'Déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables[☆]

Bradley S. Paye[*]

Terry College of Business, University of Georgia, Athens, GA 30602, USA

Abstract

Aggregate stock return volatility is both persistent and countercyclical. This paper tests whether it is possible to improve volatility forecasts at monthly and quarterly horizons by conditioning on additional macroeconomic variables. I find that several variables related to macroeconomic uncertainty, time-varying expected stock returns, and credit conditions Granger cause volatility. It is more difficult to find evidence that forecasts exploiting macroeconomic variables outperform a univariate benchmark out-of-sample. The most successful approaches involve simple combinations of individual forecasts. Predictive power associated with macroeconomic variables appears to concentrate around the onset of recessions.

JEL classification: G12; C22

Keywords: Conditional volatility, Realized volatility, Granger causality, Forecast evaluation, Forecast combination

June 11, 2012

1. Introduction

What drives secular variation in stock return volatility? In a seminal paper, Schwert (1989) considers several potential explanations, including the possibility that volatility fluctuates with the level of economic activity. Although Schwert (1989) finds only limited support for links between volatility and macroeconomic activity, subsequent papers report more encouraging evidence. This large body of literature is difficult to digest, as different studies examine different forecasting variables and apply different econometric approaches.[1]

Understanding the robustness and magnitude of links between macroeconomic variables and volatility represents an important empirical question in finance. From a risk-management perspective, understanding how future aggregate stock market volatility responds to changing macroeconomic conditions is critical for stress-testing and computing value-at-risk over longer horizons. From an asset allocation perspective, quantities that forecast volatility become state variables in investors' portfolio decisions. Finally, characterizing the extent and pattern of time series variation in volatility is important for determining the appropriate stylized facts against which asset pricing models should be evaluated.

Consistent with Schwert (1989) and other existing research, I find that stock return volatility behaves countercyclically. Empirical measures of business conditions, such as the growth rate of gross domestic product (GDP), co-move closely with sign-inverted measures of stock return volatility. From a forecasting perspective, the strong contemporaneous relation between volatility and business conditions implies that lagged volatility provides an efficient indicator of the economic state. Whether it is possible to improve forecast performance by conditioning on additional macroeconomic and financial variables is unclear. To be successful, such variables must capture information beyond that already contained in lagged volatility. This paper provides a broad assessment of the ability of macroeconomic and financial variables to improve volatility forecasts at monthly and quarterly horizons.

Recent literature identifies several channels that could drive time variation in volatility. These include time-varying volatility in shocks to fundamentals (e.g., Bansal and Yaron, 2004), nonlinear relations between time-varying expected returns and the business cycle (Mele, 2007), learning effects related to investors' uncertainty about fundamentals (e.g., Veronesi, 1999), and amplification of shocks to asset markets via financial intermediation (Brunnermeier and Pedersen, 2009). This body of theoretical work motivates the

---

[1] Relevant papers include Campbell (1987), Breen, Glosten, and Jagannathan (1989), Shanken (1990), Glosten, Jagannathan, and Runkle (1993), Whitelaw (1994), Harvey (2001), Marquering and Verbeek (2004), Lettau and Ludvigson (2010), Ludvigson and Ng (2007), Engle and Rangel (2008), Engle, Ghysels, and Sohn (2008), and Campbell and Diebold (2009).

set of forecasting variables considered in the paper. These include a measure of corporate payout, several interest rate and return spreads, a measure of changes in bank leverage, measures of current and expected economic growth, a direct proxy for time-varying expected returns, volatilities for two key macroeconomic series, and two ratios for the aggregate economy: consumption to wealth and investment to capital stock.

I emphasize an out-of-sample econometric approach, although in-sample results appear for reference and comparison. This focus parallels the recent emphasis on out-of-sample inference in the literature on stock return predictability, where an active debate continues regarding the extent to which returns are predictable.[2] The paper distinguishes between two alternative notions of out-of-sample forecast improvement. The first focuses on properties of the data generating process: Do macroeconomic variables Granger cause volatility, such that volatility depends upon these variables conditional on past volatility? The second interpretation adopts a normative stance: Do volatility models that incorporate macroeconomic variables improve the accuracy of out-of-sample forecasts?

To see that these alternative interpretations are distinct, suppose that the conditional volatility of stock returns depends on some macroeconomic variable, so that this variable Granger causes volatility. Out-of-sample forecasts exploiting the variable could nevertheless under-perform forecasts based on a (misspecified) model that omits it. This is because there is a bias-variance trade-off at play. The conditional bias reduction afforded by including the macroeconomic predictor might not offset increased forecast variance related to parameter estimation.

