On the Real-Time Predictive Content of Financial Conditions Indices for Growth*

Aaron J. Amburgey† and Michael W. McCracken‡

January 18, 2022

Abstract

We provide evidence on the real-time predictive content of the National Financial Conditions Index (NFCI), for conditional quantiles of U.S. real GDP growth. Our work is distinct from the literature in two specific ways. First, we construct (unofficial) real-time vintages of the NFCI. This allows us to conduct out-of-sample analysis without introducing the kind of look-ahead biases that are naturally introduced when using a single current vintage. We then develop methods for conducting asymptotic inference on tests of equal tick loss between nested quantile regression models when the data are subject to revision. We conclude by evaluating the real-time predictive content of NFCI vintages for quantiles of real GDP growth. While our results largely reinforce the literature, we find gains to using real-time vintages leading up to recessions — precisely when policymakers need such a monitoring device.

JEL Nos.: C12, C32, C38, C52
Keywords: out-of-sample forecasts, real-time data, quantiles

*We are grateful to Scott Brave, Todd Clark, and Ana Galvão for helpful comments as well as participants at the Friendly Faces workshop, the 2021 CFE conference, and seminar participants at the Federal Reserve Bank of St. Louis. The views expressed here are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

†Federal Reserve Bank of St. Louis. Address: P.O. Box 442. St. Louis, MO 63166, U.S.A. Tel: 314-444-4229. Email: aaron.amburgey@stls.frb.org

‡Federal Reserve Bank of St. Louis. Address: P.O. Box 442. St. Louis, MO 63166, U.S.A. Tel: 314-444-8594. Email: michael.w.mccracken@stls.frb.org

1 Introduction

Building on work by De Nicoló and Lucchetta (2017) and Alessandri and Mumtaz (2017), Adrian, Boyarchenko, and Giannone (2019; ABG) observe that current financial conditions have a substantial negative effect on the lower quantiles of future real GDP (RGDP) growth but have limited effect elsewhere in its conditional distribution. This observation has lead to a surge of interest in growth-at-risk including, but not limited to, work by Coe and Vahey (2020), who take a much longer historical context; Carriero, Clark, and Marcellino (2020), who consider the role of conditional volatility; Brownlees and Souza (2021), who take an international perspective, and Reichlin, Ricco, and Hasenzagl (2020), who provide a negative assessment, concluding that financial conditions have little predictive content for quantiles conditional on other macroeconomic information.

Although each of these papers bring their own angle to the literature on growth-at-risk, they all share one thing in common: at least part of their analysis uses a current vintage of the National Financial Conditions Index (NFCI) in a pseudo out-of-sample framework to form predictions of RGDP growth. While instructive, by taking a pseudo out-of-sample approach they do not ascertain whether the NFCI would have been useful for monitoring tail risk in real time. To be fair, vintages of the NFCI only became available in May 2011 and hence could not have been used leading into the Great Recession or any other U.S. recession prior to the COVID-19-induced recession of 2020. Even so, several of the series used to construct the NFCI depend on GDP directly (e.g., the corporate debt-to-GDP ratio) and hence pseudo out-of-sample analyses potentially lead to look-ahead biases when predicting GDP growth. In addition, as emphasized by Brownlees and Souza (2021), since the NFCI is estimated using a mixed frequency dynamic factor model, using the most recent vintage avoids filtering uncertainty at the sample endpoints — an issue that is unavoidable when using real-time data.

With this in mind, we make three contributions to the literature on growth-at-risk with an eye towards the real-time nature of the problem. First, we construct (unofficial) real-time weekly vintages of the NFCI from January 1988 through May 2011 with values that date back to January 1971. The vintages are constructed using most of the financial series in the official version as well as real-time GNP and GDP vintage data. They cover the same time span as the official NFCI, are weekly, and are constructed using mixed frequency state-space code written by Scott Brave and Andrew Butters - the originators and curators of the official NFCI. We then show that while the unofficial vintages are highly correlated with the current official vintages, end-of-sample differences exist, making them less smooth than current vintage values.

This leads to our second contribution. We provide analytical and Monte Carlo evidence on tests of equal expected tick loss for nested quantile models when data are subject to revision. The vast majority of the literature on tests of equal predictive ability, including West (1996), Clark and McCracken (2001), Giacomini and Rossi (2010), and Rossi and Sekhposyan (2011), assume that the observables, used to form and evaluate the forecast, are unrevised. One could form an argument that the tests developed in Diebold and Mariano (1995) and Giacomini and White (2006) are robust to the presence of revisions, but they introduce other complicating issues: either parameter estimation is not present or they require that a finite rolling window of observations is used to estimate parameters.

In contrast, Clark and McCracken (2009) develop tests of equal predictive ability under quadratic loss, between OLS-estimated nested linear models, when the predictors and predictands are subject to revision. They find that the presence of revisions can lead to dramatically different asymptotic distributions associated with tests of equal accuracy. Specifically, they find that test statistics with otherwise non-standard asymptotic distributions become asymptotically normal in the presence of predictable revisions. We find that the same result applies to nested quantile models evaluated under tick loss.[1]

Having developed a method for evaluating forecast accuracy under tick loss in the presence of revisions, our final contribution is a thorough and methodologically sound evaluation of the real-time predictive content of the NFCI vintages for RGDP growth. We find that adding a single lag of the NFCI to a baseline Quantile Autoregressive (QAR) model with one lag provides robust, substantial performance gains when predicting the lower tails of RGDP growth. While this result only reinforces much of the existing literature, we also find some evidence that the real-time vintages can lead to more accurate forecasts. Specifically, we find that real-time vintages have superior predictive content leading into recessions.

[1] Corradi, Fosten, and Gutknecht (2020) constructs tests of equal tick loss between non-nested models but do not consider data that are subject to revision.

While tentative, we speculate that current vintages smooth over turning point information in order to obtain lower factor volatility. Vintages leading into a recession may be more volatile, but they cannot smooth through a recession that has yet to occur and hence retain their signal.

Before proceeding, it is worth emphasizing that we are not the first to highlight the real-time nature of growth-at-risk. Delle Monache, De Polis, and Petrella (2021) use a current vintage of the NFCI for part of their analysis but then perform a related exercise using a real-time data set composed of the individual series underlying the NFCI. Caldara, Cascaldi-Garcia, Cuba-Borda, and Loria (2020) also consider the NFCI for part of their analysis but also construct their own real-time macroeconomic and financial factors based on a modest collection of five macroeconomic and four financial series. In the context of the euro area, Ferrara, Mogliani, and Sahuc (2020) gather real-time macroeconomic and financial data to predict growth-at-risk at the daily frequency. Nevertheless, our paper stands out, first and foremost, by formally constructing (unofficial) historical vintages of the NFCI and using those to conduct the forecasting exercises. In addition, in order to assess these forecasts, we develop analytical tools for conducting asymptotic inference in tests of equal tick loss in the presence of data revisions.

The remainder of the paper proceeds as follows. Section 2 describes construction of the unofficial vintages of the NFCI and provides an assessment of their validity. Section 3 presents our proposed method for constructing tests of equal expected tick loss when revisions are present. Section 4 presents Monte Carlo evidence on the finite sample size and power of these tests. Section 5 we assesses the predictive content of the NFCI in a real-time environment while also noting its behavior leading into recessions. Section 6 concludes.

2 Real-Time (Unofficial) Vintages of the NFCI

As we note in the introduction, the increasingly large literature on growth-at-risk typically uses current vintages of the NFCI when forming retrospective quantile predictions of growth in RGDP. In this section we first describe our approach to constructing a sequence of unofficial real-time vintages of the NFCI. We then assess the quality of these vintages relative to the current official NFCI vintages. Finally, we analyze the evolution of the real-time NFCI (RTNFCI) over time and provide motivation for the proceeding sections.

2.1 Constructing the Vintages

The NFCI is an indicator of overall financial conditions that is updated weekly by the Federal Reserve Bank of Chicago. It was constructed by Brave and Butters (2012) using the large approximate dynamic factor framework of Doz, Giannone, and Reichlin (2006). One advantage of this framework is that it allows the authors to create a weekly latent indicator that also uses monthly and quarterly series. Over time, the authors have substituted a few of the original series, but these changes have not substantially altered the NFCI. Currently, the NFCI contains 46 weekly, 33 monthly, and 26 quarterly series.

To construct the NFCI vintages, we first gather weekly data vintages for January 6, 1988, through May 18, 2011 (the week before the first official vintage was released). For each vintage, the data extend back to January 1971 when available. We obtain financial and survey data from Haver, Bloomberg, and FRED, as well GNP and GDP vintages from ALFRED. We use 41 weekly, 25 monthly, and 26 quarterly of the original NFCI series. Thus, for the last vintage constructed, 92 of the original 105 series are included.[2] For the financial and survey-based series, we assume any revisions are negligible. In order to ensure the real-time nature of the series, we use the release timing information provided by the original source of each series. For GNP and GDP, the only series with considerable data revisions, we use the latest release as of each weekly vintage date.

Of the 92 series we use to construct the NFCI vintages, only six date back to January 1971. Furthermore, only 50 of the series contain data before 1988, the date of our first constructed vintage. While the model is designed to handle missing values, to avoid the inclusion of noisy information, we need a rule of thumb for how many observations a series must have before it is included in each vintage. We include series that have two years worth of observations, that is 104, 24, and 8 observations for data with a weekly, monthly, and quarterly frequency, respectively. Figure 1 plots the number of series available for the first vintage of each year for both our unofficial vintages and the hypothetical official vintages.[3]

[2]For the three delinquency rate series with NFCI mnemonics DBC, DCLOSE, and DHE, we substitute American Banking Associate (ABA) data for the FRED series DRCCLT100S, DRCLACBS, and DRSREACBS, respectively.

[3]We thank Scott Brave for providing the availability information of each series in the NFCI.

Notice that while 50 of the series contain data before 1988, not all these series are used in the first vintage, due to our two-year inclusion rule. The first vintage that we construct contains 45 series, while the last contains all 92 of our series. Most of the series are added within the first 12 years, with 85 series included by the year 2000. The hypothetical official NFCI vintages also start with 45 available series and end with all 105 series. From 1988 to 2001, the maximum number of extra series the hypothetical official NFCI vintages contain is two. However, the availability of extra series increases to five in 2002 and continues to increase until 2011 where all 13 extra series are included. More detailed information on the availability of indicators is provided in Appendix A.

For a given weekly vintage of data, we estimate the latent factor using the Mixed Frequency State Space (MFSS) toolbox in MATLAB put together by Brave, Butters, and Kelley (2020), which uses the same mixed frequency state-space approach of Brave and Butters (2012). Although it was assembled by the curators of the NFCI, there are some slight differences in estimation methods. Namely, the system parameters in the MFSS are found through optimization algorithms (e.g., gradient ascent and simplex based methods), while the official NFCI is estimated via the EM algorithm of Shumway and Stoffer (1982) and Watson and Engle (1983).[4] We consider this difference in methodology to be insignificant, especially relative to the large advantages gained by using code written by the authors of the official NFCI.

For each vintage, we first transform the series according to the codes provided by the NFCI curators.[5] We then demean and normalize each transformed series to unit variance. After the data preparation steps, we drop series with insufficient observations according to the two-year rule outlined earlier in this section. We set up the state-space model with the same assumptions as Brave and Butters (2012). Specifically, we assume there is a single latent weekly factor that follows an AR(15) process. To deal with the low-frequency series that are not point-in-time sampling types, the MFSS toolbox requires us to define their accumulator type following Harvey (1989). We simply use the accumulators provided by

[4]We use the default settings of the MFSS toolbox with the exception that we only use gradient ascent. When iterating between gradient ascent and simplex based methods, we lose substantial computational time despite the fact that the resulting estimates are nearly identical.

[5]There are four exceptions where we choose different transformation codes. The series with NFCI mnemonics MLIQ10, MBOND, MMF, and CBPER are kept as levels in the official NFCI. We choose to transform each of these by taking the first difference, due to their lack of stationarity in levels.

the NFCI curators. Finally, we are ready to estimate the model parameters. As noted in Brave, Butters, and Kelley (2020), parameter initialization plays an important role in finding the correct optima. With this in mind, we fastidiously follow the NFCI parameter initialization methods given by Brave and Butters (2012). In early versions of the official NFCI, the authors dropped the first two years (i.e., 1971 and 1972) of estimated data to avoid any issues with the initialization. We follow this methodology for two reasons: 1) for several vintages, the model returned strange values for the first several values of the latent factor, and 2) we believe this more accurately represents how the NFCI would have been given in vintages that precede the first official one.

Creating the 1220 unofficial weekly vintages between January 1988 and May 2011 presents a computational obstacle. While we would like to fully estimate the model parameters for each vintage, the optimization methods of the MFSS often take a long time to converge, making this approach impractical. However, failing to re-estimate the parameters with the most recently available data in real time poses accuracy risks to the validity of our NFCI vintages. We try to find a middle ground, and instead re-estimate parameters in the first week of each month. For off weeks, we apply the most recently estimated parameters to the updated data. Because parameter estimates vary by only small amounts from week to week, reusing parameter estimates across a maximum of four vintages does not create any significant complications.

2.2 Assessing the Vintages

Although we closely follow the methodology of the official NFCI curators, it is important to assess whether our unofficial vintages are comparable to the official versions. This is especially the case since our estimates do not use all 105 of the series used in the NFCI. Relatedly, Brave and Butters (2012) do robustness checks on their NFCI estimates and find that when they use only 39 of the original series, the resulting latent factor hardly resembles the official NFCI. Since we use 3-month averages of the NFCI vintages in the empirical analysis later, our assessment focuses on these 3-month averages. Unreported assessments of the raw weekly vintages yield comparable results.

Figure 2 plots a few select vintages against each other from 1971 to 2021. The blue, orange, and yellow lines represent three of our unofficial vintages, while the purple and green lines represent two official vintages. The official 2021 vintage is the only vintage to contain data from 1971 to 1972, since, as noted in Subsection 2.1, even early official vintages dropped these first two years of observations. Ignoring the first two years, all five of the vintages look strikingly similar. For the most part, the first four vintages seem to slowly converge to the official 2021 series, with the large spikes in the 1970s and 1980s increasing with each revision. Large deviations are common for the newest values of each vintage (e.g., in the 1/27/1988 vintage where the 1987-1988 values are quite a bit lower than in the other vintages). This feature is in-line with remarks made by Brownlees and Souza (2021) concerning the high filtering uncertainty at sample endpoints for dynamic factor models. While the 1988 vintage looks the most different from the official vintages, it is important to note that only one of the original NFCI series that would have otherwise been available for this period is missing from our data set. Thus, it is unlikely that the hypothetical official vintage would have deviated significantly from our estimate.

We also calculate the correlation coefficients between the available observations of each vintage. For brevity, we only report the coefficients between the first vintage of even years, which are provided in Table 1. This table gives further evidence that, despite having fewer series, our unofficial vintages are highly correlated with the official ones, with correlations ranging from 0.9 to, effectively, 1. Unsurprisingly, the correlation coefficients tend to decrease as vintages get further apart. Nonetheless, even the 1988 and 2020 vintages have a correlation of 0.93. Focusing on the bottom-right portion of the table, the vintages from 2010-2020 (including our unofficial 2010 and 2011 vintages) all have a correlation of at least 0.98. The large positive correlations between our 2010 and 2011 vintages and the official vintages, suggest that the 13 official series that are missing from our data set do not significantly alter the latent factor.

2.3 The Real-Time Evolution of the NFCI

The results reported in Figure 2 and Table 1 suggest that our unofficial vintages of the NFCI are quite close to the official versions. As such, we feel confident using our unofficial vintages from 1988-2011 in conjunction with the official vintages from 2011-2021 to analyze the importance of the NFCI as a real-time monitoring device. In the remainder of the paper we refer to this conjunction of all real-time vintages as the real-time NFCI or RTNFCI, while we call the most recent 2021 vintage in our sample the Expost NFCI or EXNFCI. Finally, we loosely use the term NFCI to refer to the index more generally (such as when we discuss specific vintages, release timing, and the revision process).

Before discussing the NFCI's usefulness in a real-time monitoring context, it is worth noting that, as shown in Figure 2, revisions to its past values can be non-trivial. To get a more comprehensive look at the magnitude of these revisions, Figure 3 plots the Expost NFCI against the latest real-time value of each NFCI vintage from 1988-2021. For example, for June 2, 2000, the orange line plots the value of the June 7, 2000, vintage (-0.06), while the blue line plots the value (0.51) of the most recent vintage in our sample.[6] One noticeable difference between the expost and real-time versions is that, from 1988-2021, the former has a variance of 0.24, while the latter is more volatile, with a variance of 0.40. There are a few factors that play into this: 1) for a given vintage and its most recent date, some of the data have not been fully revised and some of the data has not been released yet; 2) especially in the first several years, new series are often added from vintage to vintage; and 3) the parameters of the dynamic factor model evolve between the vintages as the model takes in more data. It is unsurprising that the left-hand side of the plot exhibits a larger portion of the higher volatility since 2) and 3) are especially potent for earlier vintages. Besides the difference in volatility, the two plots roughly mimic each other with a correlation of 0.88 – lower than those found in Table 1 but not drastically so.

Finally, we provide some initial evidence on the usefulness of the EXNFCI and RTNFCI in the context of a collection of out-of-sample RGDP quantile forecasting exercises. We do so by giving a big picture overview with an eye towards identifying which features of the quantile regressions are most relevant. Specifically, for $\tau = 1$ (4), we forecast quantiles of annualized RGDP growth (average annualized RGDP growth)[7] with origins of 1988Q1 through 2019Q4 (2019Q1). Notice that we do not include the COVID-19 period in our sample. We exclude this period because large outliers occur at the horizons 2020Q2 and 2020Q3, which shroud the patterns exhibited from 1988-2019. We continue discussion of the COVID-19 period in Section 5.

