p < CP^{\frac{3}{8}}$), while r_P in this case may only grow up to rate $P^{\frac{1}{2}-2c_2}$ (if $c_2 = \frac{1}{8}$: $r_p < CP^{\frac{1}{4}}$). Finally, note that the mixing condition is slightly weaker than in West (1996) who imposes a mixing coefficient of size $-3d/(d-1)$ and thus allows for less time dependence in the data. Finally, Assumption 5 rules out that, as S and T grow to infinity, any violations of monotonicity remain undetected by spacing moment comparisons $k_S^{c_1}$ periods apart from each other. This ensures that the null hypothesis H_0 remains unaffected in the limit by using this spacing.

Before we state the main theoretical result, we point out that there are two additional assumptions in the Supplementary Material, Assumptions SM1 and SM2, which are required for the statement in Theorem 1 below. More specifically, Assumption SM1 contains rather standard factor model regularity conditions adapted to the recursive parameter estimation case, many of which are similar or even identical to Gonçalves et al. (2017). By contrast, Assumption SM2 is needed to formally relax a bounded support assumption made by Chernozhukov et al. (2014) for the case of dependent data for illustrative purposes. However, these are not presented here due to their long and mainly technical nature. The following Theorem establishes the validity of the bootstrap critical values using the BMB procedure:

Theorem 1. *Recall the definition of the constants and assume that Assumptions 1, 2, 3, 4, 5 hold, as well as SM1 and SM2 from the Supplementary Material. Then, there exist positive constants c, C such that under H_0:*

$$Pr\left(U^* > c^{BMB}(\alpha)\right) \leq \alpha + CP^{-c}, \tag{10}$$

where $c^{BMB}(\alpha)$ is the corresponding critical value at level α from the Block Multiplier Bootstrap procedure described before. If $E[\Delta L_t(\theta_{i+k}, \theta_i)] = 0$ for all $i, k \in \mathcal{C}_S$, then:

$$\left|Pr\left(U^* > c^{BMB}(\alpha)\right) - \alpha\right| \leq CP^{-c}. \tag{11}$$

Theorem 1 yields critical values for U^* in such a way that the test has asymptotic size $\alpha \in (0,1)$. Note however that the bounds in Theorem 1 are non-asymptotic in the sense that the bound CP^{-c} holds for a fixed value of P and κ (and the fixed parameters). The second part of Theorem 1, namely Equation (11), shows that when all moment inequalities are binding, the asymptotic size of the test coincides with the nominal size α.

3.3 Critical Values with Moment Inequality Selection

When the number of non-binding and thus uninformative moment inequalities with $\mathrm{E}\left[\Delta L_t(\theta_{i+k}, \theta_i)\right] < 0$ is large, one-step critical values described in the previous section may become too conservative. 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. This selection procedure is similar in spirit to the Superior Predictive Ability test of Hansen (2005), which overcomes potential conservativeness of White (2000)'s reality check test for comparing a (finite) number of forecast models by ignoring irrelevant models in the construction of the limiting distribution.

For the moment selection, we adopt a two-step method proposed by Chernozhukov et al. (2014) for the Multiplier Bootstrap to the BMB used in that 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:

$$\widehat{J}_{BMB} = \left\{ i, k \in \mathcal{C}_S : \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} L_t(\widehat{\theta}_{i+k,t}, \widehat{\theta}_{i,t}) > -2c^{BMB}(\beta_P) \right\}.$$

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 $\epsilon_1, \ldots, \epsilon_m$ independent of the data $\{L_t(\widehat{\theta}_{i+k,t}, \widehat{\theta}_{i,t})\}_{t=R+1}^{T}$.

2. Construct the BMB statistic $W_{\widehat{J}_{BMB}}$ as:

$$\begin{cases} \max\limits_{i,k \in \widehat{J}_{BMB}} \left(\frac{1}{\sqrt{m_P q_P}} \sum_{h=1}^{m_P} \epsilon_h \sum_{j \in I_h} \left(\Delta L_j(\widehat{\theta}_{i+k,j}, \widehat{\theta}_{i,j}) - \frac{1}{P} \sum\limits_{t=R+1}^{T} \Delta L_t(\widehat{\theta}_{i+k,t}, \widehat{\theta}_{i,t}) \right) \right) & \text{if } \widehat{J}_{BMB} \text{ not empty} \\ 0 & \text{if } \widehat{J}_{BMB} \text{ empty} \end{cases}$$

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

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

Theorem 2. *Recall the definition of the constants and suppose that the assumptions of Theorem 1 hold. Moreover, suppose that* $\sup\limits_{P \geq 1} \beta_P < \alpha/2$ *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:

$$Pr\left(U^* > c^{BMB,2S}(\alpha)\right) \leq \alpha + CP^{-c}, \tag{12}$$

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 $E[\Delta L_t(\theta_{i+k}, \theta_i)] = 0$ for all $i, k \in \mathcal{C}_S$, then:

$$Pr\left(U^* > c^{BMB,2S}(\alpha)\right) \leq \alpha - 3\beta_P - CP^{-c}, \tag{13}$$

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

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 $\mathrm{E}[\Delta L_t(\theta_{i+k}, \theta_i)] = 0$ for all $1 \leq l \leq \kappa$. 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

In this section we investigate the finite sample properties of the test statistic in Equation (8). Of particular interest is the performance of the test across different sample sizes, with different specifications of parameter and factor estimation. We therefore propose two set-ups: the first specifies y_t to be a function of a small set of non-factor regressors. This allows us to assess the performance of the test in a lower-dimensional set-up more similar to the MIDAS and bridge equation approaches, and also allows us to shut down the factor estimation component of the results. The second set-up has both parameter and factor estimation which allows us to investigate the test in high dimensions with factor estimation error. We will check sensitivity to the in-sample and out-of-sample splits, R and P also when the test is constructed using different sets of moment inequalities: the full set of $S(S-1)/2$ inequalities, the restricted set using spacings of width k_S^c described in Section 3.1, and the adjacent-only inequalities version akin to the $\mathbf{\Delta^e}$ test of Patton and Timmermann (2012).