I conduct two econometric tests comparing the out-of-sample forecasting performance of a benchmark model with a model augmented with one or more of the predictor variables. Both tests involve the out-of-sample difference in mean square prediction error (MSPE) relative to the benchmark. The first test, proposed by Giacomini and White (2006), is equivalent to the Diebold and Mariano (1995) test for equal predictive ability. The second test, proposed by Clark and West (2007), adds an adjustment term to the out-of-sample difference in MSPE that accounts for parameter estimation noise. The key difference between the two testing frameworks lies in the specification of the null hypothesis. In the Clark and West (2007) framework, the null hypothesis involves the population difference in MSPE between the two nested models. By contrast, the null hypothesis in the Giacomini and White (2006) framework relates to the forecasting method and

---

[2]Goyal and Welch (2008) find little evidence that common stock return forecasting regressions succeed out-of-sample. Campbell and Thompson (2008) find that imposing economically motivated constraints on coefficients and return forecasts delivers forecast improvements relative to the historical average. Rapach, Strauss, and Zhou (2010) find that combining individual return forecasts improves out-of-sample performance.

explicitly incorporates parameter estimation as a source of forecast error. The Clark and West (2007) test is appropriate when the underlying research question involves Granger causality, whereas the Giacomini and White (2006) test is appropriate for addressing the normatively oriented question of whether one forecast performs better than the other.

The empirical evidence from in-sample forecasting regressions is encouraging. Several variables appear to Granger cause volatility, including the commercial paper-to-Treasury spread, the default spread, a bond return spread, and the ratio of investment to capital in the aggregate economy. The null of no predictability is also rejected for a kitchen sink specification that includes the full set of predictors. Although the statistical evidence for Granger causality is compelling, the economic significance of the predictive power afforded by these variables is relatively small. These findings are robust to several alternative sample periods, with the strength of evidence for predictability being strongest between the 1950s and early 1980s.

Out-of-sample evidence regarding Granger causality largely confirms the results from in-sample predictive regressions. The Clark and West (2007) test for Granger causality implicates essentially the same variables that are significant based on in-sample regressions. The null of no Granger causality is also typically rejected for the kitchen sink model. Results are somewhat sensitive; however, to inclusion of the economically volatile 1970s in the out-of-sample evaluation period. Specifically, the out-of-sample evidence for Granger causality is weaker over the period 1982–2010 relative to the equally long period 1972–2000 that includes the 1970s.

The evidence for superior predictive ability (in the Giacomini and White sense) is mixed. Taken one at a time, the individual predictors fail to generate statistically significant improvements in forecast accuracy relative to the benchmark. The heavily parameterized kitchen sink model also fairs poorly. In some cases, this model significantly under-performs the benchmark. Not all of the evidence is negative, though. Combinations of the underlying univariate forecasts often statistically outperform the benchmark. A simple equal-weighted combination of the univariate forecasts often succeeds. These findings are consistent with Rapach, Strauss, and Zhou (2010), who consider similar combination schemes in the context of stock return forecasts. There are two caveats. First, the associated out-of-sample R^2 improvements relative to the benchmark are relatively small. Second, evidence of superior performance is weaker when the out-of-sample evaluation period does not include the 1970s.

The pattern of results highlights the distinction between the two types of out-of-sample tests applied in the paper. In many instances, the Clark and West test rejects, while the Giacomini and White test fails to

reject, or vice versa. Results for the kitchen sink model provide a useful illustration. Under the null hypothesis of no Granger causality, this model is expected to substantially under-perform the benchmark, because it includes a large number of spurious variables. Despite underwhelming out-of-sample R^2 values, the kitchen sink model often performs better than expected, and consequently the Clark and West test rejects the null. To expand upon this point, I develop a simple Monte Carlo experiment. Simulated data are calibrated to match the key empirical features of return volatility and a typical macroeconomic predictor. In one case, the simulated variable Granger causes volatility, while a second case imposes the null of no Granger causality. For empirically realistic sample sizes, the simulation results illustrate a pattern of divergence between the Giacomini and White (2006) and Clark and West (2007) tests consistent with that observed in the data.

Weak forecasting results during the Great Moderation suggest connections between the business cycle and relative forecast performance. Time series plots of cumulative forecast performance relative to the benchmark display a striking pattern for the more successful predictors: Forecast improvements appear to concentrate around the onset of recessions. Relative forecast performance, therefore, is countercyclical.[3] Unfortunately, the linkage between forecast performance and recessions is complex, as different variables exhibit different patterns of forecast performance around recessions. By exploiting information in multiple predictors, the combined forecasts generate a more consistent pattern of forecast improvements.