[6]The official NFCI is released the first Wednesday after the reference week. We follow this pattern throughout.

[7]We evaluate forecasts with the advance release of GDP, although we performed this exercise for both the third and current estimates and obtained similar results.

Forecast origins take place on the first day of the second month of each quarter (e.g., for the forecast origin 1988Q1 and $\tau = 1$, we forecast 1988Q1 using the information available on February 1, 1988). We use models including just an intercept, an intercept with the once lagged NFCI, an intercept with lagged RGDP growth, and an intercept with both lagged GDP and lagged NFCI. Thus the unrestricted model takes the form $\hat{y}^{(\alpha)}_{s,\tau} = x'_s\hat{\beta}^{(\alpha)}$ for $x_s = (1, y_s, z_s)'$, where y_s is RGDP growth, z_s is the NFCI, and $\alpha \in (0, 1)$ is the chosen conditional quantile. The restricted models are obtained by setting the coefficients on y_s and/or z_s to zero. For each model including the NFCI, there are an additional four permutations resulting from: 1) whether we use the EXNFCI or the RTNFCI and 2) whether we use calendar quarter averages of the NFCI (e.g., for Q1 the average of January-March) or 3-month averages of the most recent information (e.g., for Q1 the average of February-April). Furthermore, for each model there are six additional permutations resulting from: 1) whether we use a rolling or recursive window and 2) the initial estimation window size, with choices of $R = 30, 60, 90$.[8] [9] Thus, for a given quantile and horizon, there are 60 forecasting models in total. We evaluate each of the 60 permutations under average tick loss:

$$P^{-1}\sum_{t=R}^{T-\tau} L(\hat{u}^{(\alpha)}_{s+\tau}) = P^{-1}\sum_{t=R}^{T-\tau}(\alpha - 1(y_{s+\tau} \leq \hat{y}^{(\alpha)}_{s,\tau}))(y_{s+\tau} - \hat{y}^{(\alpha)}_{s,\tau}).$$

Given the $j = 1, ..., 60$ permutations, we use an ANOVA-type regression to identify the importance of the considered treatment effects. This takes the form

$$\begin{aligned} Y_{j,(\alpha,\tau)} =& \gamma_0 + \gamma_1 Roll_j + \gamma_2 R_{60,j} + \gamma_3 R_{90,j} + \gamma_4 GDP_j + \gamma_5 EXNFCI_j + \gamma_6 RTNFCI_j \\ &+ \gamma_7(EXNFCI_j * Cal_j) + \gamma_8(RTNFCI_j * Cal_j) + \epsilon_j, \end{aligned}$$

where for a fixed horizon τ and quantile α, $Y_{j,(\alpha,\tau)}$ is the natural logarithm of mean loss for model j. All RHS variables, other than the intercept, represent indicators of one of the treatments. $Roll_j$ indicates that rolling windows are used for estimation. $R_{60,j}$ and $R_{90,j}$ indicate window sizes of 60 and 90 are used at the initial forecast origin. GDP_j indicates that RGDP growth is included. $EXNFCI_j$ and $RTNFCI_j$ indicate that the corresponding

[8] Under the rolling scheme, these are the sizes of the estimation sample at each forecast origin. Under the recursive scheme, these represent the number of observations used up to and including the initial forecast origin.

[9] When $R = 90$ the first forecast origin for $\tau = 1$ (4) is 1993Q3 (1995Q1). The forecast origin starts later due to not having 90 observations to regress on from 1971-1988.

indicators are included. Cal_j indicates that the EXNFCI or RTNFCI uses calendar averages rather than the staggered 3-month average approach described earlier. Table 2 reports the resulting coefficients and p-values of each regressor for $\tau = 1, 4$ and five distinct quantiles $\alpha = 0.05, 0.10, 0.50, 0.90, 0.95$. Coefficients represent the average percentage change in mean loss associated with the inclusion of the given dummy variable in the model.

Strikingly, for both forecast horizons, in the lower tails (i.e., $\alpha = 0.05$ and $\alpha = 0.1$) the EXNFCI and RTNFCI dummy variables all have negative coefficients of -0.35 or lower, and all but one of these cases are statistically significant at the 5% level. In other words, for all of these models, the inclusion of the EXNFCI or RTNFCI is associated with at least a 35% decrease in mean loss. On the other hand, the coefficients of the EXNFCI and RTNFCI dummy variables tend to be positive in the upper tails (i.e., $\alpha = 0.9$ and $\alpha = 0.95$). These results support much of the existing literature, including ABG, which show that the NFCI is a useful tool for predicting downside risks to GDP, but not upside risks. Notably, the RTNFCI performs similarly (and sometimes better) than the EXNFCI. Though it is less decisive, the GDP inclusion indicator has negative coefficients at every horizon and quantile albeit not always significantly so. Somewhat surprisingly, as indicated by the insignificant coefficients in the final two rows, the choice of quarterly aggregation of the weekly NFCI (i.e., current 3-month vs. calendar quarter) appears to not play an important role.

Moving on to estimation-oriented indicators, both the R_{60} and R_{90} dummy variables have negative coefficients in the lower tails when $\tau = 1$ and for all quantiles when $\tau = 4$, suggesting that models with larger window sizes tend to perform better in these instances. Importantly, the rolling window indicator typically has statistically significant positive coefficients in the lower tail, suggesting that the recursive scheme is preferred. In contrast, when $\tau = 1$, these coefficients are negative in the upper tail, suggesting that the rolling scheme is preferred.

3 Real-Time Inference on Equal Tick Loss

As noted in the introduction, our goal is to assess the real-time predictive content of financial conditions, specifically the NFCI, for U.S. RGDP growth. In the previous section we addressed one aspect of this issue: obtaining real-time (unofficial) vintages of the NFCI that avoid look-ahead biases. Having obtained these vintages, we now address a second aspect of this issue: how to conduct inference in tests of equal predictive ability — between two nested quantile regression models — when data are subject to revision.

To understand the problem, note that the quantile regressions in ABG largely come from two models: one that uses an intercept and one lag of RGDP growth as predictors and one that also includes one lag of the NFCI and hence nests the baseline model. In this framework, the issue of data revisions arises in two distinct ways as we move across forecast origins. First, the lagged predictors are both subject to revision. For RGDP, this is due to the regular revision process that the Bureau of Economic Analysis conducts as more information is gathered on the reference quarter. For the NFCI, the revisions arise not only due to filtering uncertainty, but also due to revisions in GDP that are used in the construction of the index. Second, when evaluating the accuracy of the quantile losses under tick loss, one needs to choose a specific future vintage of realized RGDP growth for the predictand, and that vintage will also remain subject to revision.

This matters because data revisions can have a dramatic effect on the asymptotic distribution of tests of out-of-sample predictive ability. In the absence of data revisions, Clark and McCracken (2001) and McCracken (2007) show that sample averages of loss differentials from nested models will not be asymptotically normal but will instead have non-standard asymptotic distributions with representations as functions of stochastic integrals. In contrast, Clark and McCracken (2009) show that, in the context of conditional mean forecasts evaluated under quadratic loss, these sample averages can be asymptotically normal depending on properties of the revision process. In the following, we adapt their framework to an environment where quantile regressions are evaluated under tick loss.

3.1 Framework

At each forecast origin $t = R, ..., T - \tau$, forecasts are constructed using current vintage data $\{y_s(t), x'_s(t)\}_{s=1}^{t}$. This data consist of a scalar predictand $y_s(t)$ and vector of predictors $x_s(t)$ associated with observations $s = 1, ..., t$. With each new vintage, a subset of the most recent observations is subject to revision. Specifically, we allow the observables to be subject to a regular revision process over a finite number of periods r with $r << t$. This is a useful approximation as it implies that, in large samples, the sequence of parameter estimates are consistent for their population counterparts - the fact that the last r observations are subject to revision is asymptotically irrelevant. Were we to permit a longer lived revision process, perhaps one that includes benchmark revisions, we would have to consider a situation in which the population parameters were time varying, not due to any structural change in the economy, but rather due to the revision process itself. Finally, when the revision process is complete and the data are final, for simplicity we drop the parenthetical and let $(y_s(t), x_s'(t)) = (y_s, x_s')$.

In the spirit of our proposed application, the τ-step-ahead conditional quantile forecasts are based on two linear models $x_{j,s}'(t)\beta_j$ $j = 1, 2$, where $x_{2,s}(t) = (x_{1,s}'(t), x_{22,s}'(t))$ and hence model 2 nests model 1 and under the null we have $\beta_2^* = (\beta_1^{*\prime}, 0')'$. For a given quantile $\alpha \in (0, 1)$, both β_1 and β_2 are re-estimated as we move across forecast origins by minimizing the relevant tick loss function.[10] When a recursive estimation window is used, this optimization takes the form

$$\hat{\beta}_{j,t}^{(\alpha)} = \arg\min_{\beta_j} \sum\nolimits_{s=1}^{t-\tau} (\alpha - 1(y_{s+\tau}(t) \leq x_{j,s}'(t)\beta_j))(y_{s+\tau}(t) - x_{j,s}'(t)\beta_j).$$

When a rolling estimation window is used, the optimization is comparable but the summation is replaced with $\sum_{s=t-R+1}^{t-\tau}$. Regardless of which estimation window is used, the quantile forecasts $x_{j,t}'(t)\hat{\beta}_{j,t}^{(\alpha)}$ are evaluated against the future realization $y_{t+\tau}(t')$ for some vintage t' such that $t' - t + \tau$ is a fixed and finite non-negative integer. This yields two sequences of P forecast errors, denoted $\hat{u}_{j,t+\tau}^{(\alpha)}(t') = y_{t+\tau}(t') - x_{j,t}'(t)\hat{\beta}_{j,t}^{(\alpha)}$ $j = 1, 2$, and subsequent loss differentials

$$f_{t+\tau}(\hat{\beta}_t^{(\alpha)}) = (\alpha - 1(\hat{u}_{1,t+\tau}^{(\alpha)}(t') \leq 0))\hat{u}_{1,t+\tau}^{(\alpha)}(t') - (\alpha - 1(\hat{u}_{2,t+\tau}^{(\alpha)}(t') \leq 0))\hat{u}_{2,t+\tau}^{(\alpha)}(t'),$$

where $\hat{\beta}_t^{(\alpha)} = (\hat{\beta}_{1,t}^{(\alpha)\prime}, \hat{\beta}_{2,t}^{(\alpha)\prime})'$.

We are interested in the asymptotic behavior of the out-of-sample average of the loss differentials $P^{-1}\sum_{t=R}^{T-\tau} f_{t+\tau}(\hat{\beta}_t^{(\alpha)})$. If the function $f_{t+\tau}(.)$ is twice continuously differentiable in $\hat{\beta}_t^{(\alpha)}$, as it is under quadratic loss, methods developed in West (1996) are directly applicable for doing so. Unfortunately, due to the presence of the indicator functions, it is not continuously differentiable under tick loss. Instead, we apply results in McCracken

[10]Under tick loss, the argmin need not be unique. In our simulations and empirical work, we apply the Frisch-Newton interior point method.

(2000) that only require that the expectation of $f_{t+\tau}(.)$, $Ef_{t+\tau}(.)$ be continuously differentiable. A simple motivating example that satisfies this property is provided in the following subsection.

Delineating the asymptotic distribution requires a bit more notation. As in West (1996) and McCracken (2000), we assume that the sequence of parameter estimates satisfy $\hat{\beta}_t - \beta^* = B(t)H(t)$ where $H(t)$ denotes a vector of zero mean moment conditions and $B(t) \to^{a.s.} B$, a full rank nonstochastic matrix. In the context of a single OLS regression in a stationary environment with residuals $\varepsilon_{s+\tau}$, $B = (Ex_t x_t')^{-1}$ and for $h_{s+\tau} = x_s \varepsilon_{s+\tau}$, $H(t)$ equals $t^{-1}\sum_{s=1}^{t-\tau} h_{s+\tau}$ or $R^{-1}\sum_{s=t-R+1}^{t-\tau} h_{s+\tau}$ for the recursive and rolling schemes, respectively. In the context of a single α-quantile regression in a strictly stationary environment with residuals $\varepsilon_{s+\tau}^{(\alpha)}$, $B = (E\varphi_\alpha(0|x_t)x_t x_t')^{-1}$ and $h_{s+\tau} = (\alpha - 1(\varepsilon_{s+\tau}^{(\alpha)} < 0))x_s$, where $\varphi_\alpha(.|x_s)$ denotes the density function of $\varepsilon_{s+\tau}^{(\alpha)}$ conditional on x_s. Note though, for our application we have two quantile regressions so that $\hat{\beta}_t^{(\alpha)} = (\hat{\beta}_{1,t}^{(\alpha)\prime}, \hat{\beta}_{2,t}^{(\alpha)\prime})'$. This issue is readily addressed by defining $B = diag(B_1, B_2)$ and $h_{s+\tau} = (h_{1,s+\tau}', h_{2,s+\tau}')'$ for $B_j = (E\varphi_\alpha(0|x_{j,t})x_{j,t}x_{j,t}')^{-1}$ and $h_{j,s+\tau} = (\alpha - 1(\varepsilon_{s+\tau}^{(\alpha)} < 0))x_{j,s}$ keeping in mind that under the null, $\varepsilon_{s+\tau}^{(\alpha)} = \varepsilon_{j,s+\tau}^{(\alpha)}$ $j = 1, 2$.

With this structure in hand, applying the results in McCracken (2000) we obtain the following expansion that separates the variation in the underlying moment being tested from the variation due to parameter estimation error:

$$P^{-1/2}\sum_{t=R}^{T-\tau} f_{t+\tau}(\hat{\beta}_t^{(\alpha)}) = P^{-1/2}\sum_{t=R}^{T-\tau} f_{t+\tau}(\beta^*) + FB(P^{-1/2}\sum_{t=R}^{T-\tau} H(t)) + o_p(1),$$

where $F = \partial Ef_{t+\tau}(\hat{\beta})/\partial\hat{\beta}|_{\hat{\beta}=\beta^*}$. This expansion is identical to one in West (1996) and hence we can apply his Theorem 4.1, which implies that as both P and R tend to infinity,

$$\begin{aligned} P^{-1/2}\sum_{t=R}^{T-\tau} f_{t+\tau}(\hat{\beta}_t^{(\alpha)}) &\to^{d} N(0,\Omega) \\ \Omega &= S_{ff} + 2\lambda_{fh}FBS_{fh}' + \lambda_{hh}FBS_{hh}F'B', \end{aligned}$$

where S_{ff} is the long-run variance of $f_{t+\tau}(\beta^*)$, S_{fh} is the long-run covariance between $f_{t+\tau}(\beta^*)$ and $h_{t+\tau}$, and both λ_{fh} and λ_{hh} are non-stochastic functions of $\lim_{P,R\to\infty} P/R = \pi$ delineated in West (1996).[11] [12] This result provides us with an option for conducting

[11]Under the recursive scheme, π is allowed to diverge, while under the rolling scheme, π must remain finite.

[12]To save space, we do not reiterate the assumptions made in the referenced papers. Loosely speaking, the results described in this section require that the fully revised observables and revisions are strictly stationary, that $\pi > 0$, and that the residuals from the quantile regressions have continuous PDFs in an open neighborhood of the origin.

asymptotically valid inference so long as Ω is positive and we can consistently estimate the elements of Ω. An extensive discussion on the estimation of these elements is provided in both West (1996) and McCracken (2000). For our application to a test of equal accuracy between nested models, a few of these elements are particularly easy to estimate. For example, under the null, the two models are equivalent and hence both S_{ff} and S_{fh} are zero since $f_{t+\tau}(\beta^*)$ is zero. We therefore only need to estimate the elements λ_{hh}, B, S_{hh}, and F and do so without prior knowledge of the revision process absent that the process is finite lived. Details on how we estimate these elements is left to section 4.

3.2 A Simple Example

Consider our motivating example, one in which we want to compare the predictive content of two nested quantile regression models that take the form

$$\begin{aligned} y_{s+1} &= \beta_{0,1} + \beta_{1,1} y_s + \varepsilon_{1,s+1}^{(\alpha)} \\ &= x'_{1,s}\beta_1 + \varepsilon_{1,s+1}^{(\alpha)} \end{aligned} \tag{1}$$

and

$$\begin{aligned} y_{s+1} &= \beta_{0,2} + \beta_{1,2} y_s + \beta_{2,2} z_s + \varepsilon_{2,s+1}^{(\alpha)} \\ &= x'_{2,s}\beta_2 + \varepsilon_{2,s+1}^{(\alpha)}. \end{aligned} \tag{2}$$

Here we have in mind that y denotes RGDP growth and z denotes the NFCI. Under the null, $\beta_{2,2} = 0$ so that the NFCI has no predictive content for the α-quantile conditional on the presence of y_t. In the following we derive the asymptotic variance Ω under a specific data generating process with an eye towards identifying whether or not Ω — and specifically F — will be non-zero.

The data generating process for y is a Gaussian AR(1) that takes the form

$$y_{s+1} = \delta_0 + \delta_1 y_s + \varepsilon_{s+1},\ \varepsilon_s \sim i.i.d.N(0, \sigma_\varepsilon^2). \tag{3}$$

This process is convenient because the quantiles are location-scale and take the form

$$Q^{(\alpha)}(y_{s+1}|y_s) = \sigma_\varepsilon \Phi^{-1}(\alpha) + \delta_0 + \delta_1 y_s,$$

where $\Phi(.)$ denotes the standard normal distribution function and $\Phi^{-1}(.)$ denotes its inverse. In the absence of revisions, the parameter estimates in the two quantile regressions will be consistent for their population counterparts such that $\beta_{0,1} = \beta_{0,2} = \sigma_\varepsilon \Phi^{-1}(\alpha) + \delta_0$, $\beta_{1,1} = \beta_{1,2} = \delta_1$, and $\beta_{2,2} = 0$. In addition, the α-quantile residuals take the form $\varepsilon_{s+1}^{(\alpha)} = \varepsilon_{s+1} - \sigma_\varepsilon \Phi^{-1}(\alpha)$ and hence are Gaussian — albeit with a nonzero mean.