4.1 Set-up 1: Parameter Estimation Only

In the first Monte Carlo set-up, y_t is a function of a vector of S non-factor variables $Z_t = [Z_{1t}, ..., Z_{St}]'$:

$$y_t = \gamma' Z_t + \varepsilon_t \tag{14}$$

for $t = 1, ..., T$ where $Z_{it} \sim i.i.d.\mathcal{N}(0, \sigma_Z^2)$ and $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \sigma_\varepsilon^2)$. The variables in Z_t are, for simplicity, taken to be at the same frequency as y_t but released at staggered intervals in the order $1, ..., S$ in a similar way as depicted in Figure 1. We assume that the nowcaster makes nowcasts at each of these points $i = 1, ..., S$ using a misspecified one-variable model which only uses variable i, namely:

$$y_t = \widetilde{\gamma}_i Z_{it} + \varepsilon_{it} \qquad i = 1, ..., S \tag{15}$$

The nowcasting model in Equation (15), which omits all but one of the S variables in the true DGP for y_t, gives a $MSFE_i$ at points $i = 1, ..., S$ of $MSFE_i = \sigma_\varepsilon^2 + \sum_{j \neq i} \gamma_j^2 \sigma_Z^2$. We can therefore vary the relative magnitudes of the sequence of parameter values $\gamma_1, ..., \gamma_S$ in order to generate different scenarios according to the null and alternative hypotheses.

In order to simulate the least favourable case under the null, where the inequalities in the null hypothesis in Equation (1) all hold with equality, we use the parameterisation:

DGPZ-N: $\gamma_i = 1$ for all $i = 1, ..., S$

which gives equal $MSFE_i$ for each i. Under the alternative hypothesis we use the parametersations:

DGPZ-A1: $\gamma_i = 1$ for $i = \lfloor S/2 \rfloor$ and $\gamma_i = 0$ for $i \neq \lfloor S/2 \rfloor$

DGPZ-A2: $\gamma_i = \sqrt{\exp(-i)}$ for all $i = 1, ..., S$

These give two different types of violations of monotonicity similar to existing studies in the monotonicity testing literature such as Ghosal et al. (2000). Specifically, in DGPZ-A1 there is a one-off spike in the γ_i coefficients around the (integer part of) point $S/2$ which causes a drop down and immediate jump up in the $MSFE_i$ profile. In DGPZ-A2 the γ_i coefficients decline as i increases, which implies that $MSFE_i$ is smoothly increasing across the whole profile.

For the various DGP parameters and sample sizes, we use a number of variables $S \in \{5, 10\}$ which allows us to assess the performance of the test in a relatively small-dimensional case. For the all-inequalities version of the test this still yields a total of $S(S-1)/2 \in \{10, 45\}$ moment inequalities,

although we will compare the performance to the adjacent-only version which uses only $S-1 \in \{4, 9\}$ inequalities. We let $\sigma_Z^2 = 1$ and therefore set $\sigma_\varepsilon^2 = \sum_{i=1}^{S} \gamma_i^2$ which fixes the signal-to-noise ratio in Equation (14) to be 1:1. The nowcaster uses recursive OLS to estimate the models described in Equation (15). We let the total sample size be $T \in \{200, 300\}$ and for the sample split $T = R + P$ we use ratios $P/R \in \{1/2, 2\}$ to allow cases where $R < P$ and $P > R$, and so R and P vary between 67 and 200.

For the bootstrap, we considered a grid of values for the small and large blocks equal to $r_P \in \{0, 1, 2\}$ and $q_P \in \{4, 5, 6, 7, 8\}$, though we only display some of these combinations for brevity. We make use of the warp speed bootstrap of Giacomini et al. (2013) which allows us to make just $B = 1$ bootstrap draw over $M = 999$ Monte Carlo replications and reduce the computational complexity of the problem.

4.2 Set-up 2: Parameter and Factor Estimation

In this next set-up we simulate both the factor model and factor-augmented model in Section 2 to give both parameter and factor estimation in the out-of-sample estimation approach. The quarterly factor-augmented regression model is a simplified version of Equation (3):

$$y_t = \beta_0' F_t + \beta_1' F_{t-1/3} + \beta_2' F_{t-2/3} + \varepsilon_t \tag{16}$$

for quarters $t = 1, 2, ..., T$ which has no additional regressors W_t. The monthly factor model has $3T$ observations on the $N \times 1$ vector of variables X_t which is exactly as in Equation (4). As above, the variables X_{jt} are released at distinct times in the order $j = 1, ..., N$, yielding $S = 3N$ nowcasts per quarter. The set-up is interesting relative to existing simulations such as Gonçalves et al. (2017) which do not allow for the ragged edge or consider multiple updates of factor estimates.

The nowcaster forms nowcasts on each data release through the three months of each quarter. The factors are estimated by recursive PCA after solving the ragged edge using an $AR(1)$ interpolation. The estimated factors are then used in the second-stage where recursive OLS is used to estimate the nowcast model, exactly as described in Section 2.2. We take a simple specification of the data generating processes in Equations (16) and (4) where the idiosyncratic errors are $u_{it} \sim i.i.d.\mathcal{N}(0, \sigma_u^2)$, there is a single factor $r = 1$ which follows the $AR(1)$ process $F_t = \rho_F F_{t-1} + v_t$ where $v_t \sim i.i.d.\mathcal{N}(0, 1 - \rho_F^2)$ so that the unconditional variance of the factors is fixed at unity. We let the factor loadings be the non-stochastic vector $\Lambda = \mathbf{1}_{N \times 1}$ and finally the factor-augmented model errors are $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \sigma_\varepsilon^2)$.

In the results which follow, we do not present simulations under the null. As mentioned in Section

2.2, a correct nowcast model specification in this set-up leads to a strictly declining MSFE profile as the contamination to the factors due to the ragged edge interpolation vanishes throughout the period as we add information from $i = 1$ to $i = S$. Thus, this downward-sloping MSFE is in the interior of the null space (with $\mathrm{E}[\Delta L_t(.)] < 0$ rather than the least favourable case where $\mathrm{E}[\Delta L_t(.)] = 0$) and we therefore expect the rejection rates to be close to zero under this specification. We confirmed this by running results for the case where $(\beta_0, \beta_1, \beta_2) = (0.25, 0.25, 0.25)$ which are available from the authors on request.