We can now begin to delineate the elements of the asymptotic variance. To do so, first note that under the null, the quantile residuals $\varepsilon_{j,s+1}^{(\alpha)} = \varepsilon_{s+1}^{(\alpha)}$ are conditionally homo-amplitudinal and hence $\varphi_\alpha(0|x_{j,s}) = \varphi_\alpha(0)$. Moreover, we know $\varphi_\alpha(0) = \sigma_\varepsilon^{-1}\phi(\Phi^{-1}(\alpha))$ and hence $B_j = (E\varphi_\alpha(0|x_{j,s})x_{j,s}x_{j,s}')^{-1}$ simplifies to $(Ex_{j,s}x_{j,s}')^{-1}/(\sigma_\varepsilon^{-1}\phi(\Phi^{-1}(\alpha)))^2$. Similarly, since the estimated models are correctly specified, we find that $S_{h_ih_j} = E((\alpha - 1(\varepsilon_{s+\tau}^{(\alpha)} < 0))^2 x_{i,s}x_{j,s}')$ simplifies to $(\alpha - \alpha^2)(Ex_{i,s}x_{j,s}')$ for $i, j = 1, 2$.

Deriving a closed form for F requires knowledge of the revision process. For analytical tractability we assume that the y variable is revised once and the revision is deterministic such that $y_t = y_t(t) + c$. We could allow c to be stochastic, but it would needlessly complicate the algebra below. For this exercise we do not consider revisions to z, but in the simulations and empirics, we estimate Ω allowing for that option.

Given this revision process, we immediately find that for $F = (F_1, -F_2)$, F_2 takes the form

$$\begin{aligned}
F_2 &= \partial E[(\alpha - 1(y_{t+1}(t') - x_{2,t}'(t)\hat{\beta}_2 \leq 0))(y_{t+1}(t') - x_{2,t}'(t)\hat{\beta}_2)]/\partial\hat{\beta}_2|_{\hat{\beta}_2 = \beta_2^*} \\
&= -E(\alpha - 1(y_{t+1}(t') - x_{2,t}'(t)\beta_2^* \leq 0))x_{2,t}'(t) \\
&= -E(\alpha - 1(u_{2,t+1}^{(\alpha)} \leq (y_{t+1} - y_{t+1}(t')) - (x_{2,t}' - x_{2,t}'(t))\beta_2^*))(x_{2,t}' - (x_{2,t}' - x_{2,t}'(t))) \\
&= -E(\alpha - 1(u_{t+1}^{(\alpha)} \leq (y_{t+1} - y_{t+1}(t')) - c\delta_1))(x_{2,t}' - (0, c, 0)),
\end{aligned}$$

while for $J' = (I_{k_1 \times k_1}, 0_{k_1 \times k_{22}})$, $F_1 = F_2 J$. This representation is instructive because it allows us to see how F may or may not be zero. If $c = 0$ or if the forecast is evaluated using revised data $y_{t+1}(t+2) = y_{t+1}$ and $\delta_1 = 0$, then F is zero. But even if $\delta_1 = 0$, so long as the forecast is evaluated using the initial release $y_{t+1}(t+1)$, F is non-zero.

We can now put the elements together and characterize the asymptotic variance. If we assume that the forecasts are evaluated against the initial release $y_{t+1}(t+1)$, we find that

$$\begin{aligned}
\Omega &= \lambda_{hh} F B S_{hh} F' B' \\
&= \lambda_{hh} F_2(-JB_1J' + B_2)S_{h_2h_2}(-JB_1J' + B_2)F_2' \\
&= \lambda_{hh}(\alpha-\alpha^2)(\sigma_\varepsilon^{-1}\phi(\Phi^{-1}(\alpha)))^{-2}F_2(-J(Ex_{1,t}x_{1,t}')^{-1}J' + (Ex_{2,t}x_{2,t}')^{-1})F_2' \\
&= \lambda_{hh}(\alpha-\alpha^2)(\alpha - \Phi(\Phi^{-1}(\alpha) + c\frac{(1-\delta_1)}{\sigma_\varepsilon}))^2(\sigma_\varepsilon^{-1}\phi(\Phi^{-1}(\alpha)))^{-2} \times \\
&\quad (0,-c,0)(-J(Ex_{1,t}x_{1,t}')^{-1}J' + (Ex_{2,t}x_{2,t}')^{-1})(0,-c,0)' \\
&= \lambda_{hh}(\alpha-\alpha^2)(\alpha - \Phi(\Phi^{-1}(\alpha) + c\frac{(1-\delta_1)}{\sigma_\varepsilon}))^2\phi(\Phi^{-1}(\alpha))^{-2}\frac{\sigma_\varepsilon^2 c^2}{\sigma_y^2}(\frac{\rho_{zy}^2}{1-\rho_{zy}^2}), \qquad (4)
\end{aligned}$$

where σ_y^2 denotes the variance of y_t and ρ_{zy} denotes the correlation between y and z.[13] The second equality arises because $F_1 = F_2J$ and $h_{1,s+1} = J'h_{2,s+1}$. The third arises because the quantile residuals are conditionally homo-amplitudinal, and under the null, both models are correctly specified. The fourth arises due to the fact that $u_{2,t+1}^{(\alpha)}$ is independent of $x_{2,t}$ and $Ex_{2,t}'(-J(Ex_{1,t}x_{1,t}')^{-1}J' + (Ex_{2,t}x_{2,t}')^{-1}) = 0$. The final equality is straightforward algebra associated with the final component in the fourth equality.

While we have emphasized the importance of F not being zero for Ω to be positive, the formula shows that other features are important as well. F can be non-zero and yet $F_2(-J(Ex_{1,t}x_{1,t}')^{-1}J' + (Ex_{2,t}x_{2,t}')^{-1}) = 0$. In this example, this occurs if $\rho_{zy} = 0$. While it is clearly possible that $F = 0$, it is also true that this example is based on a DGP and models that cause the formula to simplify tremendously. If the quantile errors were not conditionally homo-altitudinal, if the revisions were anything but independent news, or if the models were misspecified, it seems very unlikely that Ω would be zero, because none of the simplifications after the third equality of (4) would exist.[14]

[13]If we evaluate against the revised value y_{t+1} the final result changes only insofar as the term $c\frac{1-\delta_1}{\sigma_\varepsilon}$ becomes $-c\delta_1/\sigma_\varepsilon$. If the two models being compared are simpler cases in which the benchmark is just an intercept and the nesting model is a $QAR(1)$, the final results only change insofar as the term $\frac{\sigma_\varepsilon^2 c^2}{\sigma_y^2}(\frac{\rho_{zy}^2}{1-\rho_{zy}^2})$ becomes c^2.

[14]The emphasis on independence is important. The literature defines a revision (c_t) as news if for $y_t(t) = y_t + c_t$, $Ec_t = 0$ and $cov(y_t(t), c_t) = 0$. In our simple example this is insufficient to make $F_2 = 0$ unless we also know that $y_t(t)$ and c_t are independent. This is not the case in Clark and McCracken (2009). The difference is that the score of a quadratic loss function is linear in the predictors, and that is not the case under tick loss.

4 Monte Carlo Evidence

In the previous section we delineated a framework for tests of predictive ability between two nested quantile regression models when the data are subject to revision. Due to the novelty of this framework, we provide Monte Carlo evidence on the finite sample size and power of the test and do so in the context of the example provided in Section 3.2. While simplistic, this example provides the advantage of having expected loss differentials and asymptotic variances that are analytically tractable, which, in turn, facilitates interpretation of the simulations. Because the models are nested, and the recursive scheme appears to be best when forecasting the lower tails of GDP, we focus on recursively estimated models and rejection frequencies in the upper tail of the asymptotic distribution of the test statistic used to conduct inference. Additional results for two-sided tests or when the rolling estimation scheme is used are provided in Appendix C.

4.1 Monte Carlo Design

There are two DGPs. For each we generate data using independent draws from normal distributions and an assumed autoregressive structure. We focus on one-step-ahead forecasts with initial sample sizes $R \in \{100, 500, 1000\}$ and out-of-sample sizes P such that $P/R \in \{0.5, 1, 2\}$. All results are based on 5000 Monte Carlo draws.

For both DGPs, the fully revised data are generated by

$$y_t = \delta_0 + \delta_1 y_{t-1} + \delta_2 z_{t-1} + u_t$$

$$z_t = \gamma_1 z_{t-1} + v_t,$$

for $t = 1, ..., T$, where $(u_t, v_t)' \sim i.i.d.\mathcal{N}(0, \Sigma)$ and $\Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}$. In most instances, the DGP parameters are loosely based on OLS estimates of AR(1) models for RGDP growth and the 3-month averaged NFCI over a Great Moderation sample.

As we did in the simple example from Section 3.2, we consider a single-period revision process for the dependent variable such that $y_t(t) = y_t(t+1) - c$ but where z is not subject to revision (i.e., $z_t(t) = z_t$). We focus on revision magnitudes of size $c = 0.5\sigma_u$ and $c = \sigma_u$ but will investigate other magnitudes later in the context of the power of the test. Accuracy is evaluated using tick loss for the quantiles $\alpha \in (0.05, 0.1, 0.5)$.[15] For each DGP, two nested quantile regression models are estimated recursively across all forecast origins. Forecasts from these two models are evaluated against the initial release of the dependent variable $y_{t+1}(t+1)$.

DGP1: In the simplest of our two DGPs, the two nested models take the form

$$\text{Restricted} : y_{t+1} = \beta_{0,1} + \varepsilon^{(\alpha)}_{1,t+1}$$

$$\text{Unrestricted} : y_{t+1} = \beta_{0,2} + \beta_{1,2} y_t + \varepsilon^{(\alpha)}_{2,t+1}.$$

For all experiments, we set $\delta_0 = 2$, $\delta_2 = 0$, and $\sigma^2_u = 5$. We set δ_1 to 0 and 0.4 for the size and power experiments, respectively.

DGP2: In the more complex of our two DGPs, the nested models take the form

$$\text{Restricted} : y_{t+1} = \beta_{0,1} + \beta_{1,1} y_t + \varepsilon^{(\alpha)}_{1,t+1}.$$

$$\text{Unrestricted} : y_{t+1} = \beta_{0,2} + \beta_{1,2} y_t + \beta_{2,2} z_t + \varepsilon^{(\alpha)}_{2,t+1}.$$

For all experiments, we set $\delta_0 = 2$, $\delta_1 = 0.2$, $\gamma_1 = 0.9$, $\sigma^2_u = 5$, $\sigma^2_v = 0.05$, and $\rho_{yz} = -0.5$. We set δ_2 to 0 and -1.6 for the size and power experiments, respectively.

In all cases, the null hypothesis of equal expected tick loss is tested using the test statistic:

$$OOS - t = \frac{P^{-1/2} \sum_{t=R}^{T-1} L(\hat{u}^{(\alpha)}_{1,t+h}) - L(\hat{u}^{(\alpha)}_{2,t+h})}{\sqrt{V}}, \tag{5}$$

where V represents an estimate of the asymptotic variance of the numerator. We consider four versions of this test statistic, each with a different estimate of this variance. The first, denoted Θ, uses the (infeasible) estimated variance of the average loss differentials across Monte Carlo draws. The second uses the theoretical asymptotic variance, Ω, given in equation (4) of Section 3.2. The third, $\hat{\Omega}$, uses an estimate of Ω that takes the form

$$\hat{\Omega} = \hat{\lambda}_{hh} \begin{pmatrix} \hat{F_1}', & -\hat{F_2}' \end{pmatrix} \begin{pmatrix} \hat{B}_1 & 0 \\ 0 & \hat{B}_2 \end{pmatrix} \begin{pmatrix} J'\hat{S}_{h_2h_2}J & J'\hat{S}_{h_2h_2} \\ \hat{S}_{h_2h_2}J & \hat{S}_{h_2h_2} \end{pmatrix} \begin{pmatrix} \hat{B}_1 & 0 \\ 0 & \hat{B}_2 \end{pmatrix} \begin{pmatrix} \hat{F}_1 \\ -\hat{F}_2 \end{pmatrix}, \tag{6}$$

where for $P/R = \hat{\pi}$, $\hat{\lambda}_{hh} = 2[1-\hat{\pi}^{-1} ln(1+\hat{\pi})]$, and J is a selection matrix.[16] $\hat{F}_i$ is constructed using $P^{-1} \sum_{t=R}^{T-1} x'_{i,t}(t)[1(y_{t+1}(t+1) \leq x'_{i,t}(t)\hat{\beta}^{(\alpha)}_{i,t}) - \alpha]$. After obtaining the residuals $\hat{\varepsilon}^{(\alpha)}_{i,s+1}$

[15]Note that for this DGP, the conditional quantiles are symmetric around the origin and hence we omit results for the upper quantiles.

[16]Under the rolling scheme, $\hat{\lambda}_{hh}$ takes the values $\hat{\pi} - \hat{\pi}^2/3$ and $1 - (3\hat{\pi})^{-1}$ when $0 \leq \hat{\pi} \leq 1$ and $1 < \hat{\pi} < \infty$ respectively.

when using the entirety of the final vintage to estimate models $i = 1, 2$, we follow White and Kim (2002) and form $\hat{B}_i^{-1}$ using $(2\hat{c}_T T)^{-1} \sum_{s=1}^{T-1} 1(|\hat{\varepsilon}_{i,s+1}^{(\alpha)}| \leq \hat{c}_T) x_{i,s}(T) x'_{i,s}(T)$, with $\hat{c}_T = 1.06\hat{\sigma}_{\varepsilon_i}^{(\alpha)} T^{-\eta}$, and $\eta \in (0, 1/2)$.[17] Finally, we estimate the long-run variance of $\hat{h}_{2,s+1} = (\alpha - 1(\hat{\varepsilon}_{2,s+1}^{(\alpha)} < 0) x_{2,s}(T)$, $\hat{S}_{h_2 h_2}$, using the Bartlett kernel and bandwidth $\lfloor 4(T/100)^{2/9} + 1 \rfloor$. The fourth version of the test statistic simply uses $\hat{S}_{dd}$, the sample variance of the out-of-sample loss differentials. All test statistics are compared against critical values from the standard normal distribution.

4.2 Monte Carlo Results

For all experiments, we report one-sided rejection frequencies at the 5% significance level. Table 3 reports results on the actual size of the test for DGP1. The top section of the panels provide rejection frequencies when using the two valid, but infeasible, versions of the test statistic. In both instances there is a tendency to underreject, more so for smaller sample sizes and quantiles. Between the two, the Ω-based tests arguably perform the best. Regardless, these two are included in the table largely to reinforce the validity of the asymptotic results described in Section 3.1.

As a practical matter, the bottom section of the panels are the ones that are most relevant for a practitioner. Contrary to the aforementioned tests, the $\hat{\Omega}$-based tests have a tendency to overreject, with sizes ranging from 5% to 11%. In accordance with the theory, the actual size of the test tends to improve with the sample size. The actual size seems to be a bit better for the larger revisions, for reasons that are not immediately clear. We also continue to observe that the actual size tends to be a bit worse as we get further into the tails. The final panel reports rejection frequencies for the $\hat{S}_{dd}$-based tests. In most instances, the rejection frequencies are well above those associated with a nominally 5% test. Even worse, the issue deteriorates as the sample size increases, with rejection frequencies nearing 50%. The root of the problem is that $\hat{S}_{dd}$ converges in probability to zero. Therefore, since the numerator of the test statistic in (5) is asymptotically normal with zero mean, as $\hat{S}_{dd}$ converges to zero we expect rejection frequencies to converge to 50% in the upper tail (and 50% in the lower tail were we doing a two-sided test).

[17] We report results for $\eta = 1/40$ but obtain broadly similar results for a range of smaller and larger values. To be clear, the results are sensitive to this parameter. See McCracken (2004) for more discussion on this choice for out-of-sample inference.

Table 4 reports the power results for DGP1. There are three clear trends common among all four of the test statistics. First, as expected, power improves with sample size. Rejection frequencies for the smallest R and P/R settings range from 15% to 97%, whereas for the largest R and P/R settings, they essentially max out at 100%. Second, at least for small samples, actual power tends to decline as we move into the tails. Third, power appears to increase with the magnitude of the revision c. This increase, however, turns out to be misleading - a point we revisit later in this section. Regardless, what is most important is that the $\hat{\Omega}$-based tests have reasonable power across all permutations of sample size and quantile.

Table 5 reports the size results for DGP2. Θ-based rejection frequencies range from 2% to 8% and improve as the sample size increases. For this DGP, Ω-based statistics substantially overreject in the smallest samples, with actual sizes ranging from 12% to 21%. Even so, there is a clear pattern of improvement as either the initial or out-of-sample size increases. $\hat{\Omega}$-based tests exhibit modest overrejections, with values ranging from 6% to 15%. However, in contrast to the size results in Table 2, there is no clear pattern of improvement as the sample size increases. At the lowest quantile and the smallest revision, the actual size tends to rise with the sample size. But for the largest quantile and revision, it is typically the case that the actual size improves as the sample size increases. Finally, as expected, the $\hat{S}_{dd}$-based tests perform poorly. While there are instances where the rejection frequency is between 3% and 6% for the lowest quantile, in nearly all other instances the rejection frequencies rise to the 20% to 50% range as the sample size increases — exactly what we saw in Table 2.