In order to simulate under the alternative, as above we assume the nowcaster repeatedly misspecifies the model for the true DGP in Equation (16). Again, they use a single-variable version of the model which in this case uses only the factor dated in the current month:

$$y_t = \widetilde{\beta}_2' F_{t-2/3} + \varepsilon_{1t} \qquad i \leq N \tag{17}$$

$$y_t = \widetilde{\beta}_1' F_{t-1/3} + \varepsilon_{2t} \qquad N+1 \leq i \leq 2N \tag{18}$$

$$y_t = \widetilde{\beta}_0' F_t + \varepsilon_{3t} \qquad 2N+1 \leq i \leq 3N \tag{19}$$

Ignoring the effect of the ragged edge interpolation, the monthly $MSFE_m$ for $m = 1, 2, 3$ are equal to $MSFE_1 = (\beta_0^2 + \beta_1^2)\sigma_F^2 + \sigma_\varepsilon^2$, $MSFE_2 = (\beta_0^2 + \beta_2^2)\sigma_F^2 + \sigma_\varepsilon^2$ and $MSFE_3 = (\beta_1^2 + \beta_2^2)\sigma_F^2 + \sigma_\varepsilon^2$. Therefore monotonicity is violated either when $\beta_2 > \beta_1$ or $\beta_1 > \beta_0$ (or both). We therefore consider two different parameterisations for the alternative with differing magnitudes of non-monotonicity:

DGPF-A1: $(\beta_0, \beta_1, \beta_2) = (0, 0.5, 1.5)$

DGPF-A2: $(\beta_0, \beta_1, \beta_2) = (0, 0.5, 3)$.

For the various DGP parameters and sample sizes, we let $\sigma_u^2 = 1$ and $\rho_F = 0.5$ in the factor model. For the nowcast model errors we let $\sigma_\varepsilon^2 = (\beta_0^2 + \beta_1^2 + \beta_2^2)$ which fixes the signal-to-noise ratio in Equation (16) to be 1:1. The same quarterly sample sizes $T \in \{200, 300\}$ are used, noting that this gives monthly sample sizes $3T \in \{600, 900\}$ for X_t. The same sample splits for R and P are maintained from the previous section. For the number of variables in X_t, we consider $N \in \{10, 20\}$. This gives a large number of $S \in \{30, 60\}$ nowcasts per quarter, which in turn yields a maximum number of moment inequalities of $S(S-1)/2 \in \{435, 1770\}$. Since there may be the issue of highly correlated factor estimates, as outlined in Equation (9) we will check the performance of the all-inequalities version of the test with that of moment inequality spacings of $k_S^c = 5$ (further combinations are considered in the Supplementary Material).

We note that, although $N = 10$ is at the lower bound of the number of series used in empirical factor nowcasting studies, this will allow us to explore the performance of the test under factor estimation error. The simulations of Gonçalves et al. (2017) show that the size and power of their test do not vary significantly with N, even if it is as low as 10. In our case we may expect more variation over N because it implies large differences in the number of inequalities.

4.3 Results

Table 1 displays the one-sided rejection rates for the set-up without factor estimation error (DGPZ-N, DGPZ-A1 and DGPZ-A2) for a nominal size of 5%. The results are displayed for the block choices $r_P \in \{0, 1, 2\}$ and $q_P \in \{4, 5\}$ and for the all-inequalities and adjacent-only versions of the test. The results show that the test has good size properties for the least favourable case under the null. For the all-inequalities version of the test, rejection rates tend to be close to the nominal size of 5%, in many cases between 4% and 6%. We see that the test becomes more conservative as the number of inequalities increases, though this is of similar magnitude to the results of Zhang and Cheng (2014), who consider a similar set-up.

In terms of the power, we also see that the test has good performance. For DGPZ-A1, for the all-inequalities version on the left of Table 1 we see power above 90% everywhere for the $T = 200$ case, increasing to unity as T increases to $T = 300$. For DGPZ-A2 the power is slightly lower at around 70% when $P = 67$ but this improves significantly with increases in the sample size. However, we find some more worrying results when using the adjacent-only inequalities test similar to the $\mathbf{\Delta^e}$ test of Patton and Timmermann (2012). Looking at the right hand side of Table 1, although the size properties seem reasonable, we see that the power of this version of the test is substantially lower for DGPZ-A2 which exhibits smoothly upwards-sloping violation of monotonicity.

In general, we conclude that the test works well in this set-up with a small number of non-factor predictors. We also suggest that it is preferable to use the full set of moment inequalities (and not only the adjacent ones), particularly as the results of Theorem 1 allow this to be large relative to the sample size. We now turn our attention to the second case with factor predictors.

In the presence of factor estimates, the results in Table 2 show that the test retains good power properties. The rejection rates are above 80% in all cases for both DGPF-A1 and DGPF-A2 where the quarterly sample size is $T = 200$ and rise to unity when $T = 300$. In analysing the effect of parameter and factor estimation error on the results, considering the case where $T = 200$ in the top half of Table 2, we are interested in what happens as (R, P) switches from $(133, 67)$ to $(67, 133)$. We