Table 6 reports the power results for DGP2. Focusing first on results with $c = 0.5\sigma_u$, power increases with sample size for each type of test statistic. The Ω- and $\hat{\Omega}$-based tests yield the highest power for all permutations while the infeasible Θ-based tests yield almost no power except when $\alpha = 0.5$ and the sample size is large. In all cases, power increases as α approaches 0.5.

The most striking result from this table is the lack of power when $c = \sigma_u$. In fact, power decreases with the sample size — with reported rejection frequencies of essentially 0% for the largest R and P/R settings. To understand the striking distinctions in power, consider Figure 4. Here, we plot expected tick loss for the restricted and unrestricted models, as a function of c/σ_u. The parameters are the same as those in the DGP2 simulations. There are four subplots corresponding to quantiles of 0.05, 0.1, 0.25, and 0.5. The expected losses of the restricted model are plotted using simulations, whereas the unrestricted model's expected losses are plotted using equation (B.1) from Appendix B. When the unrestricted model's expected losses are lower than that of the restricted model, we should expect power in the right tail of the asymptotic distribution of the test statistic in (5), whereas we should expect power in the left tail when the opposite is true. Unsurprisingly, for each plot the unrestricted model's expected losses are minimized when there are no revisions and increase with the absolute magnitude of the revisions. When $\alpha = 0.5$, expected losses are symmetric, but for $\alpha < 0.5$ the magnitude of expected losses increases more steeply with positive revisions than negative ones. The expected loss for the restricted model varies with α and σ_u in a similar fashion as for the unrestricted model, but are slightly less sensitive to changes in σ_u.

The plots contextualize the rejection frequencies reported in Table 6. Figure 4 reveals that when $c = 0.5\sigma_u$, the expected loss of the unrestricted model is slightly lower than that of the restricted model, and hence the Monte Carlo simulations yield power in the upper tail. Moreover, as α increases the expected loss differential increases as well, and this is reflected in Table 6 – as we move from $\alpha = 0.05$ to $\alpha = 0.5$ the test's power increases. On the other hand, when $c = \sigma_u$, the two models have expected loss differentials of -0.0385, -0.0540, -0.0698, and -0.0598 for the quantiles 0.05, 0.1, 0.25, and 0.5, respectively. In each case, the expected loss differential is negative and thus, as seen in Table 6, we should not expect to see power in the upper tail.

Table 7 reports the rejection frequencies from the same DGP when we use two-sided tests. The rejection frequencies when $c = \sigma_u$ are much higher than those in Table 6, clearly stemming from rejections in the lower tail. Amalgamating the findings in Figure 4 and Tables 6 and 7, we conclude that for our experiment design, in the presence of large revisions, power can actually materialize in the left tail despite the restricted model being incorrectly specified under the alternative.

Altogether, we find that the proposed test of equal tick loss can be reasonably well sized and exhibit substantial power. Even so, the simulation evidence exposes some weaknesses. Actual size and power tend to be worse the further out we go in the tails. Not unexpectedly, the actual size of the test tends to improve with the sample size. In addition, properties of the revision process are likely to affect actual size and power irrespective of sample size — and in the case of power, can lead to unexpected deviations from the null hypothesis. For this reason, in the following empirical section, all tests are conducted against a two-sided alternative.

5 Empirical Results

In this section, we use the vintage data and the approach to inference described above, to evaluate whether in real time the NFCI is a useful tool for monitoring downside risk to RGDP growth. Specifically, we use real-time vintage data for RGDP and the NFCI to construct quantile forecasts using the two nested models described in equations (1) and (2) from Section 3.2. We then evaluate the relative forecast accuracy of these two models, under tick loss, based on the asymptotic inference delineated in Sections 3.1 and 4.

We obtain GNP, GDP, NFCI, and NBER recession data from the FRED and ALFRED databases. The data date back to 1971Q1, and we obtain vintages from January 1988 through August 2021. Throughout we generalize by using the term GDP despite using GNP in vintages before December 1991. Forecast origins take place at the start of the second month of a given quarter, and we evaluate forecasts for the horizons $\tau = 1, ..., 12$, with an emphasis on horizons 1 and 4, those discussed in ABG. For example, the first forecast origin is February 1, 1988, and for $\tau = 1$ (12) the target dates are 1988Q1 (1990Q4). For the NFCI and RTNFCI, we use non-calendar 3-month averages, that is, averages that use the most recent 3 months of observations at a given origin. To evaluate tick loss, we use the initial release of GDP but find comparable results when using the third estimate and a recent vintage (August 26, 2021). Note that our formal evaluation does not include periods associated with the COVID-19 pandemic for reasons that we later clarify.

We evaluate the forecasts using the same approach as was considered in the Monte Carlo simulations from the preceding section. That is, we estimate the asymptotic variance Ω using the nonparametric method presented in (6) and its components as described in the subsequent paragraph. We consider both recursive and rolling forecasts and evaluate each using the same methods, with the distinction that the respective formulas for λ_{hh} differ. For some window sizes and horizons, the first forecast origin takes places later than February 1, 1988, due to an insufficient number of observations for the direct multi-step estimation. For example, when $R = 90$ and $\tau = 1$ the first forecast origin is November 1, 1993, and when $R = 60$ and $\tau = 12$, the first forecast origin is February 1, 1989.

Tables 8 and 9 present results for the comparison of the two nested quantile forecasting models when using the recursive and rolling schemes, respectively. In the first section of panels we use the current vintage NFCI (i.e., EXNFCI), while in the second we use the real-time vintages of the NFCI (i.e., RTNFCI). For a given unrestricted model, window size, α quantile, and horizon, we report the percentage change in the average tick loss between the unrestricted and restricted models as well as the p-values associated with the $OOS - t$ test statistic delineated in (5). Note that we use a two-sided test due to our findings in Section 4, which indicated that revisions can result in the restricted model outperforming the unrestricted model even when the NFCI has predictive content. While our focus is on the lower tails of the conditional distribution of RGDP growth, for completeness we consider two quantiles in the lower and upper tails as well as the median. In addition, for robustness, we consider three distinct initial sample sizes, 30, 60, and 90, for estimation of the parameters. For brevity, we only report the results for horizons 1 and 4, but discuss results for other horizons in subsequent exercises.

Focusing first on Table 8, we find that both the EXNFCI and RTNFCI models (i.e., the unrestricted models) tend to perform better than the QAR(1) model in the lower tail at both horizons. In all cases, performance gains for $R = 60$ and 90 are larger than that for $R = 30$, with the larger window sizes displaying statistical significance in all cases of the lowest two reported α quantiles. In most cases, we also find significant deterioration in the unrestricted models' performance relative to the QAR(1) in the upper tail. For the most part, the results in Table 9 mirror those in Table 8. There are a few exceptions, however. In the lower tails, the unrestricted models tend to have the lowest relative average tick loss for $R = 30$. Perhaps most notably, for some permutations of the upper tail forecasts, the EXNFCI and RTNFCI models outperform the QAR(1) at a statistically significant level.

Recall that these results do not include the COVID-19 period; the reasons for which are illustrated in Figure 5. Each panel gives the cumulative sum of the loss differentials between two specified models for a given α-quantile and forecast horizons 1 and 4. The first, second, and third columns give the differentials between the QAR(1) and the EXNFCI, the QAR(1) and the RTNFCI, and the EXNFCI and the RTNFCI models, respectively. The first, second, and third rows give the differentials for α quantiles 0.05, 0.5, and 0.95, respectively. We choose to report the results when the models are estimated using the recursive scheme and $R = 60$. The first choice is due to the recursive scheme's superiority over the rolling scheme in the lower tails, while the second choice strikes a balance between using a larger window size and maximizing the availability of forecast origins.[18]

The trends in the first two panels of the first row are unsurprisingly similar. Absent the COVID-19 period, there is a clear and continued benefit to appending the EXNFCI or RTNFCI to the QAR(1) model. In the COVID-19 period, the cumulative sum of the loss differentials for both horizons in these two panels experience steep declines, suggesting that the unrestricted models performed worse than the QAR(1). Moreover, the third panel of the first row suggests that the RTNFCI performed slightly better than the EXNFCI for $\tau = 4$, whereas the two models performed similarly for $\tau = 1$. Finally, reinforcing earlier findings, the second and third rows of Figure 5 indicate that neither of the unrestricted models provides performance advantages over the QAR(1) in the median and upper quantiles.

In the lower tail, it appears that some of the largest improvements in the RTNFCI model occur near recessions, especially at the longer horizon. On the other hand, the differentials between the QAR(1) and EXNFCI seem to change less predictably near recessions. To get a better feel for this consider Figure 6. In the first panel, we plot the average percentage change in relative tick loss when $\alpha = 0.05$ at horizons ranging from 1 to 12 quarters ahead of any recessionary period.[19] On average we find that the RTNFCI model is more accurate than the QAR(1) model for horizons roughly a year out from any recessionary period but then deteriorates. However, the EXNFCI model does worse than the QAR(1) at every horizon.[20] Our findings suggest that the RTNFCI performs better than the EXNFCI at

[18]Specifically, when $R = 90$, the first origin takes place after the 1990-91 recession. In unreported results, we find that using $R = 90$ gives similar results for overlapping forecast periods.

[19]We compute these results with an OLS regression where the LHS is the loss differentials between the two models scaled by the average loss of the denominator model (noted in the legend) and the RHS is an intercept and a recession dummy variable that is equal to 1 when a recession period is anywhere between the forecast origin and target date. A little algebra shows that adding the resulting intercept and slope coefficient gives the average relative percentage change in tick loss between the two models for forecasts when the recession dummy is equal to 1.

[20]As can be seen in Figure 5, a large contributor to this outcome is the 1990-91 recession period, where the NFCI model performs much worse than the QAR(1). In unreported results, we compute these results

horizons less than 10 quarters leading into any recessionary period.

In the second panel of Figure 6, we report results from a comparable exercise but in the context of predicting the likelihood of a recession rather than the lower tail of RGDP growth. Specifically, we use maximum likelihood to estimate three Probit models: the baseline that only includes an intercept and one lag of the current vintage of RGDP, one that adds one lag of the current vintage of the NFCI, and one that adds one lag of the RTNFCI. Quality of the predictions is evaluated using the area under the receiving operator characteristics curve (AUROC).[21] Briefly, the AUROC is a measure of the quality of the trade-off between the true and the false positive rates of a model's predictions of the likelihood of a recession. A value greater than 50% means the model is better than a random guess, while one less than 50% implies the model is worse. The panel indicates that all three models are an improvement over simply guessing at the shortest horizons. Even so, the baseline quickly loses predictive content, and the model including the EXNFCI loses most of its predictive content after one year. In contrast, the RTNFCI model retains predictive content up to nearly two years.[22] This finding stands in sharp contrast to arguments made by Reichlin, Ricco, and Hasenzagl (2020) that the NFCI has "little advanced information of recessions beyond what is already included in real economic indicators."

6 Conclusion

In this paper we emphasize the real-time nature of the growth-at-risk literature and do so in two ways. First, we construct real-time vintages of the NFCI using the same methodology of its authors, and demonstrate that these vintages possess similar characteristics to those of their official counterpart. We then describe how to properly implement tests of equal predictive ability between nested quantile regression models when the data are subject to revision, and show that these tests display reasonable size and power properties in finite samples. With these two tools in hand, we then use the unofficial real-time vintages, in conjunction with the already available official vintages, to evaluate the out-of-sample predictive content of the NFCI while avoiding the look-ahead bias present in previous exercises

when origins before 1992 are excluded and find that the NFCI performs better than the QAR(1) for $\tau = 1, 2,$ and 4.

[21] See Berge and Jordà (2011) for further discussion on ROC curves.

[22] We reach similar conclusions when using the diagonal elementary score of Bouallègue et al. (2019).

of this nature. We find that accounting for the real-time nature of the monitoring process does not significantly alter the results found in Adrian, Boyarchenko, and Giannone (2019). Real-time quantile forecasts of the lower tail of U.S. RGDP growth are more accurate when you include the NFCI, even when conditioning on lagged RGDP growth. Finally, despite the fact that the real-time vintages exhibit more variability and are subject to revision, we find evidence that the real-time vintages provide improved forecast accuracy leading into recessions — precisely when monitoring downside risks is most important.

References

Adrian, Tobias, Boyarchenko, Nina, and Domenico Giannone (2019), "Vulnerable Growth," *American Economic Review* 109, 1263-1289.

Alessandri, Piergiorgio, and Haroon Mumtaz (2017), "Financial Conditions and Density Forecasts for US Output and Inflation," *Review of Economic Dynamics* 24, 66-78.

Berge, Travis and Òscar Jordà (2011), "Evaluating the Classification of Economic Activity into Recessions and Expansions," *American Economic Journal: Macroeconomics* 3, 246-277.

Bouallègue, Zied Ben, Magnusson, Linus, Haiden, Thomas, and David S. Richardson (2019), "Monitoring trends in ensemble forecast performance focusing on surface variables and high-impact events," Quarterly Journal of the Royal Meteorological Society, 144, 1463-1473.

Brave, Scott A., and Andrew R. Butters (2012), "Diagnosing the Financial System: Financial Conditions and Financial Stress," *International Journal of Central Banking* 8, 191-239.

Brave, Scott A., Butters, Andrew R., and David Kelley (2020), "A Practitioner's Guide and MATLAB Toolbox for Mixed Frequency State Space Models," *Journal of Statistical Software*, forthcoming.

Brownless, Christian, and André Souza (2021), "Backtesting Global Growth-at-Risk," *Journal of Monetary Economics* 118, 312-30.

Caldara, Dario, Cascaldi-Garcia, Danilo, Cuba-Borda, Pablo, and Francesca Loria (2020), "Understanding Growth-at-Risk; A Markov Switching Approach," Working Paper.

Carriero, Andrea, Clark, Todd E., and Massimiliano Marcellino (2020), "Nowcasting Tail Risks to Economic Activity with Many Indicators," Federal Reserve Bank of Cleveland Working Paper 20-13R2.

Catherine, Doz, Giannone, Domenico, and Lucrezia Reichlin (2006), "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models," ECB Working Paper No. 674.

Clark, Todd E., and Michael W. McCracken (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics* 105, 85-110.

Clark, Todd E., and Michael W. McCracken (2009), "Tests of Equal Predictive Ability with Real-Time Data," *Journal of Business and Economic Statistics* 27, 441-54.

Coe, Patrick J., and Shaun P. Vahey (2020), "Financial Conditions and the Risks to Economic Growth in the United States since 1875," CAMA Working Paper No. 36.

Corradi, Valentina, Fosten, Jack, and Daniel Gutknecht (2020), "Conditional Quantile Coverage: An Application to Growth-at-Risk," Working Paper.

De Nicoló, Gianni, and Marcella Lucchetta (2017), "Forecasting Tail Risks," *Journal of Applied Econometrics* 32, 159-170.

Delle Monache, Davide, De Polis, Andrea, and Ivan Petrella (2021), "Modeling and Forecasting Macroeconomic Downside Risk," Bank of Italy Working Paper No. 1324.

Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253-263.

Ferrara, Laurent, Mogliani, Matteo, and Jean-Guillaume Sahuc (2020), "High-Frequency Monitoring of Growth-at-Risk," CAMA Working Paper 97/2020.

Giacomini, Rafaella, and Barbara Rossi (2010), "Forecast Comparisons in Unstable Environments," *Journal of Applied Econometrics* 25, 595-620.

Giacomini, Rafaella, and Halbert White (2006), "Tests of Conditional Predictive Ability," *Econometrica* 74, 1545-1578.

Harvey, Andrew C. (1989), "Forecasting, Structural Time Series Models and the Kalman Filter," Cambridge University Press.

McCracken, Michael W. (2000), "Robust Out-of-Sample Inference," *Journal of Econometrics* 99, 195-223.

McCracken, Michael W. (2004), "Parameter Estimation and Tests of Equal Forecast Accuracy between Non-nested Models," *International Journal of Forecasting* 20, 503-514.

McCracken, Michael W. (2007), "Asymptotics for Out-of-Sample Tests of Granger Causality," *Journal of Econometrics* 140, 719-752.

Reichlin, Lucrezia, Ricco, Giovanni, and Thomas Hasenzagl (2020), "Financial Variables as Predictors of Real Growth Vulnerability," Deutsche Bundesbank Discussion Papers 05/2020.

Rossi, Barbara, and Tatevik Sekhposyan (2011), "Understanding Models' Forecasting Performance," *Journal of Econometrics* 164, 158-172.

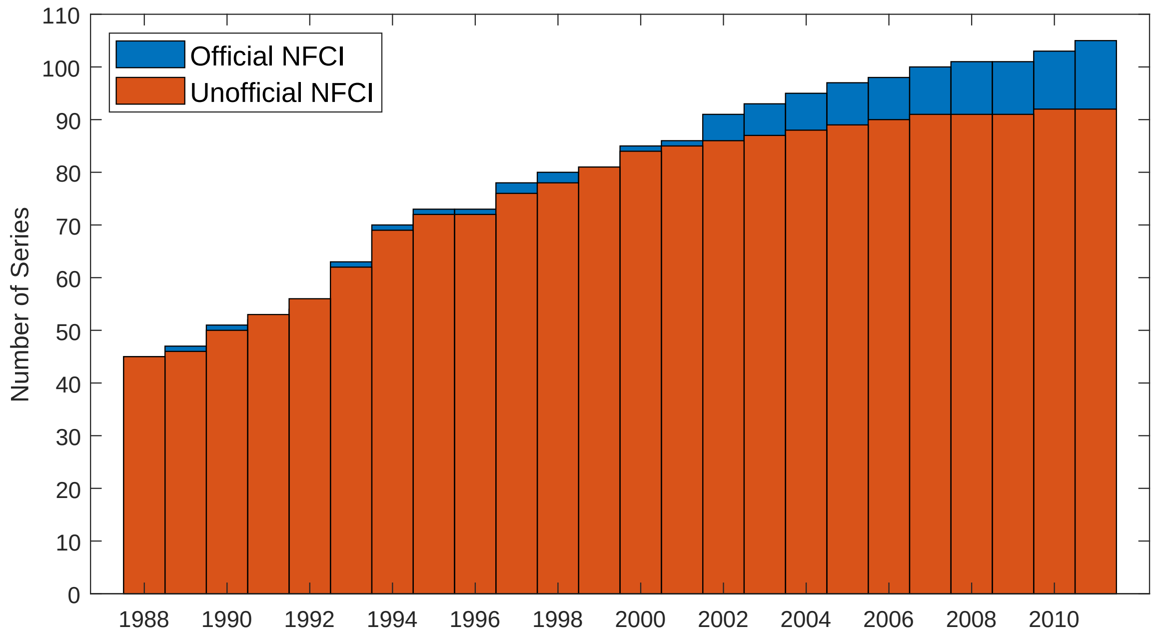
Shumway, Robert H., and David S. Stoffer (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis* 3, 253-264.

Watson, Mark W., and Robert F. Engle (1983), "Alternative Algorithms for the Estimation of Dynamic Factor, Mimic and Varying Coefficient Regression Models" *Journal of Econometrics* 23, 385-400.

West, Kenneth D. (1996), "Asymptotic Inference About Predictive Ability," *Econometrica* 64, 1067-1084.

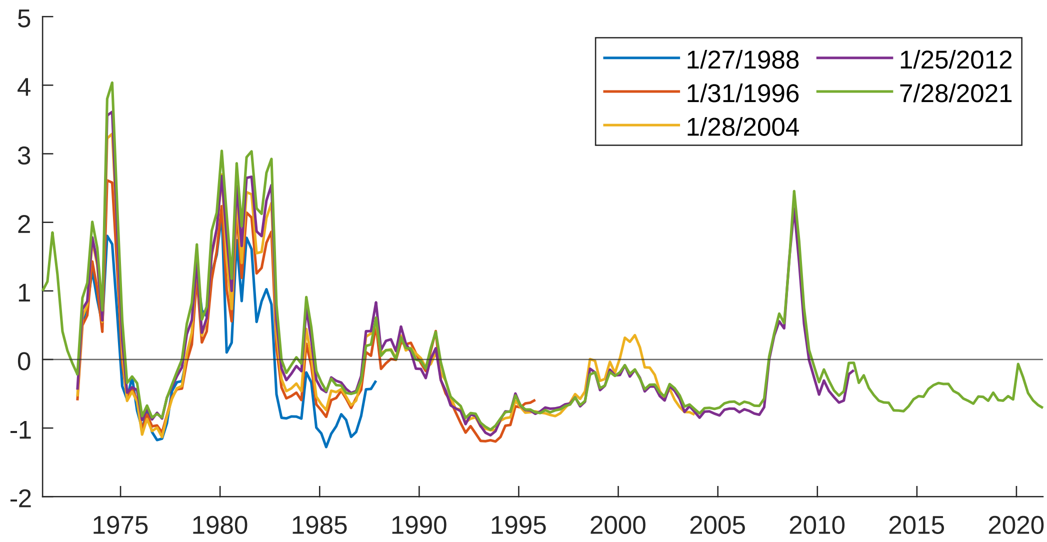
White, Halbert and Tae-Hwan Kim (2002), "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regression," University of California at San Diego, Economics Working Paper Series.

Figure 1: Number of Series Available by Vintage Date



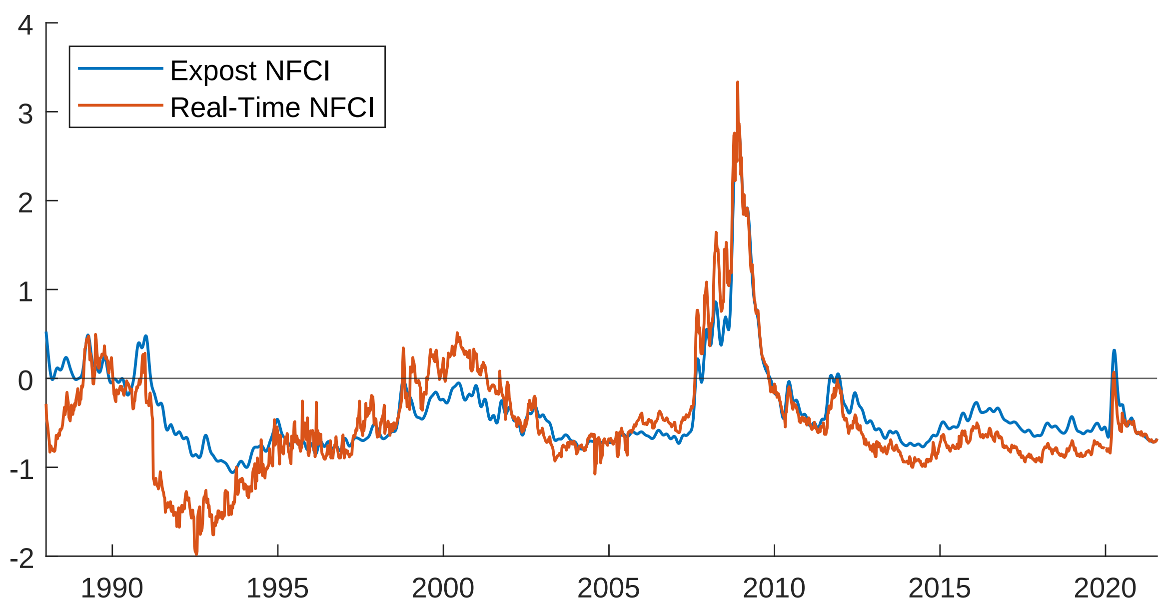
Notes: This figure provides the number of series available for the first vintage of each year in our unofficial NFCI as well as that of the hypothetical official NFCI vintages that would have been released before May 25, 2011. For a given vintage, series are only considered available if at least two years worth of observations would have been released.

Figure 2: NFCI 3-Month Average Vintage Comparisons



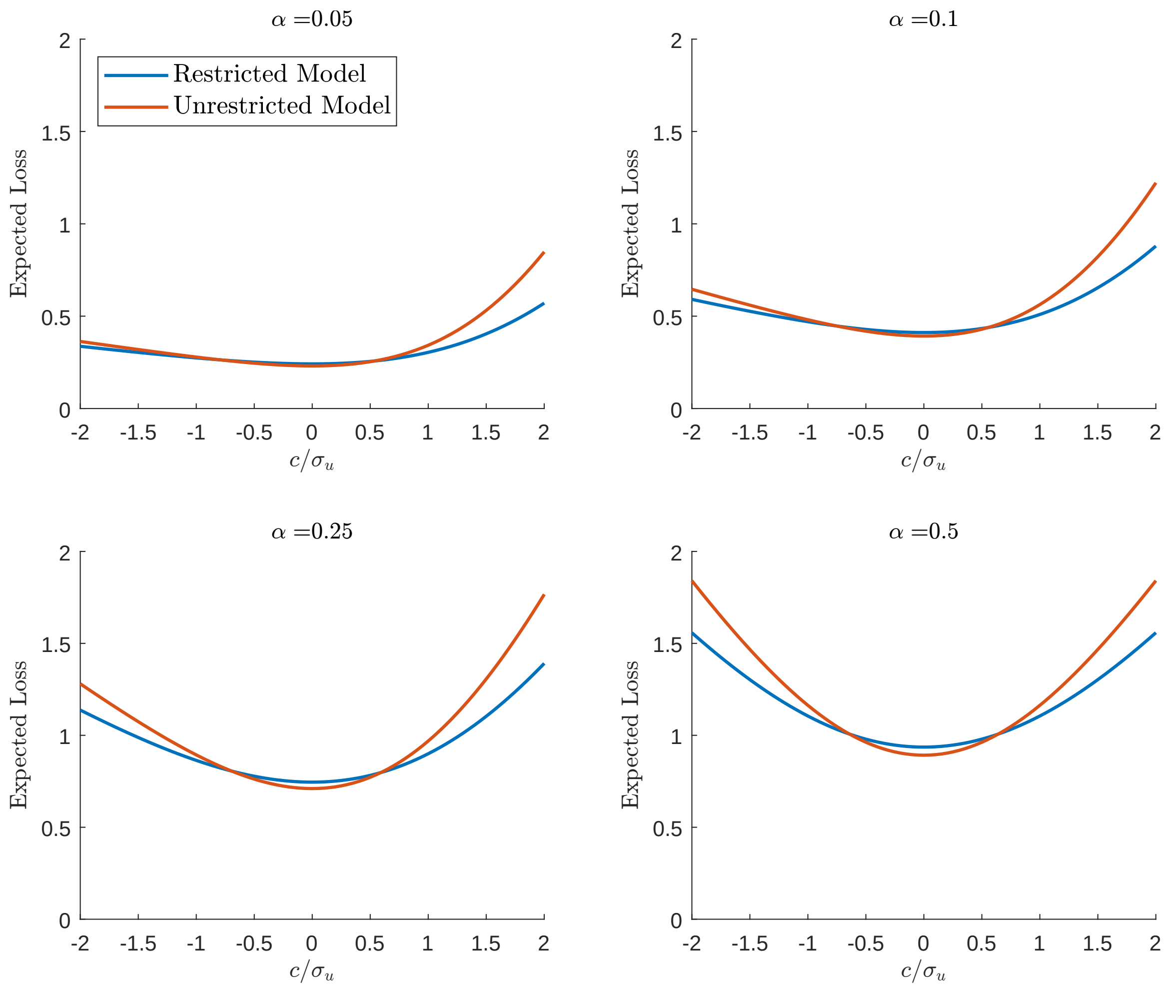
Notes: This figure compares the 3-month averages of five different NFCI vintages. The 1988, 1996, and 2004 vintages are unofficial vintages that we constructed, while the 2012 and 2021 vintages are official vintages produced by the NFCI curators.

Figure 3: Expost vs. Real-Time NFCI



Notes: This figure compares a plot of a recent vintage of the NFCI (i.e., Expost), to a plot of the endpoints of each consecutive NFCI vintage (i.e., Real-Time).

Figure 4: Expected Losses, Power DGP2

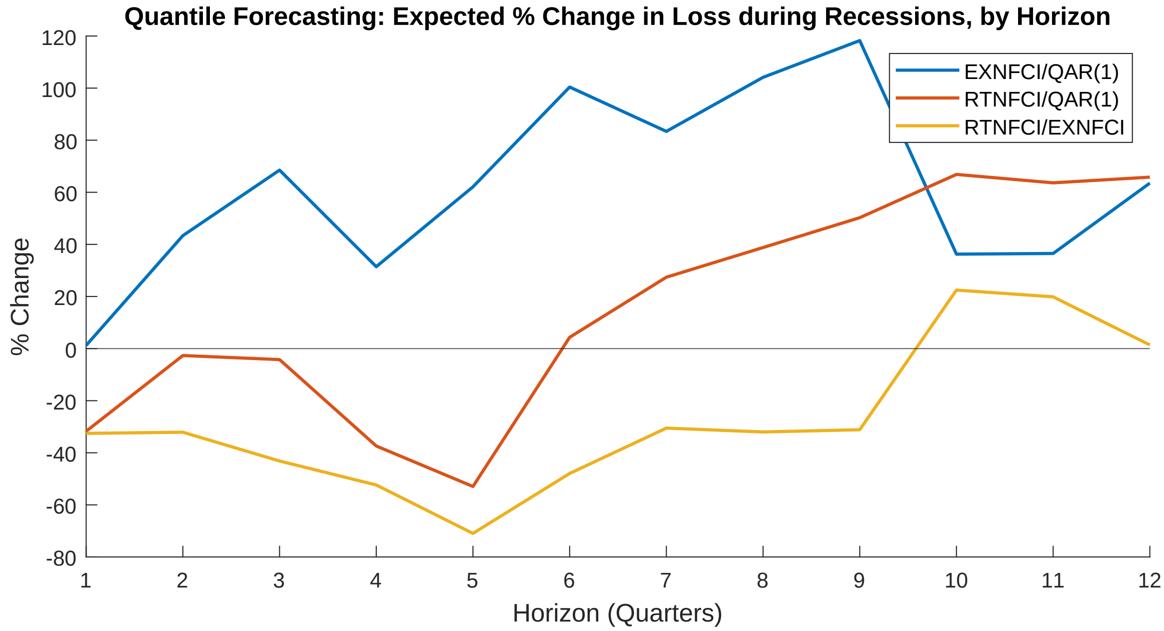


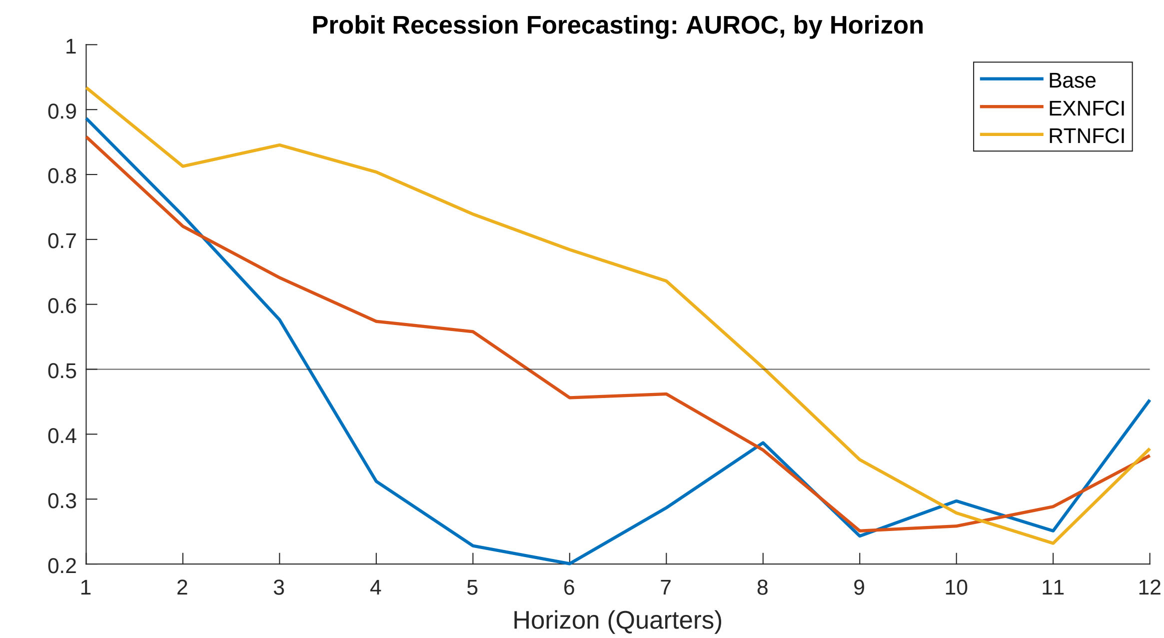
Notes: This figure plots the expected tick loss of the restricted and unrestricted models of the DGP2 power simulations over a range of revision sizes. Each panel provides the results for a specified α-quantile.

Figure 5: Cumulative Sum of Loss Differentials

Notes: Each subplot gives the cumulative sum of loss differentials between the specified models over the forecast origins 1988Q1-2021Q2. For example, plots with the title "QAR(1) vs EXNFCI" give the cumulative sum of loss differentials between the QAR(1) and EXNFCI models. We only report results from the recursive window estimation scheme with a window size of R = 60. The first, second, and third rows of plots give the results when $\alpha = 0.05$, $\alpha = 0.5$, and $\alpha = 0.95$, respectively.

Figure 6: Recession-Related Model Performance





Notes: In the first panel, each line represents the expected percentage change of the tick loss between the two specified $\alpha = 0.05$ GDP quantile forecasting models over a range of horizons. For example, for one-step-ahead forecasts, the RTNFCI model is associated with $\sim 30\%$ lower tick loss than in the QAR(1) and EXNFCI models. Results are only reported for the recursive estimation scheme with a window size of R = 60. In the second panel, each line represents the AUROC values associated with each recession forecasting model over a range of horizons.

Table 1: 3-Month Averages of NFCI Vintages, Correlations

	Jan-88	**Jan-90**	**Jan-92**	**Jan-94**	**Jan-96**	**Jan-98**	**Jan-00**	**Jan-02**	**Jan-04**	**Jan-06**	**Jan-08**	**Jan-10**	**Jan-12**	**Jan-14**	**Jan-16**	**Jan-18**	**Jan-20**
Jan-88	1.00	1.00	0.98	0.92	0.95	0.91	0.93	0.94	0.94	0.94	0.93	0.91	0.92	0.92	0.92	0.93	0.93
Jan-90		1.00	0.97	0.91	0.95	0.90	0.92	0.94	0.93	0.94	0.93	0.92	0.92	0.92	0.92	0.93	0.93
Jan-92			1.00	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.97	0.97	0.96
Jan-94				1.00	0.98	1.00	0.98	0.97	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.97
Jan-96					1.00	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.97
Jan-98						1.00	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98
Jan-00							1.00	1.00	1.00	0.99	0.99	0.98	0.98	0.98	0.98	0.97	0.96
Jan-02								1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.98	0.97	0.97
Jan-04									1.00	1.00	1.00	0.99	0.99	0.98	0.98	0.98	0.97
Jan-06										1.00	1.00	0.99	0.99	0.99	0.98	0.98	0.98
Jan-08											1.00	0.99	0.99	0.98	0.98	0.97	0.97
Jan-10												1.00	0.99	1.00	1.00	0.99	0.99
Jan-12													1.00	1.00	1.00	0.99	0.98
Jan-14														1.00	1.00	1.00	0.99
Jan-16															1.00	1.00	0.99
Jan-18																1.00	1.00
Jan-20																	1.00

Notes: This table provides the correlation coefficients between 3-month averages of NFCI vintages. For brevity, we only include the first vintage of even years.