Table 1: Rejection Frequencies - Parameter Estimation Only

		All Inequalities						Adjacent-Only					
$T = 200, (R, P) = (133, 67)$								$T = 200, (R, P) = (133, 67)$					
			$\kappa = 10$			$\kappa = 45$			$\kappa = 4$			$\kappa = 9$	
		DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2
$r_P = 0$	$q_P = 4$	0.041	0.923	0.688	0.048	0.918	0.719	0.056	0.932	0.409	0.043	0.914	0.426
$r_P = 0$	$q_P = 5$	0.038	0.932	0.672	0.040	0.930	0.719	0.061	0.927	0.447	0.029	0.902	0.462
$r_P = 1$	$q_P = 4$	0.046	0.929	0.709	0.034	0.921	0.728	0.052	0.927	0.471	0.028	0.919	0.465
$r_P = 1$	$q_P = 5$	0.039	0.941	0.672	0.031	0.932	0.734	0.050	0.916	0.458	0.035	0.898	0.457
$r_P = 2$	$q_P = 4$	0.034	0.926	0.646	0.044	0.935	0.707	0.063	0.925	0.481	0.025	0.886	0.447
$r_P = 2$	$q_P = 5$	0.039	0.892	0.646	0.049	0.919	0.656	0.042	0.910	0.485	0.028	0.902	0.411
$T = 200, (R, P) = (67, 133)$								$T = 200, (R, P) = (67, 133)$					
			$\kappa = 10$			$\kappa = 45$			$\kappa = 4$			$\kappa = 9$	
		DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2
$r_P = 0$	$q_P = 4$	0.042	1.000	0.927	0.028	0.997	0.954	0.051	0.999	0.695	0.049	0.999	0.680
$r_P = 0$	$q_P = 5$	0.048	1.000	0.935	0.032	0.997	0.948	0.059	0.999	0.691	0.040	0.999	0.646
$r_P = 1$	$q_P = 4$	0.040	1.000	0.950	0.023	0.996	0.945	0.049	1.000	0.680	0.036	0.999	0.667
$r_P = 1$	$q_P = 5$	0.041	1.000	0.935	0.020	0.998	0.942	0.047	0.999	0.703	0.038	1.000	0.668
$r_P = 2$	$q_P = 4$	0.048	1.000	0.926	0.022	0.996	0.932	0.043	1.000	0.733	0.039	1.000	0.635
$r_P = 2$	$q_P = 5$	0.040	1.000	0.926	0.028	0.997	0.944	0.046	0.999	0.665	0.032	0.999	0.666
$T = 300, (R, P) = (200, 100)$								$T = 300, (R, P) = (200, 100)$					
			$\kappa = 10$			$\kappa = 45$			$\kappa = 4$			$\kappa = 9$	
		DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2
$r_P = 0$	$q_P = 4$	0.037	0.990	0.843	0.027	0.992	0.888	0.035	0.984	0.618	0.052	0.989	0.559
$r_P = 0$	$q_P = 5$	0.058	0.990	0.862	0.030	0.995	0.875	0.043	0.986	0.627	0.046	0.986	0.559
$r_P = 1$	$q_P = 4$	0.046	0.991	0.880	0.035	0.988	0.867	0.038	0.976	0.591	0.037	0.985	0.583
$r_P = 1$	$q_P = 5$	0.047	0.991	0.871	0.027	0.995	0.876	0.029	0.984	0.582	0.043	0.986	0.534
$r_P = 2$	$q_P = 4$	0.056	0.990	0.872	0.027	0.995	0.850	0.029	0.981	0.553	0.045	0.985	0.521
$r_P = 2$	$q_P = 5$	0.050	0.992	0.830	0.023	0.993	0.859	0.044	0.984	0.629	0.027	0.992	0.499
$T = 300, (R, P) = (100, 200)$								$T = 300, (R, P) = (100, 200)$					
			$\kappa = 10$			$\kappa = 45$			$\kappa = 4$			$\kappa = 9$	
		DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2	DGPZ-N	DGPZ-A1	DGPZ-A2
$r_P = 0$	$q_P = 4$	0.048	1.000	0.992	0.034	1.000	0.992	0.040	1.000	0.853	0.041	1.000	0.860
$r_P = 0$	$q_P = 5$	0.039	1.000	0.993	0.031	1.000	0.996	0.050	1.000	0.888	0.051	1.000	0.844
$r_P = 1$	$q_P = 4$	0.038	1.000	0.991	0.031	1.000	0.992	0.035	1.000	0.832	0.042	1.000	0.832
$r_P = 1$	$q_P = 5$	0.035	1.000	0.987	0.032	1.000	0.993	0.039	1.000	0.853	0.056	1.000	0.828
$r_P = 2$	$q_P = 4$	0.045	1.000	0.994	0.032	1.000	0.993	0.045	1.000	0.840	0.050	1.000	0.858
$r_P = 2$	$q_P = 5$	0.047	1.000	0.992	0.026	1.000	0.991	0.056	1.000	0.886	0.047	1.000	0.860