Table 2: ANOVA (N = 60)

	$h=1$					$h=4$				
Treatment\ α	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
Rolling	0.111*	0.053*	-0.039*	-0.148*	-0.094*	0.061	0.102*	-0.033*	0.014	0.011
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.09)	(0.00)	(0.01)	(0.75)	(0.87)
R = 60	-0.132	-0.049	0.022*	0.05	0.064	-0.193	-0.132	-0.017*	-0.105	-0.156
	(0.08)	(0.66)	(0.00)	(0.08)	(0.06)	(0.66)	(0.56)	(0.01)	(0.68)	(0.84)
R = 90	-0.154*	-0.115*	-0.036*	0.033	0.012	-0.353*	-0.232*	-0.105*	-0.171*	-0.284*
	(0.00)	(0.00)	(0.00)	(0.13)	(0.73)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
GDP	-0.052	-0.032	-0.033*	-0.09*	-0.064*	-0.034	-0.028	-0.028*	-0.02	-0.005
	(0.08)	(0.09)	(0.00)	(0.00)	(0.03)	(0.34)	(0.28)	(0.03)	(0.65)	(0.94)
EXNFCI	-0.578*	-0.388*	0.007	0.047	0.077	-0.49	-0.456*	0.048*	0.113	0.135
	(0.00)	(0.00)	(0.86)	(0.19)	(0.20)	(0.22)	(0.00)	(0.02)	(0.63)	(0.97)
RTNFCI	-0.54*	-0.35*	0.019	0.036	0.072	-0.648*	-0.539*	0.026	0.127*	0.155
	(0.00)	(0.00)	(0.08)	(0.19)	(0.09)	(0.00)	(0.00)	(0.06)	(0.03)	(0.08)
EXNFCI*Cal	0.082	0.038	0.006	-0.004	0.01	0.072	0.052	0.015	-0.005	-0.058
	(0.08)	(0.20)	(0.56)	(0.89)	(0.82)	(0.20)	(0.20)	(0.48)	(0.94)	(0.58)
RTNFCI*Cal	0.002	-0.005	-0.003	-0.008	-0.01	0.07	0.063	0.018	0.012	0.016
	(0.96)	(0.87)	(0.79)	(0.78)	(0.82)	(0.22)	(0.13)	(0.37)	(0.87)	(0.88)

Notes: This table provides the coefficients and p-values from an ANOVA regression on the natural logarithm of mean tick loss. Coefficients represent the percentage change in mean loss associated with the indicator. Asterisks indicate statistically significant values at the 5% level.

Table 3: Size Tests, DGP1

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.01	0.01	0.01	0.01	0.01	0.01	0.06	0.04	0.03	0.02	0.01	0.01
	500	0.02	0.02	0.02	0.02	0.02	0.02	0.04	0.04	0.03	0.03	0.02	0.02
	1000	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.03	0.03	0.03
0.1	100	0.02	0.01	0.02	0.02	0.01	0.01	0.05	0.04	0.03	0.02	0.02	0.02
	500	0.03	0.03	0.03	0.03	0.03	0.03	0.05	0.04	0.04	0.04	0.03	0.03
	1000	0.03	0.03	0.03	0.04	0.04	0.04	0.05	0.04	0.04	0.04	0.04	0.04
0.5	100	0.03	0.03	0.02	0.03	0.03	0.03	0.06	0.04	0.03	0.04	0.03	0.03
	500	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.04	0.04	0.04	0.04
	1000	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.04	0.04	0.05	0.04	0.04

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.11	0.10	0.09	0.10	0.09	0.08	0.03	0.03	0.04	0.16	0.22	0.29
	500	0.10	0.09	0.08	0.08	0.07	0.06	0.11	0.16	0.22	0.41	0.44	0.45
	1000	0.09	0.09	0.07	0.07	0.06	0.06	0.20	0.27	0.33	0.46	0.47	0.47
0.1	100	0.10	0.10	0.10	0.09	0.08	0.07	0.04	0.05	0.08	0.23	0.30	0.35
	500	0.10	0.09	0.08	0.07	0.07	0.06	0.17	0.23	0.29	0.44	0.46	0.46
	1000	0.09	0.08	0.07	0.06	0.06	0.06	0.26	0.33	0.38	0.47	0.48	0.48
0.5	100	0.11	0.10	0.09	0.08	0.07	0.06	0.08	0.11	0.16	0.34	0.39	0.42
	500	0.08	0.07	0.07	0.06	0.06	0.05	0.25	0.31	0.36	0.44	0.45	0.45
	1000	0.07	0.07	0.07	0.06	0.06	0.05	0.34	0.38	0.41	0.45	0.45	0.46

Notes: The DGP is delineated in Section 3.2. For size tests, $\delta_1 = 0$. α represents the quantile that each model is forecasting. R and P give the number of observations in the first forecast origin and number of forecasts, respectively. c defines the revision size. Θ, Ω, $\hat{\Omega}$, and $\hat{S}_{dd}$ are the simulation-based, theoretical, non-parametric, and sample variances, respectively. All tests are compared against one-sided standard normal critical values. The number of Monte Carlo replications is 5000, and the nominal size is 5%.

Table 4: Power Tests, DGP1

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.15	0.23	0.41	0.28	0.43	0.67	0.50	0.59	0.74	0.33	0.39	0.53
	500	0.52	0.81	0.98	0.89	0.99	1.00	0.92	0.99	1.00	0.94	0.99	1.00
	1000	0.82	0.98	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
0.1	100	0.20	0.33	0.58	0.38	0.60	0.86	0.62	0.75	0.89	0.50	0.63	0.81
	500	0.70	0.95	1.00	0.98	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00
	1000	0.95	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99
0.5	100	0.40	0.63	0.89	0.70	0.91	0.99	0.87	0.96	1.00	0.89	0.97	1.00
	500	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.55	0.69	0.84	0.69	0.80	0.90	0.21	0.34	0.58	0.54	0.79	0.96
	500	0.97	1.00	1.00	1.00	1.00	1.00	0.74	0.94	1.00	0.99	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00
0.1	100	0.71	0.85	0.95	0.84	0.92	0.97	0.30	0.51	0.78	0.73	0.93	0.99
	500	0.99	1.00	1.00	1.00	1.00	1.00	0.88	0.99	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	0.99
0.5	100	0.91	0.98	1.00	0.97	0.99	1.00	0.62	0.85	0.98	0.95	1.00	1.00
	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	0.99	1.00	0.99	0.99	1.00	1.00	0.99	1.00	1.00	0.99

Notes: See the notes to Table 3. For power tests, $\delta_1 = 0.4$.

Table 5: Size Tests, DGP2

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.03	0.02	0.02	0.04	0.04	0.03	0.19	0.14	0.10	0.12	0.09	0.08
	500	0.03	0.03	0.03	0.04	0.04	0.04	0.11	0.08	0.06	0.07	0.05	0.05
	1000	0.03	0.03	0.03	0.04	0.04	0.04	0.08	0.06	0.05	0.06	0.06	0.05
0.1	100	0.03	0.03	0.03	0.05	0.05	0.05	0.19	0.15	0.11	0.12	0.10	0.08
	500	0.03	0.03	0.03	0.05	0.05	0.05	0.11	0.08	0.07	0.08	0.07	0.07
	1000	0.04	0.04	0.04	0.05	0.05	0.05	0.08	0.07	0.06	0.07	0.06	0.06
0.5	100	0.05	0.05	0.05	0.08	0.08	0.08	0.21	0.17	0.12	0.16	0.14	0.12
	500	0.05	0.05	0.05	0.07	0.07	0.07	0.12	0.10	0.08	0.09	0.08	0.08
	1000	0.05	0.05	0.05	0.06	0.06	0.06	0.10	0.08	0.07	0.08	0.08	0.07

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.07	0.07	0.06	0.11	0.11	0.11	0.03	0.03	0.03	0.12	0.16	0.21
	500	0.09	0.09	0.10	0.12	0.11	0.11	0.06	0.08	0.10	0.30	0.35	0.40
	1000	0.10	0.10	0.11	0.11	0.10	0.10	0.10	0.13	0.17	0.38	0.43	0.45
0.1	100	0.07	0.07	0.07	0.13	0.13	0.12	0.04	0.04	0.05	0.19	0.25	0.31
	500	0.09	0.09	0.09	0.11	0.10	0.09	0.09	0.11	0.14	0.37	0.42	0.46
	1000	0.10	0.10	0.09	0.09	0.08	0.08	0.14	0.18	0.23	0.43	0.47	0.49
0.5	100	0.09	0.08	0.09	0.14	0.15	0.12	0.10	0.11	0.11	0.35	0.41	0.47
	500	0.11	0.12	0.10	0.10	0.08	0.07	0.18	0.22	0.28	0.49	0.52	0.53
	1000	0.10	0.10	0.08	0.07	0.07	0.06	0.24	0.28	0.32	0.50	0.52	0.52

Notes: See the notes to Table 3. For size tests, $\delta_2 = 0$.

Table 6: Power Tests, DGP2

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.03	0.03	0.03	0.01	0.00	0.00	0.30	0.28	0.26	0.08	0.05	0.03
	500	0.05	0.05	0.05	0.00	0.00	0.00	0.38	0.37	0.38	0.03	0.01	0.00
	1000	0.06	0.06	0.06	0.00	0.00	0.00	0.43	0.42	0.44	0.01	0.00	0.00
0.1	100	0.05	0.04	0.04	0.01	0.01	0.00	0.33	0.32	0.33	0.09	0.05	0.02
	500	0.06	0.07	0.08	0.00	0.00	0.00	0.44	0.45	0.48	0.03	0.01	0.00
	1000	0.08	0.09	0.12	0.00	0.00	0.00	0.50	0.51	0.54	0.01	0.00	0.00
0.5	100	0.07	0.08	0.11	0.02	0.01	0.00	0.44	0.46	0.51	0.12	0.08	0.05
	500	0.12	0.18	0.28	0.00	0.00	0.00	0.60	0.66	0.75	0.05	0.02	0.00
	1000	0.18	0.29	0.45	0.00	0.00	0.00	0.70	0.78	0.87	0.02	0.00	0.00

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.15	0.16	0.19	0.10	0.09	0.07	0.08	0.07	0.07	0.06	0.05	0.04
	500	0.32	0.34	0.35	0.06	0.03	0.01	0.11	0.12	0.13	0.02	0.01	0.00
	1000	0.39	0.38	0.39	0.03	0.01	0.00	0.12	0.13	0.15	0.01	0.00	0.00
0.1	100	0.19	0.22	0.26	0.12	0.10	0.06	0.09	0.10	0.11	0.07	0.07	0.05
	500	0.39	0.40	0.42	0.06	0.02	0.01	0.13	0.15	0.19	0.02	0.01	0.00
	1000	0.44	0.45	0.47	0.02	0.00	0.00	0.17	0.19	0.24	0.01	0.00	0.00
0.5	100	0.30	0.35	0.42	0.17	0.14	0.08	0.16	0.19	0.23	0.12	0.10	0.07
	500	0.52	0.57	0.65	0.08	0.03	0.01	0.26	0.34	0.46	0.04	0.02	0.00
	1000	0.62	0.70	0.79	0.03	0.00	0.00	0.35	0.47	0.63	0.01	0.00	0.00

Notes: See the notes to Table 3. For power tests, δ_2 = -1.6.

Table 7: Two-Sided Power Tests, DGP2

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.07	0.07	0.06	0.09	0.10	0.13	0.69	0.66	0.61	0.42	0.41	0.42
	500	0.06	0.05	0.05	0.18	0.24	0.35	0.72	0.66	0.61	0.63	0.67	0.76
	1000	0.06	0.05	0.05	0.30	0.42	0.59	0.73	0.67	0.64	0.77	0.86	0.93
0.1	100	0.06	0.06	0.06	0.09	0.10	0.13	0.68	0.66	0.62	0.45	0.43	0.46
	500	0.06	0.06	0.06	0.21	0.29	0.43	0.74	0.68	0.65	0.69	0.75	0.84
	1000	0.06	0.06	0.08	0.36	0.53	0.74	0.76	0.70	0.67	0.83	0.91	0.96
0.5	100	0.06	0.06	0.07	0.08	0.09	0.13	0.71	0.69	0.66	0.50	0.49	0.50
	500	0.08	0.12	0.19	0.19	0.29	0.47	0.79	0.77	0.79	0.73	0.77	0.86
	1000	0.11	0.19	0.34	0.34	0.55	0.78	0.83	0.83	0.87	0.85	0.92	0.96

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.26	0.26	0.28	0.31	0.33	0.38	0.07	0.08	0.09	0.22	0.34	0.47
	500	0.48	0.46	0.43	0.68	0.73	0.84	0.10	0.11	0.12	0.56	0.71	0.86
	1000	0.55	0.50	0.47	0.85	0.91	0.96	0.11	0.12	0.14	0.73	0.88	0.96
0.1	100	0.28	0.29	0.32	0.34	0.37	0.41	0.08	0.09	0.12	0.28	0.41	0.53
	500	0.51	0.47	0.46	0.75	0.80	0.88	0.12	0.13	0.16	0.62	0.77	0.90
	1000	0.57	0.52	0.49	0.88	0.94	0.98	0.14	0.15	0.19	0.78	0.92	0.98
0.5	100	0.32	0.37	0.41	0.38	0.39	0.42	0.15	0.18	0.20	0.40	0.48	0.56
	500	0.56	0.57	0.63	0.76	0.78	0.87	0.22	0.29	0.39	0.65	0.76	0.89
	1000	0.65	0.69	0.76	0.88	0.93	0.96	0.29	0.40	0.55	0.78	0.92	0.97

Notes: See the notes to Table 6. All tests are compared against two-sided standard normal critical values.

Table 8: CM Test of Equal Tick Loss, Recursive

		h = 1					h = 4				
Model	$R \setminus \alpha$	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
EXNFCI	30	-0.101*	-0.179*	0.004	-0.057*	0.005*	0.026	-0.108	0.015*	0.068*	0.15*
		(0.00)	(0.00)	(0.36)	(0.00)	(0.00)	(0.51)	(0.06)	(0.00)	(0.03)	(0.00)
	60	-0.169*	-0.231*	0.013	0.015	0.013*	-0.079*	-0.142*	0.081*	0.036*	0.062
		(0.00)	(0.00)	(0.14)	(0.52)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.15)
	90	-0.132*	-0.207*	-0.028*	0.014	0.024*	-0.134*	-0.172*	0.045	-0.01*	0.1*
		(0.00)	(0.00)	(0.00)	(0.55)	(0.00)	(0.00)	(0.00)	(0.19)	(0.00)	(0.01)
RTNFCI	30	-0.104*	-0.152*	-0.017	-0.042*	0.008	-0.104*	-0.215*	-0.029	0.073*	0.168*
		(0.00)	(0.00)	(0.27)	(0.01)	(0.20)	(0.00)	(0.00)	(0.48)	(0.00)	(0.00)
	60	-0.133*	-0.194*	0.039	0.019	0.023*	-0.169*	-0.231*	0.075	0.053*	0.051
		(0.00)	(0.00)	(0.18)	(0.60)	(0.00)	(0.00)	(0.00)	(0.22)	(0.01)	(0.43)
	90	-0.099*	-0.177*	-0.018	-0.002	0.027*	-0.18*	-0.227*	0.016	-0.021*	0.097*
		(0.00)	(0.00)	(0.27)	(0.93)	(0.00)	(0.00)	(0.00)	(0.71)	(0.01)	(0.03)

Notes: This table gives the percentage change in the average tick loss when lagged EXNFCI or RTNFCI is added to a baseline QAR(1) model on RGDP growth (e.g., -0.101 denotes roughly a 10% lower average tick loss in the unrestricted model than in the restricted model). P-values associated with the two-sided OOS-t tests are provided in parenthesis. Asterisks indicate statistical significance at the 5% level. Results are reported for the horizons $h = 1$ and $h = 4$, the quantiles $\alpha \in \{0.05, 0.1, 0.5, 0.9, 0.95\}$, and initial window sizes of $R \in \{30, 60, 90\}$. All models are estimated with a recursive scheme.

Table 9: CM Test of Equal Tick Loss, Rolling

		h = 1					h = 4				
Model	$R \setminus \alpha$	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
EXNFCI	30	-0.271*	-0.434*	0.007	0.044*	0.288*	-0.107*	-0.192*	0.016*	0.889*	1.105*
		(0.00)	(0.00)	(0.27)	(0.00)	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.00)
	60	-0.046*	-0.332*	-0.002	-0.107*	-0.072*	-0.152*	-0.234*	-0.026	-0.07*	-0.143
		(0.00)	(0.00)	(0.49)	(0.00)	(0.00)	(0.00)	(0.00)	(0.13)	(0.00)	(0.24)
	90	-0.231*	-0.161*	-0.055*	0.013	-0.015	-0.226*	-0.267*	-0.02	-0.093*	0.016
		(0.00)	(0.00)	(0.00)	(0.43)	(0.13)	(0.00)	(0.00)	(0.17)	(0.00)	(0.91)
RTNFCI	30	-0.572*	-0.445*	0.014	-0.025	0.228*	-0.07*	-0.117*	0.046	0.8*	1.015*
		(0.00)	(0.00)	(0.17)	(0.12)	(0.00)	(0.01)	(0.00)	(0.06)	(0.00)	(0.01)
	60	-0.038*	-0.323*	0.01	-0.077	-0.051*	-0.139*	-0.197*	-0.037	-0.004	-0.077
		(0.00)	(0.00)	(0.57)	(0.11)	(0.00)	(0.00)	(0.00)	(0.39)	(0.89)	(0.63)
	90	-0.195*	-0.211*	-0.031	0.015	-0.015*	-0.162*	-0.195*	-0.014	-0.08*	0.02
		(0.00)	(0.00)	(0.06)	(0.37)	(0.00)	(0.00)	(0.00)	(0.64)	(0.00)	(0.88)

Notes: See the notes to Table 8. All models are estimated with a rolling scheme.

Appendix

A NFCI Series Availability

Financial Indicator	Mnemonic	Start Date	Frequency	RTNFCI
10-yr Constant Maturity Treasury yield	D10	08-Jan-1971	W	
10-yr Interest Rate Swap/Treasury yield spread	SWAP10	03-Apr-1987	W	
10-yr/2-yr Treasury yield spread	SPR210	20-Aug-1971	W	
1-mo. Asset-backed/Financial commercial paper spread	ABCP	05-Jan-2001	W	
1-mo. BofAML Option Volatility Estimate Index	VOL1	08-Apr-1988	W	
1-mo. Nonfinancial commercial paper A2P2/AA credit spread	A2P2	10-Jan-1997	W	
1-yr/1-mo. LIBOR spread	TERM	10-Jan-1986	W	
20-yr Treasury/State & Local Government 20-yr GO bond spread	MBOND	08-Jan-1971	W	
2-yr Interest Rate Swap/Treasury yield spread	SWAP2	03-Apr-1987	W	
2-yr/3-mo. Treasury yield spread	SPR23M	08-Jan-1971	W	
30-yr Conforming Mortgage/10-yr Treasury yield spread	MINC	02-Apr-1971	W	
30-yr Jumbo/Conforming fixed rate mortgage spread	JINC	12-Jun-1998	W	
3-mo. BofAML Swaption Volatility Estimate Index	VOL3	06-Dec-1996	W	
3-mo. Eurodollar spread (LIBID-Treasury)	LIBID	08-Jan-1971	W	
3-mo. Eurodollar, 10-yr/3-mo. swap, 2-yr and 10-yr Treasury Open Interest	OINT	23-Jun-1995	W	
3-mo. Financial commercial paper/Treasury bill spread	CBILL	08-Jan-1971	W	
3-mo. Overnight Indexed Swap (OIS)/Treasury yield spread	SWAP3M	19-Sep-2003	W	
3-mo. TED spread (LIBOR-Treasury)	TED	06-Jun-1980	W	
3-mo./1-wk AA Financial commercial paper spread	CTERM	10-Jan-1997	W	
3-mo./1-wk Treasury Repo spread	RTERM	24-May-1991	W	
ABA Value of Delinquent Bank Card Credit Loans/Total Loans	DBC	26-Feb-1999	M	
ABA Value of Delinquent Consumer Loans/Total Loans	DCLOSE	26-Feb-1999	M	
ABA Value of Delinquent Home Equity Loans/Total Loans	DHE	26-Feb-1999	M	
ABA Value of Delinquent Noncard Revolving Credit Loans/Total Loans	DOTH	26-Feb-1999	M	X
Advanced Foreign Economies Trade-weighted US Dollar Value Index	USD	12-Jan-1973	W	
Agency MBS Repo Delivery Failures Rate	FAILSMBS	07-Oct-1994	W	

Agency Repo Delivery Failures Rate	FAILSA	07-Oct-1994	W	
BofAML 3-5 yr AAA CMBS OAS spread	CMBS	02-Jan-1998	W	
BofAML High Yield/Moody's Baa corporate bond yield spread	HY	07-Nov-1986	W	
BofAML Home Equity ABS/MBS yield spread	ABSSPREAD	05-Jul-1991	W	
Bond Market Association Municipal Swap/20-yr Treasury yield spread	MSWAP	07-Jul-1989	W	
Broker-dealer Debit Balances in Margin Accounts	BDG	29-Jan-1971	M	
CBOE Market Volatility Index VIX	VIX	05-Jan-1990	W	
CMBS Issuance (Relative to 12-mo. MA)	CMBSI	28-Dec-1990	M	X
CME E-mini S&P Futures Market Depth	EQUITYLIQ	04-Jan-2008	W	X
CME Eurodollar/CBOT T-Note Futures Market Depth	RATELIQ	01-Feb-2008	W	X
COMEX Gold/NYMEX wTI Futures Market Depth	COMMODLIQ	04-Jan-2008	W	X
Commercial Bank 24-mo. Personal Loan/2-yr Treasury yield spread	CBPER	05-May-1972	Q	
Commercial Bank 48-mo. New Car Loan/2-yr Treasury yield spread	CBCAR	05-May-1972	Q	
Commercial Bank C&I Loans/Total Assets	CITA	02-Mar-1973	M	
Commercial Bank Consumer Loans/Total Assets	CONTA	02-Mar-1973	M	
Commercial Bank Noncurrent/Total Loans	NCL	28-Jun-1985	Q	X
Commercial Bank Real Estate Loans/Total Assets	RTA	02-Mar-1973	M	
Commercial Bank Securities in Bank Credit/Total Assets	STA	02-Mar-1973	M	
Commercial Bank Total Unused C&I Loan Commitments/Total Assets	DCOMM	29-Jun-1990	Q	X
Commercial Paper Outstanding	CG	10-Nov-1995	W	
Consumer Credit Outstanding	CCG	29-Jan-1971	M	
CoreLogic National House Price Index	LPH	02-Apr-1976	M	
Corporate Securities Repo Delivery Failures Rate	FAILSC	05-Oct-2001	W	
Counterparty Risk Index (formerly maintained by Credit Derivatives Research)	CPR	13-Sep-2002	W	X
FDIC Volatile Bank Liabilities	GVL	01-Jul-1994	Q	X
Fed funds and Reverse Repurchase Agreements/Total Assets of Commercial Banks	ITA	30-Mar-1973	M	
Fed Funds/Overnight Agency Repo rate spread	REPOA	24-May-1991	W	
Fed Funds/Overnight MBS Repo rate spread	REPOMORT	24-May-1991	W	
Fed Funds/Overnight Treasury Repo rate spread	REPO	24-May-1991	W	
Federal, state, and local debt outstanding/GDP	STLOC	02-Apr-1971	Q	
Finance Company Owned & Managed Receivables	FG	29-Jan-1971	M	
FRB Commercial Property Price Index	CPH	02-Apr-1971	Q	
FRB Senior Loan Officer Survey: Increasing spreads on Large C&I Loans	SPCILARGE	13-Jul-1990	Q	
FRB Senior Loan Officer Survey: Increasing spreads on Small C&I Loans	SPCISMALL	13-Jul-1990	Q	
FRB Senior Loan Officer Survey: Tightening Standards on CRE Loans	CRE	12-Oct-1990	Q	

FRB Senior Loan Officer Survey: Tightening Standards on Large C&I Loans	CILARGE	13-Jul-1990	Q	
FRB Senior Loan Officer Survey: Tightening Standards on RRE Loans	RRE	12-Oct-1990	Q	
FRB Senior Loan Officer Survey: Tightening Standards on Small C&I Loans	CISMALL	13-Jul-1990	Q	
FRB Senior Loan Officer Survey: willingness to Lend to Consumers	CWILL	15-Jan-1971	Q	
Household debt outstanding/PCE Durables and Residential Investment	HH	02-Apr-1971	Q	
ICE BofAML ABS/5-yr Treasury yield spread	CTABS	01-Feb-1991	M	
ICE BofAML Financial/Corporate Credit bond spread	CTF	31-Jan-1997	M	
ICE BofAML Mortgage Master MBS/10-year Treasury yield spread	CTMBS	27-Jan-1989	M	
Markit High Yield (HY) 5-yr Senior CDS Index	LHY	07-Jan-2005	W	X
Markit Investment Grade (IG) 5-yr Senior CDS Index	LIG	01-Oct-2004	W	X
MBA Serious Delinquencies	MDQ	30-Jun-1972	Q	
Money Stock: MZM	MG	01-Mar-1974	M	
Moody's Baa corporate bond/10-yr Treasury yield spread	BAA	03-Jan-1986	W	
NACM Survey of Credit Managers: Credit Manager's Index	NACMM	15-Feb-2002	M	
Net Notional Value of Credit Derivatives5	DNET	07-Nov-2008	W	X
New State & Local Government Debt Issues (Relative to 12-mo.h MA)	MBONDGR	27-Feb-2004	M	
New US Corporate Debt Issuance (Relative to 12-mo. MA)	BONDGR	01-Jan-1988	M	
New US Corporate Equity Issuance (Relative to 12-mo. MA)	STKGR	01-Jan-1988	M	
NFIB Survey: Credit Harder to Get	SMALL	02-Nov-1973	M	
Nonfinancial business debt outstanding/GDP6	NFC	02-Apr-1971	Q	
Nonmortgage ABS Issuance (Relative to 12-mo. MA)	ABSI	29-Dec-2000	M	X
On-the-run vs. Off-the-run 10-yr Treasury liquidity premium	MLIQ10	04-Jan-1985	W	
Repo Market Volume (Repurchases+Reverse Repurchases of primary dealers)	REPOGR	07-Oct-1994	W	
S&P 500 Financials/S&P 500 Price Index (Relative to 2-yr MA)	FINS	06-Sep-1991	W	
S&P 500, NASDAQ, and NYSE Market Capitalization/GDP	MCAP	28-Jun-1985	Q	
S&P 500, S&P 500 mini, NASDAQ 100, NASDAQ mini Open Interest	OEQ	24-Sep-1999	W	
S&P US Bankcard Credit Card: 3-mo. Delinquency Rate	CCDQ	28-Feb-1992	M	
S&P US Bankcard Credit Card: Excess Rate Spread	CCINC	31-Jan-1992	M	

S&P US Bankcard Credit Card: Receivables Outstanding	CRG	28-Feb-1992	M	
Total Agency and GSE Assets/GDP	GSE	30-Dec-1983	Q	
Total Assets of ABS issuers/GDP	TABS	30-Dec-1983	Q	
Total Assets of Broker-dealers/GDP	SBD	02-Apr-1971	Q	
Total Assets of Finance Companies/GDP	FC	02-Apr-1971	Q	
Total Assets of Funding Corporations/GDP	FCORP	02-Apr-1971	Q	
Total Assets of Insurance Companies/GDP	INS	02-Apr-1971	Q	
Total Assets of Pension Funds/GDP	PENS	02-Apr-1971	Q	
Total MBS Issuance (Relative to 12-mo. MA)	MBSI	29-Dec-2000	M	X
Total Money Market Mutual Fund Assets/Total Long-term Fund Assets	MMF	28-Dec-1984	M	
Total REIT Assets/GDP	REIT	02-Apr-1971	Q	
Treasury Repo Delivery Fails Rate	FAILS	07-Oct-1994	W	
UM Household Survey: Auto Credit Conditions Good/Bad spread	CARSPREAD	24-Feb-1978	M	
UM Household Survey: Durable Goods Credit Conditions Good/Bad spread	DURSPREAD	27-Jan-1978	M	
UM Household Survey: Mortgage Credit Conditions Good/Bad spread	HOUSSPREAD	24-Feb-1978	M	
wilshire 5000 Stock Price Index	W500	29-Jan-1971	W	
CBOE Crude Oil Volatility Index, OVX	SPOVX	18-May-2007	W	

Notes: This table lists each financial indicator in the NFCI, as well as its mnemonic, start date, frequency, and whether it is excluded from the RTNFCI (denoted with an X). There are 106 listed series due to the removal and inclusion of series MBOND and SPOVX after the initial release of the NFCI, respectively. The original series with mnemonics DBC, DCLOSE, and DHE were replaced with the quarterly FRED series DRCCLT100S, DRCLACBS, and DRSREACBS in the RTNFCI. For more information on these indicators, see Brave and Butters (2011).

B Expected Tick Loss and F

Here we derive the population-level expected tick loss $E(L(\hat{u}^{(\alpha)}_{2,t+1}(t')))|_{\hat{\beta}_2=\beta_2^*}$ with an eye towards deriving F explicitly. The expected tick loss proves useful when we discuss the test's power in Section 4. For this reason, we allow $\beta_{2,2}$ to be non-zero (i.e., equation (3) is augmented by $\delta_2 z_t$) and hence z has predictive content for the α quantile given y.

Recall that the forecast error is given by

$$\begin{aligned}
\hat{u}^{(\alpha)}_{2,t+1}(t') &= y_{t+1}(t+1) - x'_{2,t}(t)\hat{\beta}_2 \\
&= (y_{t+1} - x'_{2,t}\beta^*_2) - (y_{t+1} - y_{t+1}(t+1)) + (x'_{2,t}\beta^*_2 - x'_{2,t}(t)\hat{\beta}_2) \\
&= u^{(\alpha)}_{2,t+1} - (y_{t+1} - y_{t+1}(t+1)) + (x'_{2,t}\beta^*_2 - x'_{2,t}(t)\hat{\beta}_2) \\
&= \varepsilon_{t+1} - \sigma_\varepsilon\Phi^{-1}(\alpha) - (y_{t+1} - y_{t+1}(t+1)) + (x'_{2,t}\beta^*_2 - x'_{2,t}(t)\hat{\beta}_2) \\
&= \varepsilon_{t+1} - \sigma_\varepsilon\Phi^{-1}(\alpha) - c + (x'_{2,t}\beta^*_2 - x'_{2,t}(t)\hat{\beta}_2).
\end{aligned}$$

Since $\varepsilon_{t+1} \sim N(0, \sigma^2_\varepsilon)$ and independent of $x_{2,t}(t)$, we know $u^{(\alpha)}_{2,t+1} \sim N(\mu^{(\alpha)}_t, \sigma^2_\varepsilon | x'_{2,t}(t))$, where $\mu^{(\alpha)}_t = -\sigma_\varepsilon\Phi^{-1}(\alpha) - c + (x'_{2,t}\beta^*_2 - x'_{2,t}(t)\hat{\beta}_2)$. The conditional normality is important because it allows us to also delineate the conditional expected tick loss $E(L(\hat{u}^{(\alpha)}_{2,t+1}(t'))|x_{2,t}(t))$, which proves useful when deriving F. The conditionally expected tick loss is

$$\begin{aligned}
E(L(\hat{u}^{(\alpha)}_{2,t+1}(t'))|x_{2,t}(t)) &= E((\alpha - 1(\hat{u}^{(\alpha)}_{2,t+1}(t') < 0))\hat{u}^{(\alpha)}_{2,t+1}(t')|x_{2,t}(t)) \\
&= \alpha E(\hat{u}^{(\alpha)}_{2,t+1}(t')|x_{2,t}(t)) - E(\hat{u}^{(\alpha)}_{2,t+1}(t')1(\hat{u}^{(\alpha)}_{2,t+1}(t') < 0)|x_{2,t}(t)).
\end{aligned}$$

The first term is simply $\alpha\mu^{(\alpha)}_t$. The second term is the expectation of a rectified Gaussian variable, which can be rewritten as $E(\hat{u}^{(\alpha)}_{2,t+1}(t')|\hat{u}^{(\alpha)}_{2,t+1}(t') < 0, x_{2,t}(t))P(\hat{u}^{(\alpha)}_{2,t+1}(t') < 0|x_{2,t}(t))$. Notice that $(\hat{u}^{(\alpha)}_{2,t+1}(t')|\hat{u}^{(\alpha)}_{2,t+1}(t') < 0, x_{2,t}(t))$ follows a conditionally normal distribution with the upper tail truncated and thus has expectation $\mu^{(\alpha)}_t - \sigma_\varepsilon \frac{\phi(-\frac{\mu^{(\alpha)}_t}{\sigma_\varepsilon}|x_{2,t}(t))}{\Phi(-\frac{\mu^{(\alpha)}_t}{\sigma_\varepsilon}|x_{2,t}(t))}$. $P(\hat{u}^{(\alpha)}_{2,t+1}(t') < 0|x_{2,t}(t)) = \Phi(-\frac{\mu^{(\alpha)}_t}{\sigma_\varepsilon}|x_{2,t}(t))$, and thus $E(\hat{u}^{(\alpha)}_{2,t+1}(t')1(\hat{u}^{(\alpha)}_{2,t+1}(t') < 0)|x_{2,t}(t)) = \mu^{(\alpha)}_t\Phi(-\frac{\mu^{(\alpha)}_t}{\sigma_\varepsilon}|x_{2,t}(t)) - \sigma_\varepsilon\phi(-\frac{\mu^{(\alpha)}_t}{\sigma_\varepsilon}|x_{2,t}(t))$. Putting all terms together we obtain

$$E(L(\hat{u}^{(\alpha)}_{2,t+1}(t'))|x_{2,t}(t)) = \mu^{(\alpha)}_t(\alpha - \Phi(-\frac{\mu^{(\alpha)}_t}{\sigma_\varepsilon}|x_{2,t}(t))) + \sigma_\varepsilon\phi(-\frac{\mu^{(\alpha)}_t}{\sigma_\varepsilon}|x_{2,t}(t)).$$

The unconditional expected loss, evaluated at the population parameter $\hat{\beta}_2 = \beta^*_2$, then simplifies to

$$EL(\hat{u}^{(\alpha)}_{2,t+1}(t'))|_{\hat{\beta}_2=\beta^*_2} = \mu^{(\alpha)}(\alpha - \Phi(-\frac{\mu^{(\alpha)}}{\sigma_\varepsilon})) + \sigma_\varepsilon\phi(-\frac{\mu^{(\alpha)}}{\sigma_\varepsilon}), \tag{B.1}$$

where $\mu^{(\alpha)} = -\sigma_\varepsilon\Phi^{-1}(\alpha) + c(\delta_1 - 1)$ and hence the expected loss depends nonlinearly on the magnitude of the revision c, an issue that can lead to power of the test arising in unexpected directions — a point we reemphasize in the Monte Carlo section.