Table 2: Rejection Frequencies - Parameter and Factor Estimation

		All Inequalities				Spacing $k_S^c = 5$			
		$T = 200, (R, P) = (133, 67)$				$T = 200, (R, P) = (133, 67)$			
		$N = 10, \kappa = 435$		$N = 20, \kappa = 1770$		$N = 10, \kappa = 300$		$N = 20, \kappa = 1485$	
		DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2
$r_P = 0$	$q_P = 4$	0.863	0.928	0.930	0.948	0.872	0.921	0.909	0.928
$r_P = 0$	$q_P = 5$	0.870	0.932	0.926	0.937	0.912	0.920	0.898	0.917
$r_P = 1$	$q_P = 4$	0.894	0.927	0.915	0.951	0.858	0.912	0.903	0.921
$r_P = 1$	$q_P = 5$	0.848	0.911	0.923	0.952	0.899	0.920	0.876	0.929
$r_P = 2$	$q_P = 4$	0.871	0.921	0.896	0.943	0.850	0.908	0.891	0.914
$r_P = 2$	$q_P = 5$	0.883	0.931	0.930	0.930	0.869	0.892	0.900	0.927
		$T = 200, (R, P) = (67, 133)$				$T = 200, (R, P) = (67, 133)$			
		$N = 10, \kappa = 435$		$N = 20, \kappa = 1770$		$N = 10, \kappa = 300$		$N = 20, \kappa = 1485$	
		DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2
$r_P = 0$	$q_P = 4$	0.995	0.997	0.998	0.998	0.997	0.997	0.999	1.000
$r_P = 0$	$q_P = 5$	0.998	0.994	0.998	0.998	0.996	0.998	1.000	1.000
$r_P = 1$	$q_P = 4$	0.995	0.994	0.998	0.998	0.997	0.997	0.999	1.000
$r_P = 1$	$q_P = 5$	0.995	0.997	0.998	0.998	0.997	0.998	0.999	1.000
$r_P = 2$	$q_P = 4$	0.994	0.997	0.998	0.998	0.997	0.997	0.999	1.000
$r_P = 2$	$q_P = 5$	0.993	0.997	0.998	0.998	0.997	0.997	0.999	1.000
		$T = 300, (R, P) = (200, 100)$				$T = 300, (R, P) = (200, 100)$			
		$N = 10, \kappa = 435$		$N = 20, \kappa = 1770$		$N = 10, \kappa = 300$		$N = 20, \kappa = 1485$	
		DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2
$r_P = 0$	$q_P = 4$	0.984	0.978	0.979	0.995	0.979	0.982	0.988	0.995
$r_P = 0$	$q_P = 5$	0.981	0.980	0.980	0.987	0.976	0.991	0.986	0.991
$r_P = 1$	$q_P = 4$	0.985	0.979	0.980	0.994	0.980	0.991	0.989	0.995
$r_P = 1$	$q_P = 5$	0.977	0.980	0.990	0.991	0.979	0.988	0.991	0.995
$r_P = 2$	$q_P = 4$	0.980	0.986	0.980	0.987	0.978	0.991	0.989	0.995
$r_P = 2$	$q_P = 5$	0.984	0.978	0.983	0.984	0.975	0.984	0.988	0.995
		$T = 300, (R, P) = (100, 200)$				$T = 300, (R, P) = (100, 200)$			
		$N = 10, \kappa = 435$		$N = 20, \kappa = 1770$		$N = 10, \kappa = 300$		$N = 20, \kappa = 1485$	
		DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2	DGPF-A1	DGPF-A2
$r_P = 0$	$q_P = 4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$r_P = 0$	$q_P = 5$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$r_P = 1$	$q_P = 4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$r_P = 1$	$q_P = 5$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$r_P = 2$	$q_P = 4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$r_P = 2$	$q_P = 5$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

see that the power of the test under DGPF-A1 and DGPF-A2 improves significantly as we move from the $(R, P) = (133, 67)$ to the $(R, P) = (67, 133)$ case. This indicates that, although R is required to be large relative to P for estimation error to be irrelevant asymptotically, in finite samples it appears that increases in power come from the increase in the number of out-of-sample points, P, in spite of larger estimation error.

Table 2 shows that the results do not change markedly when comparing the all-inequalities version of the test to those with spacings $k_S^c = 5$. This indicates that this may not be a serious issue in practical applications. In the Supplementary Material, we provide a wide range of further simulations which explore this in more detail, including an alternative set-up where monotonicity is violated when the nowcaster has a horizon-specific prediction bias.

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 five GDP subcomponents. This builds on the literature of nowcasting the individual components of GDP, studies of 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. (2017) 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 Method

The quarterly real GDP variable and its five sub-components (consumption, investment, government expenditure, exports and imports) was accessed from the Haver Analytics database USECON. For the monthly variables, we select a dataset of $N = 62$ predictors from the FRED-MD database, all transformed to stationary using the transformations recommended by McCracken and Ng (2016). A full description of these variables, along with information regarding their publication lag, can be found in the Supplementary Material. The variables were selected to be similar to the "medium" case of Bańbura and Modugno (2014) who show that factors estimated from the largest form of the Stock and Watson-type dataset, like FRED-MD, tend to perform worse than smaller datasets in nowcasting GDP as they contain too many price and financial series. Our database mostly contains data on

production, surveys, employment, prices and housing series. This dataset is somewhat larger than many existing factor model nowcasting studies for the U.S. such as Antolín Díaz et al. (2017).

Regarding the timeliness of the data flow, the first release of the target GDP series occurs just under a month after the end of the reference quarter, for example Q1 is released near the end of April. The most timely monthly data is the consumer sentiment index which the only variable to be released before the end of the observation month. This is shortly followed by the financial series which are available without lag as they are aggregated from daily data. Employment data are also very timely, whereas industrial production, housing and loans variables are released some weeks after the end of the observation month.

This information is used to generate a calendar of releases which determines the sequence of nowcasts we use to predict each quarter's GDP growth. Starting from the first day of the quarter and finishing around 118 days later when GDP is released, we consider every unique release of data. This includes cases where data are released at different times in the same day, such is the case with the retail sales and industrial production variables. Counting up the distinct releases in the 3 months of the nowcast period and the 28 days of the backcast period before GDP is released gives a total of $S = 53$ different nowcast updates per quarter.

We use data on all months and quarters between 1978Q2 and 2016Q1, which is the common available sample for all variables. This gives a total of $T = 152$ quarterly observations and $3T = 456$ monthly observations. We use the recursive out-of-sample method described in Section 2.2 with different splits of R and P, namely $(R, P) \in \{(101, 51), (87, 65), (76, 76), (65, 87), (51, 101)\}$ which correspond to ratios of $P/R \in \{1/2, 3/4, 1, 4/3, 2\}$.

At every prediction point $t = R + 1, ..., T$ and at every nowcast point $i = 1, ..., S$, in the first step we solve the ragged edge problem using a simple $AR(1)$ interpolation as in Kim and Swanson (2017). We then estimate the factors by PCA as in Equation (6). We estimate the number of factors r using the IC_{p2} information criterion of Bai and Ng (2002) based on the first window of data for the case $P/R = 1$ and fix this number throughout the out-of-sample experiment. This results in $r = 4$ factors which is between the $r = 8$ factors found by McCracken and Ng (2016) on the full set of $N = 134$ FRED-MD variables, and $r = 1$ used in the small-scale factor model of Antolín Díaz et al. (2017).

We make nowcasts using the quarterly regression model:

$$y_t = \mu + \rho y_{t-1} + \beta_0' F_t + \beta_1' F_{t-1/3} + \beta_2' F_{t-2/3} + \varepsilon_t \tag{20}$$

which is equivalent to Equation (3) with $W_t = [1, y_{t-1}]'$, so the must-have regressors W_t are just a

constant and a lag of y_t. Since the lagged quarter of GDP is not available in the first 28 days of the nowcast period, as with the monthly variables we solve the ragged edge by substituting y_{t-1} with the prediction from an $AR(1)$ model until the data point is released.