To derive F_2, it's useful to note that the conditional expected loss is continuously differentiable in β_2. Hence, we interchange the differential and unconditional expectation operators to obtain

$$\begin{aligned}
F_2 &= E[\partial E(L(\hat{u}^{(\alpha)}_{2,t+1}(t'))|x_{2,t}(t))/\partial\hat{\beta}_2]|_{\hat{\beta}_2=\beta_2^*} \\
&= \alpha\partial\mu_t^{(\alpha)}/\partial\hat{\beta}_2 - \Phi(-\frac{\mu_t^{(\alpha)}}{\sigma_\varepsilon}|x_{2,t}(t))\partial\mu_t^{(\alpha)}/\partial\hat{\beta}_2 \\
&\quad -\mu_t^{(\alpha)}\partial\Phi(-\frac{\mu_t^{(\alpha)}}{\sigma_\varepsilon}|x_{2,t}(t)))/\partial\hat{\beta}_2 + \sigma_\varepsilon\partial\phi(-\frac{\mu_t^{(\alpha)}}{\sigma_\varepsilon}|x_{2,t}(t))/\partial\hat{\beta}_2]|_{\hat{\beta}_2=\beta_2^*} \\
&= E[-\alpha x_{2,t}(t) + \Phi(-\frac{\mu_t^{(\alpha)}}{\sigma_\varepsilon}|x_{2,t}(t)))x_{2,t}(t) \\
&\quad -\frac{\mu_t^{(\alpha)}}{\sigma_\varepsilon}\phi(-\frac{\mu_t^{(\alpha)}}{\sigma_\varepsilon}|x_{2,t}(t))x'_{2,t}(t) + \phi'(-\frac{\mu_t^{(\alpha)}}{\sigma_\varepsilon}|x_{2,t}(t))x'_{2,t}(t)]|_{\hat{\beta}_2=\beta_2^*} \\
&= -E(\alpha - 1(\hat{u}_{2,t+1}(t') < 0))x'_{2,t}(t),
\end{aligned}$$

where $\phi'(.)$ denotes the derivative of the standard normal PDF. In the fourth equality, that the latter two right-hand-side terms cancel, is based on straightforward algebra associated with a standard normal density. Given F_2, $F = (F_1, -F_2)$ can be obtained by recalling that $x_{1,t}(t) = J'x_{2,t}(t)$ and under the null $\hat{u}_{2,t+1}(t') = \hat{u}_{1,t+1}(t')$.

C Monte Carlo Simulation Tables

Table C1: Two-Sided Size Tests, DGP1

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.07	0.08	0.08	0.07	0.07	0.07	0.24	0.19	0.16	0.10	0.09	0.08
	500	0.07	0.07	0.07	0.06	0.06	0.06	0.12	0.10	0.09	0.07	0.06	0.06
	1000	0.07	0.07	0.06	0.06	0.06	0.05	0.11	0.09	0.09	0.06	0.06	0.05
0.1	100	0.08	0.08	0.08	0.07	0.07	0.07	0.19	0.16	0.14	0.09	0.08	0.07
	500	0.07	0.07	0.07	0.06	0.06	0.06	0.11	0.10	0.09	0.07	0.07	0.06
	1000	0.06	0.07	0.07	0.05	0.06	0.06	0.09	0.08	0.08	0.06	0.06	0.06
0.5	100	0.07	0.07	0.07	0.06	0.06	0.06	0.14	0.11	0.09	0.07	0.06	0.05
	500	0.07	0.06	0.06	0.06	0.05	0.05	0.10	0.09	0.07	0.06	0.06	0.06
	1000	0.07	0.06	0.06	0.05	0.05	0.05	0.08	0.07	0.07	0.06	0.06	0.06

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.28	0.25	0.21	0.17	0.14	0.10	0.08	0.14	0.23	0.39	0.55	0.69
	500	0.17	0.13	0.10	0.08	0.07	0.06	0.27	0.42	0.57	0.84	0.89	0.91
	1000	0.13	0.10	0.08	0.08	0.07	0.06	0.46	0.61	0.73	0.91	0.93	0.94
0.1	100	0.24	0.22	0.18	0.14	0.11	0.09	0.10	0.19	0.31	0.52	0.68	0.78
	500	0.15	0.12	0.09	0.08	0.07	0.06	0.39	0.54	0.67	0.89	0.92	0.92
	1000	0.10	0.09	0.08	0.07	0.06	0.06	0.56	0.71	0.81	0.94	0.94	0.95
0.5	100	0.18	0.15	0.11	0.09	0.07	0.06	0.21	0.31	0.44	0.72	0.83	0.87
	500	0.09	0.08	0.07	0.06	0.06	0.05	0.57	0.73	0.81	0.93	0.94	0.95
	1000	0.07	0.07	0.06	0.06	0.06	0.05	0.75	0.84	0.89	0.96	0.96	0.97

Notes: See the notes to Table 3. All tests are compared against two-sided standard normal critical values.

Table C2: Two-Sided Power Tests, DGP1

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.10	0.16	0.29	0.19	0.32	0.55	0.50	0.54	0.68	0.25	0.28	0.38
	500	0.40	0.70	0.95	0.79	0.97	1.00	0.91	0.98	1.00	0.89	0.97	1.00
	1000	0.72	0.96	1.00	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
0.1	100	0.13	0.23	0.45	0.27	0.47	0.75	0.61	0.71	0.85	0.41	0.51	0.69
	500	0.58	0.89	1.00	0.94	1.00	1.00	0.97	1.00	1.00	0.99	1.00	1.00
	1000	0.89	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99
0.5	100	0.29	0.50	0.81	0.57	0.82	0.98	0.84	0.94	0.99	0.83	0.94	0.99
	500	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.61	0.69	0.82	0.67	0.76	0.85	0.12	0.23	0.45	0.42	0.71	0.93
	500	0.97	1.00	1.00	1.00	1.00	1.00	0.63	0.89	0.99	0.99	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.99	1.00	1.00	1.00	1.00
0.1	100	0.73	0.83	0.94	0.81	0.89	0.95	0.19	0.38	0.68	0.63	0.89	0.99
	500	0.99	1.00	1.00	1.00	1.00	1.00	0.82	0.98	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	0.99	0.98	1.00	1.00	1.00	1.00	0.99
0.5	100	0.89	0.97	1.00	0.96	0.98	1.00	0.50	0.77	0.97	0.92	1.00	1.00
	500	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	0.99	1.00	0.99	0.99	1.00	1.00	0.99	1.00	1.00	0.99

Notes: See the notes to Table 4. All tests are compared against two-sided standard normal critical values.

Table C3: Two-Sided Size Tests, DGP2

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.06	0.07	0.07	0.06	0.06	0.06	0.52	0.44	0.37	0.20	0.15	0.12
	500	0.07	0.07	0.07	0.06	0.06	0.06	0.25	0.19	0.15	0.10	0.09	0.08
	1000	0.07	0.07	0.06	0.06	0.06	0.06	0.17	0.14	0.12	0.09	0.08	0.07
0.1	100	0.07	0.06	0.07	0.06	0.06	0.06	0.45	0.37	0.28	0.16	0.13	0.11
	500	0.06	0.07	0.07	0.07	0.06	0.06	0.22	0.17	0.14	0.11	0.09	0.08
	1000	0.07	0.07	0.07	0.07	0.07	0.06	0.16	0.13	0.11	0.10	0.09	0.08
0.5	100	0.06	0.06	0.06	0.07	0.07	0.07	0.37	0.28	0.20	0.17	0.13	0.11
	500	0.06	0.07	0.07	0.06	0.06	0.06	0.17	0.13	0.11	0.10	0.09	0.07
	1000	0.07	0.07	0.06	0.06	0.06	0.06	0.14	0.11	0.09	0.08	0.08	0.07

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.19	0.19	0.19	0.17	0.15	0.13	0.08	0.09	0.12	0.20	0.28	0.37
	500	0.18	0.19	0.19	0.15	0.13	0.11	0.14	0.21	0.28	0.52	0.62	0.73
	1000	0.19	0.19	0.17	0.14	0.13	0.12	0.21	0.30	0.40	0.67	0.76	0.82
0.1	100	0.16	0.16	0.17	0.15	0.14	0.12	0.08	0.10	0.15	0.25	0.36	0.46
	500	0.17	0.17	0.16	0.13	0.11	0.10	0.19	0.27	0.36	0.62	0.72	0.80
	1000	0.18	0.15	0.13	0.11	0.10	0.09	0.29	0.39	0.50	0.76	0.83	0.87
0.5	100	0.13	0.15	0.14	0.15	0.13	0.09	0.16	0.19	0.21	0.46	0.54	0.63
	500	0.13	0.13	0.10	0.08	0.06	0.05	0.32	0.40	0.48	0.75	0.82	0.86
	1000	0.12	0.10	0.07	0.06	0.05	0.04	0.41	0.52	0.62	0.85	0.90	0.91

Notes: See the notes to Table 5. All tests are compared against two-sided standard normal critical values.

Table C4: Size Tests, DGP1, Rolling

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.01	0.00	0.00
	500	0.01	0.01	0.01	0.02	0.01	0.01	0.03	0.02	0.01	0.02	0.02	0.01
	1000	0.02	0.02	0.01	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02
0.1	100	0.01	0.01	0.00	0.01	0.01	0.00	0.03	0.02	0.01	0.01	0.01	0.00
	500	0.02	0.02	0.01	0.03	0.02	0.02	0.04	0.03	0.02	0.03	0.02	0.02
	1000	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.03	0.02	0.03	0.03	0.03
0.5	100	0.03	0.03	0.02	0.03	0.03	0.02	0.05	0.04	0.03	0.04	0.03	0.02
	500	0.03	0.03	0.03	0.03	0.04	0.04	0.05	0.04	0.03	0.04	0.04	0.03
	1000	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.07	0.04	0.02	0.06	0.04	0.02	0.02	0.01	0.00	0.06	0.07	0.05
	500	0.07	0.06	0.03	0.06	0.05	0.03	0.05	0.07	0.06	0.33	0.33	0.28
	1000	0.07	0.05	0.04	0.05	0.05	0.04	0.13	0.15	0.14	0.40	0.39	0.35
0.1	100	0.07	0.06	0.04	0.07	0.05	0.03	0.02	0.01	0.01	0.13	0.15	0.13
	500	0.08	0.06	0.04	0.06	0.05	0.04	0.10	0.12	0.12	0.39	0.37	0.34
	1000	0.07	0.06	0.04	0.06	0.06	0.04	0.20	0.23	0.21	0.43	0.41	0.40
0.5	100	0.11	0.10	0.06	0.08	0.06	0.04	0.07	0.07	0.07	0.30	0.32	0.28
	500	0.08	0.07	0.05	0.06	0.05	0.04	0.22	0.25	0.25	0.42	0.41	0.39
	1000	0.07	0.06	0.05	0.05	0.05	0.05	0.31	0.34	0.33	0.44	0.44	0.44

Notes: See the notes to Table 3. Models are estimated with a rolling scheme.

Table C5: Power Tests, DGP1, Rolling

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.13	0.19	0.31	0.23	0.37	0.68	0.42	0.48	0.62	0.22	0.23	0.34
	500	0.52	0.86	1.00	0.88	1.00	1.00	0.91	0.99	1.00	0.92	0.99	1.00
	1000	0.84	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
0.1	100	0.18	0.31	0.56	0.34	0.60	0.92	0.57	0.70	0.86	0.42	0.53	0.78
	500	0.71	0.98	1.00	0.98	1.00	1.00	0.98	1.00	1.00	0.99	1.00	1.00
	1000	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.5	100	0.38	0.63	0.93	0.68	0.92	1.00	0.85	0.95	1.00	0.87	0.97	1.00
	500	0.98	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99
	1000	1.00	0.99	0.99	1.00	1.00	0.99	1.00	0.99	0.99	1.00	1.00	0.99

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.47	0.57	0.71	0.58	0.68	0.79	0.16	0.21	0.34	0.42	0.65	0.89
	500	0.96	1.00	1.00	1.00	1.00	1.00	0.71	0.95	1.00	0.99	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00
0.1	100	0.65	0.79	0.92	0.78	0.86	0.95	0.26	0.42	0.68	0.67	0.90	0.99
	500	0.99	1.00	1.00	1.00	1.00	1.00	0.88	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
0.5	100	0.89	0.97	1.00	0.96	0.99	1.00	0.60	0.85	0.99	0.95	1.00	1.00
	500	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99
	1000	1.00	0.99	0.99	1.00	1.00	0.99	1.00	0.99	0.99	1.00	1.00	0.99

Notes: See the notes to Table 4. Models are estimated with a rolling scheme.

Table C6: Size Tests, DGP2, Rolling

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.02	0.02	0.01	0.04	0.04	0.03	0.20	0.15	0.10	0.12	0.10	0.09
	500	0.03	0.02	0.02	0.03	0.03	0.03	0.11	0.08	0.06	0.06	0.05	0.05
	1000	0.03	0.03	0.02	0.04	0.04	0.04	0.08	0.06	0.05	0.05	0.05	0.05
0.1	100	0.03	0.03	0.02	0.05	0.06	0.06	0.19	0.15	0.11	0.12	0.10	0.10
	500	0.03	0.03	0.03	0.05	0.05	0.05	0.10	0.07	0.06	0.07	0.06	0.06
	1000	0.04	0.04	0.03	0.04	0.05	0.05	0.08	0.07	0.06	0.06	0.06	0.06
0.5	100	0.04	0.04	0.04	0.08	0.10	0.12	0.21	0.18	0.15	0.16	0.15	0.18
	500	0.05	0.06	0.05	0.07	0.08	0.09	0.13	0.10	0.09	0.09	0.10	0.10
	1000	0.05	0.05	0.05	0.06	0.06	0.07	0.09	0.08	0.07	0.08	0.07	0.07

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.08	0.07	0.06	0.12	0.11	0.11	0.03	0.02	0.01	0.10	0.13	0.13
	500	0.09	0.08	0.08	0.12	0.11	0.09	0.05	0.05	0.04	0.28	0.31	0.31
	1000	0.10	0.10	0.08	0.11	0.11	0.09	0.08	0.09	0.08	0.35	0.38	0.37
0.1	100	0.07	0.07	0.07	0.13	0.13	0.14	0.03	0.03	0.02	0.15	0.20	0.23
	500	0.08	0.09	0.08	0.11	0.10	0.10	0.07	0.08	0.07	0.34	0.38	0.40
	1000	0.09	0.10	0.09	0.09	0.09	0.09	0.12	0.13	0.14	0.42	0.44	0.45
0.5	100	0.09	0.10	0.11	0.15	0.17	0.19	0.09	0.07	0.06	0.31	0.38	0.45
	500	0.11	0.12	0.10	0.10	0.09	0.08	0.17	0.19	0.19	0.47	0.50	0.52
	1000	0.10	0.10	0.08	0.07	0.06	0.06	0.21	0.24	0.26	0.49	0.50	0.50

Notes: See the notes to Table 5. Models are estimated with a rolling scheme.

Table C7: Power Tests, DGP2, Rolling

		Θ-based						Ω-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.03	0.03	0.02	0.01	0.00	0.00	0.27	0.23	0.17	0.07	0.03	0.01
	500	0.05	0.04	0.04	0.00	0.00	0.00	0.37	0.32	0.28	0.02	0.00	0.00
	1000	0.05	0.05	0.07	0.00	0.00	0.00	0.40	0.37	0.35	0.01	0.00	0.00
0.1	100	0.04	0.04	0.03	0.01	0.00	0.00	0.31	0.28	0.24	0.08	0.03	0.01
	500	0.06	0.08	0.09	0.00	0.00	0.00	0.43	0.42	0.41	0.02	0.00	0.00
	1000	0.07	0.09	0.14	0.00	0.00	0.00	0.47	0.47	0.51	0.00	0.00	0.00
0.5	100	0.07	0.09	0.12	0.01	0.01	0.00	0.43	0.43	0.48	0.12	0.06	0.01
	500	0.13	0.20	0.39	0.00	0.00	0.00	0.59	0.68	0.82	0.04	0.01	0.00
	1000	0.19	0.33	0.64	0.00	0.00	0.00	0.69	0.80	0.93	0.01	0.00	0.00

		$\hat{\Omega}$-based						$\hat{S}_{dd}$-based					
		$c = 0.5\sigma_u$			$c = \sigma_u$			$c = 0.5\sigma_u$			$c = \sigma_u$		
		P/R			P/R			P/R			P/R		
α	R	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
0.05	100	0.13	0.13	0.11	0.09	0.06	0.02	0.06	0.04	0.01	0.05	0.02	0.01
	500	0.31	0.29	0.25	0.05	0.01	0.00	0.09	0.06	0.03	0.02	0.00	0.00
	1000	0.37	0.33	0.30	0.02	0.00	0.00	0.10	0.08	0.06	0.01	0.00	0.00
0.1	100	0.17	0.17	0.17	0.10	0.07	0.02	0.08	0.06	0.03	0.05	0.03	0.01
	500	0.37	0.37	0.34	0.05	0.01	0.00	0.12	0.11	0.07	0.02	0.00	0.00
	1000	0.41	0.40	0.42	0.01	0.00	0.00	0.14	0.13	0.12	0.00	0.00	0.00
0.5	100	0.29	0.32	0.38	0.17	0.10	0.03	0.15	0.14	0.12	0.11	0.06	0.02
	500	0.51	0.58	0.71	0.06	0.01	0.00	0.25	0.30	0.42	0.03	0.01	0.00
	1000	0.62	0.70	0.86	0.02	0.00	0.00	0.33	0.44	0.65	0.01	0.00	0.00

Notes: See the notes to Table 6. Models are estimated with a rolling scheme.