In generating the sets of moment inequalities for our test, we measure nowcast error losses using MSFE. We first consider using the full set of all $S(S-1)/2$ pairwise moment inequalities, which gives a high-dimensional set-up with $\kappa = 1378$. As in the Monte Carlo section, we will compare our results to the adjacent-only version of the test, similar to the $\mathbf{\Delta^e}$ test of Patton and Timmermann (2012), and the version with inequality spacing of $k_S^c = 5$ (we checked the results for $k_S^c \in \{1, 2\}$ which were very similar and therefore omitted). We determine the block lengths using the optimal procedure outlined in the Supplementary Material and generate $B = 399$ bootstrap draws.

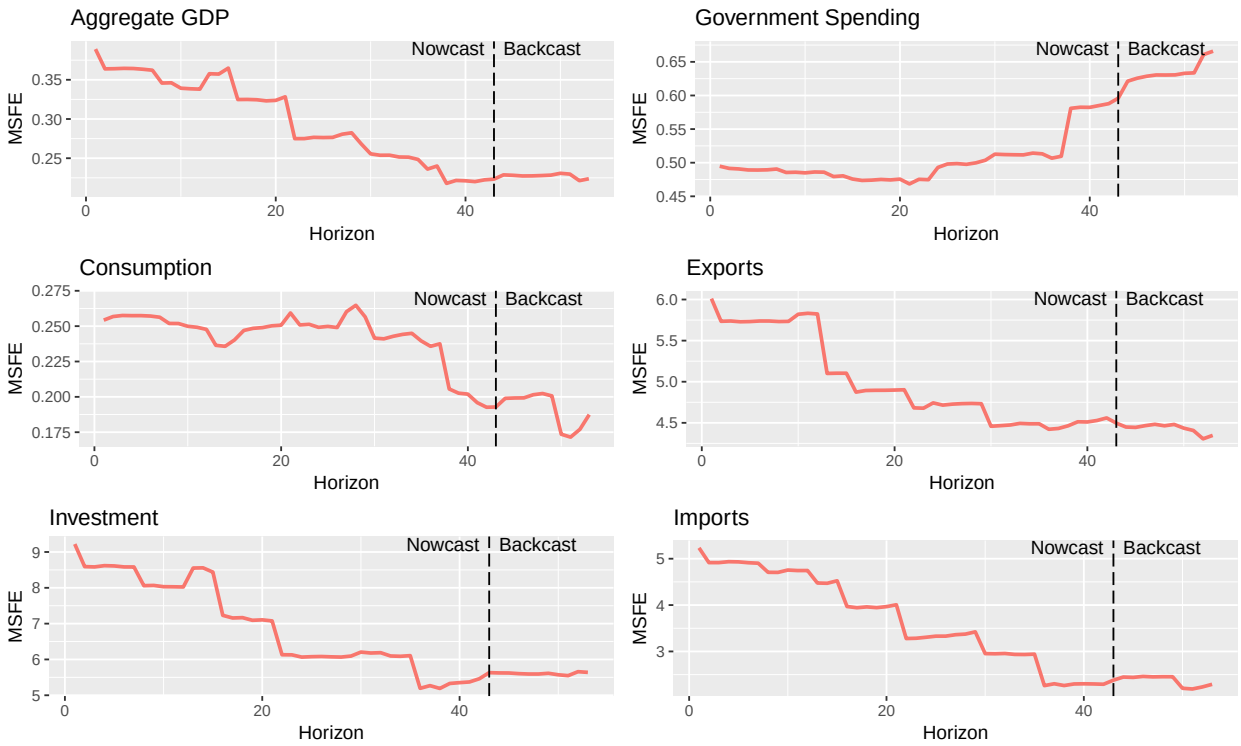
We also ran other configurations to explore the robustness of the empirical results and assess how the test performs. Firstly, the rolling estimation scheme is used in place of the recursive scheme. Next, we consider a reduced dataset of $N = 9$, similar to the "small" case of Bańbura and Modugno (2014), to observe any differences resulting from dataset composition. We also explore grouping data releases together into days (not updating nowcasts within the day) and also in sets of 10 days, rather than using each unique release timing. We next explore the sensitivity to the date span used by providing results with a post-1984 sample, motivated by evidence such as Breitung and Eickmeier (2011) who find a break in the factor loadings in around 1984, attributed to the 'Great Moderation'. Finally, we check the results when the asymmetric Linex loss function is used instead of squared error loss.

5.2 Results

The graphs in Figure 2 depict the evolution of MSFE for the aggregate real GDP growth variable and the five GDP sub-components, for the scenario where $(R, P) = (76, 76)$. Upon graphical inspection of the results, the top-left panel of Figure 2 shows that the nowcasts for aggregate real GDP growth improve throughout the period, with the quarterly MSFE dropping from 0.35% to below 0.25% before flattening in the backcast period.

For consumption, investment, exports and imports the MSFE profiles are also clearly downwards-sloping. In some cases there are jumps such as for exports at $i = 13$ when the previous quarter's exports data is released, and for investment at $i = 16$ when a group of employment-related series are released simultaneously. However in all of these series, including aggregate GDP, there are some minor violations of monotonicity with small segments of rising MSFE. Our test is able to determine whether or not these deviations from monotonicity are statistically significant or are merely features

Figure 2: MSFE by Nowcast Horizon for the Sample with $(R, P) = (76, 76)$



Notes: Horizon refers to the point throughout the prediction period at which we estimate the factor model and forecasting model.

caused by small-sample heterogeneity in the data.

The most striking feature is that of the government spending component, where MSFE is increasing throughout most of the prediction period, from 0.5% to more than 0.65%. Since this increase seems to be sizeable, we expect our test to reject the null hypothesis of monotonicity.

We now turn to the results of our test, which can formally detect whether these MSFE profiles are monotonically declining rather than simply relying only on the graphical evidence in Figure 2. The results are displayed in Table 3, where the first two sets of columns use our test with the full set of inequalities and then with spacing $k_S^c = 5$. The last set of columns uses the adjacent-only comparisons as in the $\mathbf{\Delta^e}$ test of Patton and Timmermann (2012).

The first thing to note is that the results of our test appear to be very stable, both across the various (R, P) combinations and also as we move from the full set of moment inequalities to the spacing $k_S^c = 5$. For the headline aggregate GDP series, the test finds no evidence of non-monotonicity with p-values around 0.7 in these versions of the test. This finding mirrors the result of Bańbura et al. (2013), though using a different method for estimating the factors for the nowcasting model, and over a different in-sample and out-of-sample period. This suggests that factor-based methods are

Table 3: Monotonicity Test Results by GDP Sub-Component

		All Inequalities ($\kappa = 1378$)				Spacing k_S^c=5 ($\kappa = 1128$)				Adjacent-only ($\kappa = 52$)			
		U^*	90%	95%	p-val	U^*	90%	95%	p-val	U^*	90%	95%	p-val
	GDP	0.265	1.967	2.217	0.774	0.181	1.535	1.983	0.840	0.196	0.726	0.956	0.612
	CONS	0.173	1.295	1.594	0.546	0.173	1.287	1.502	0.541	0.067	0.381	0.551	0.614
$R = 101$	INV	6.270	52.073	64.052	0.689	5.785	50.598	61.723	0.764	5.995	22.622	26.974	0.619
$P = 51$	GOV	1.040	1.230	1.428	0.140	1.040	1.248	1.431	0.168	0.729	0.764	0.930	0.113
	EXP	1.099	18.825	23.441	0.910	1.099	17.232	22.655	0.842	0.704	7.679	9.440	0.880
	IMP	1.826	34.240	43.349	0.724	1.826	36.066	48.349	0.719	0.849	11.525	14.622	0.807
	GDP	0.198	1.618	2.150	0.820	0.166	1.643	2.040	0.827	0.154	0.570	0.763	0.614
	CONS	0.228	1.157	1.621	0.534	0.228	1.203	1.505	0.494	0.110	0.430	0.490	0.569
$R = 87$	INV	5.376	42.177	57.087	0.784	5.376	42.251	52.243	0.749	4.833	20.961	25.953	0.654
$P = 65$	GOV	1.667	1.389	1.794	0.065	1.667	1.353	1.617	0.035	0.633	0.755	0.920	0.165
	EXP	1.342	17.922	22.967	0.832	1.342	17.471	22.589	0.837	0.788	5.178	6.168	0.885
	IMP	1.773	26.047	33.398	0.694	1.773	31.115	42.130	0.684	0.749	8.954	10.122	0.835
	GDP	0.234	1.485	1.900	0.724	0.166	1.486	1.790	0.767	0.171	0.706	0.963	0.629
	CONS	0.253	1.271	1.519	0.499	0.253	1.239	1.543	0.459	0.098	0.307	0.366	0.634
$R = 76$	INV	4.660	40.626	54.963	0.817	4.341	24.879	30.361	0.797	4.609	19.036	23.488	0.644
$P = 76$	GOV	1.721	1.172	1.546	0.028	1.721	1.152	1.337	0.013	0.621	0.637	0.742	0.108
	EXP	1.203	16.574	21.071	0.865	1.203	18.736	22.659	0.905	0.752	6.954	8.852	0.862
	IMP	1.733	13.570	16.132	0.732	1.733	14.509	16.912	0.717	0.768	9.291	12.065	0.862
	GDP	0.201	1.545	1.830	0.840	0.134	1.686	2.079	0.892	0.143	0.606	0.779	0.727
	CONS	0.263	1.145	1.370	0.486	0.263	1.072	1.307	0.494	0.100	0.396	0.516	0.594
$R = 65$	INV	4.271	46.331	55.836	0.845	4.271	44.610	54.225	0.867	4.147	23.796	31.785	0.742
$P = 87$	GOV	1.540	1.150	1.379	0.030	1.540	1.281	1.619	0.058	0.529	0.658	0.794	0.206
	EXP	1.432	16.053	21.526	0.797	1.432	16.732	21.077	0.822	0.739	7.468	8.651	0.897
	IMP	3.160	29.758	37.303	0.624	3.160	12.351	14.857	0.689	1.491	9.394	12.084	0.579
	GDP	0.194	1.176	1.558	0.792	0.130	1.164	1.433	0.900	0.152	0.507	0.645	0.652
	CONS	0.236	0.856	1.157	0.461	0.236	0.960	1.297	0.506	0.103	0.254	0.288	0.576
$R = 51$	INV	4.400	28.539	35.535	0.799	4.400	50.441	61.825	0.812	4.197	13.485	17.895	0.569
$P = 101$	GOV	1.466	1.167	1.347	0.033	1.466	1.127	1.386	0.038	0.497	0.568	0.709	0.150
	EXP	1.362	11.554	14.767	0.842	1.294	11.412	13.557	0.827	1.244	5.034	6.261	0.704
	IMP	3.802	24.678	33.122	0.469	3.802	22.969	29.892	0.529	1.886	7.335	9.029	0.579

Notes: The p-value is the one-sided rejection probability based on the empirical bootstrap distribution.

able to generate monotonically improving nowcasts of U.S. real GDP growth irrespective of the factor estimation method or data span. The results for consumption, investment, imports and exports are very similar to the result for aggregate GDP, with large p-values as expected.

In the case of government spending the first two versions of the test find significant evidence of non-monotonicity with many rejections at the 5% level, with the exception of the smallest sample size with $P = 51$. This finding shows that our test is able to reject in scenarios where there are clear violations of monotonicity. It is interesting to note that the adjacent-only version of the test, however, does not reject the null at the 5% significance level in any configuration. This aligns with the results of the Monte Carlo simulations where very low power was detected for upward-sloping MSFE profiles.

Finally, we also find that our results are robust when switching to the rolling estimation scheme and when reducing the dataset to $N = 9$. These results, all of which can be found in the Supplementary Material, show even stronger rejection of the null in the government spending component with many at the 1% significance level. The adjacent-only version of the test still fails to reject the null across almost every configuration. The findings are also very similar when we only make $S = 12$ nowcasts per quarter by updating the nowcasts every 10 days, even though the maximum number of moment inequalities drops to $\kappa = 66$. The idea of using the post-1984 sample motivated by possible structural breaks in the factor loadings also has very minimal bearing on the results. Lastly, even when the nowcaster has asymmetric Linex loss with aversion to negative prediction errors, they come to the same conclusion regarding nowcast monotonicity.

6 Conclusion

This paper proposes a test to formally assess the performance of 'big data' nowcasting methods. In particular, we focus on models involving estimated factors which have become a leading case in the high-dimensional empirical nowcasting literature, though the test can readily deal with small-dimensional set-ups like bridge equations and MIDAS models. The test uses moment inequalities to evaluate the monotonicity of metrics for nowcast accuracy such as MSFE. 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.

The main contribution of this paper is in extending 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 grow with the sample size. We also show that their moment selection pro-

cedure to rule out non-binding inequalities remains applicable in our context. We then provide the conditions under which both parameter estimation error and factor estimation error are asymptotically negligible in this high dimensional setting when using the pseudo out-of-sample approach of West (1996). We derive conditions required for using factor-augmented models of Stock and Watson (2002a,b) and Bai and Ng (2006) in this nowcast monotonicity test, with the factors estimated by Principal Components Analysis (PCA), after solving the issue of the 'ragged edge'.

We illustrate the finite sample performance of our test through Monte Carlo simulations, and provide an in-depth empirical application to nowcasting the growth rate of U.S. aggregate real GDP, and the growth rate of five GDP subcomponents. Our test confirms that there is no statistical evidence of non-monotonicity in the MSFE of aggregate GDP growth as more information is added throughout the nowcast period. However, the test does pick up violations of monotonicity in the government spending component of GDP. This result suggests that the factor-augmented model may be misspecified for nowcasting government spending, and that alternative models should be considered.

References

Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud (2014). Nowcasting GDP in Real Time: A Density Combination Approach. *Journal of Business & Economic Statistics 32*(1), 48–68.

Altissimo, F., R. Cristadoro, M. Forni, M. Lippi, and G. Veronese (2010). New Eurocoin: Tracking Economic Growth in Real Time. *The Review of Economics and Statistics 92*(4), 1024–1034.

Antolín Díaz, J., T. Drechsel, and I. Petrella (2017). Tracking the Slowdown in Long-Run GDP Growth. *The Review of Economics and Statistics 99*(2), 343–356.

Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge Models to Forecast the Euro Area GDP. *International Journal of Forecasting 20*(3), 447–460.

Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica 74*(4), 1133–1150.

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, pp. 195–236. Amsterdam: North-Holland.

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.

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

Bragoli, D. and J. Fosten (2018). Nowcasting Indian GDP. *Oxford Bulletin of Economics and Statistics 80*(2), 259–282.

Breitung, J. and S. Eickmeier (2011). Testing for Structural Breaks in Dynamic Factor Models. *Journal of Econometrics 163*(1), 71–84.

Bühlmann, P. and H. Künsch (1999). Block Length Selection in the Bootstrap for Time Series. *Computational Statistics and Data Analysis 31*, 295–310.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Testing Many Moment Inequalities. cemmap Working Paper CWP65/13, Massachusetts Institute of Technology.

Clark, T. E. and M. W. McCracken (2009). Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts. *International Economic Review 50*(2), 363–395.

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

Clements, M. P. and A. B. Galvão (2009). Forecasting US Output Growth Using Leading Indicators: an Appraisal Using MIDAS Models. *Journal of Applied Econometrics 24*(7), 1187–1206.

Corradi, V. and N. R. Swanson (2014). Testing for Structural Stability of Factor Augmented Forecasting Models. *Journal of Econometrics 182*(1), 100–118.

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

Doz, C., D. Giannone, and L. Reichlin (2011). A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering. *Journal of Econometrics 164*(1), 188–205.

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.

Giacomini, R., D. N. Politis, and H. White (2013). A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators. *Econometric Theory 29*(3), 567–589.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The Real-time Informational Content of Macroeconomic Data. *Journal of Monetary Economics 55*(4), 665–676.

Gonçalves, S., M. W. McCracken, and B. Perron (2017). Tests of Equal Accuracy for Nested Models with Estimated Factors. *Journal of Econometrics 198*(2), 231–252.

Hall, P., J. Horowitz, and B. Jing (1995). On Blocking Rules For the Bootstrap with Dependent Data. *Biometrika 82*, 561–574.

Hansen, P. R. (2005). A Test for Superior Predictive Ability. *Journal of Business & Economic Statistics 23*(4), 365–380.

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

Kim, H. H. and N. R. Swanson (2017). Methods for Backcasting, Nowcasting and Forecasting Using Factor-MIDAS: With An Application To Korean GDP. *Journal of Forecasting (forthcoming)*.

Knotek, E. S. and S. Zaman (2017). Nowcasting US Headline and Core Inflation. *Journal of Money, Credit and Banking 49*(5), 931–968.

Luciani, M. and L. Ricci (2014). Nowcasting Norway. *International Journal of Central Banking 10*(4), 215–248.

Marcellino, M., M. Porqueddu, and F. Venditti (2016). Short-Term GDP Forecasting With a Mixed-Frequency Dynamic Factor Model With Stochastic Volatility. *Journal of Business & Economic Statistics 34*(1), 118–127.

McCracken, M. and S. Ng (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics Economic Statisics 34*(4), 574—-589.

McCracken, M. W. (2000). Robust Out-of-sample Inference. *Journal of Econometrics 99*(2), 195–223.

Patton, A. J. and A. Timmermann (2012). Forecast Rationality Tests Based on Multi-horizon Bounds. *Journal of Business & Economic Statistics 30*(1), 1–17.

Schumacher, C. (2016). A Comparison of MIDAS and Bridge Equations. *International Journal of Forecasting 32*(2), 257–270.

Stock, J. H. and M. W. Watson (2002a). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics 20*(2), 147–162.

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.