The main findings survive a wide array of robustness checks. These include using alternative empirical proxies for stock return volatility, using alternative dynamic specifications for the benchmark forecasts, forecasting the level as opposed to the log of stock return volatility, and employing a recursive versus rolling estimation scheme.

Stock return volatility is persistent, while stock returns are not. Notwithstanding this important difference, there are striking similarities between results established here and findings reported in the extensive literature on forecasting stock returns. Specifically, this paper shows that, from an ex post perspective, certain macroeconomic and financial variables help explain time variation in stock return volatility. At the same time, the additional predictive power afforded by these variables is small, and forecasting ability appears to concentrate during a subperiod from the 1950s through the early 1980s. Similar statements well-characterize recent findings in the equity premium literature.[4] Viewed from this perspective, the results

---

[3]This parallels findings in the return forecasting literature. Henkel, Martin, and Nardari (2011) employ a regime switching model and find that return predictability is concentrated in recessions. Rapach, Strauss, and Zhou (2010) find that out-of-sample return forecasting gains are concentrated in low growth episodes.

[4]Spiegel (2008) provides a recent synopsis of this literature.

reported here conjure a distinct sense of déjà vu.

This paper is among several recent studies that embody a resurgence of interest in connections between macroeconomic conditions and stock return volatility. Related work includes Adrian and Shin (2010), David and Veronesi (2009), Engle, Ghysels, and Sohn (2008), and Ludvigson and Ng (2007). Relative to these studies, the present paper focuses on applying recently developed econometric techniques to evaluate the extent to which macroeconomic and financial variables improve volatility forecasts out-of-sample. While the papers adopt different perspectives, important commonalities exist.

The remainder of the paper proceeds as follows. Section 2 discusses the statistical properties of stock return volatility and analyzes connections between stock return volatility and the level of economic activity. Section 3 discusses theoretical explanations for time variation in stock return volatility and describes the set of forecasting variables used in the analysis. Sections 4 and 5 present empirical results, the former from an in-sample perspective and the latter using an out-of-sample research design. Section 6 concludes.

2. Stock return volatility and the business cycle

This section discusses the measurement of stock return volatility and characterizes the relation between stock return volatility and the business cycle.

2.1. Measuring stock return volatility

The conditional variance of a portfolio return is based on ex ante expectations and is fundamentally unobservable. A regression-based approach to modeling conditional volatility relies on an ex post measurement of variance. Following Taylor (1986), French, Schwert, and Stambaugh (1987), and Schwert (1989), I sum squared daily returns to construct a proxy for the variance of excess returns on the Standard & Poor's (S&P) 500 index at both monthly and quarterly sampling frequencies:

$$RV(t) = \sum_{i=1}^{N_t} R_{i,t}^2, \tag{1}$$

where N_t denotes the number of trading days in the tth month or quarter and $R_{i,t}$ indicates the daily excess return on the S&P 500 index on the ith trading day of the tth period.

The notation $RV(t)$ emphasizes the connection between Eq. (1) and the realized variance literature that employs intraday returns to measure return variation. Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2002) show that, as the intraperiod sampling frequency increases, Eq. (1)

converges in probability to the quadratic variation of a frictionless, arbitrage-free asset price process.[5]

2.2. Statistical properties of stock return volatility

Aggregate stock return volatility is positively skewed and leptokurtotic. The leptokurtosis is partially attributable to several concentrated episodes of stock return volatility, including the October 1987 stock market crash and the dramatic plunge in equity values associated with the financial crisis of 2008. This paper focuses on forecasting volatility using linear models estimated by ordinary least squares (OLS). When regression errors are non-normal and fat-tailed, OLS might be inferior to nonlinear, robust estimation approaches. Taking the natural logarithm of realized volatility results in a series that is approximately Gaussian, as shown by Andersen, Bollerslev, Diebold, and Ebens (2001). Consequently, the subsequent empirical analysis focuses on modeling and forecasting the natural logarithm of annualized volatility, $LVOL(t) \equiv ln(\sqrt{mRV(t)})$, where m corresponds to the number of periods within a year (four for quarterly sampling and 12 for monthly sampling).[6] The upper panel of Fig. 1 presents a time series plot of quarterly $LVOL(t)$. Volatility is clearly persistent, a fact well established in the literature.

2.3. The relation between stock return volatility and the business cycle

Stock return volatility tends to be higher during recessions relative to expansions. For example, Schwert (1989) regresses volatility on a dummy variable that takes the value of one during National Bureau of Economic Research (NBER) recessions and finds volatility to be significantly higher during these periods. The middle and bottom panels of Fig. 1 provide visual evidence regarding the cyclical properties of stock return volatility. These panels provide time series plots of sign-inverted log stock return volatility alongside real GDP growth at the quarterly frequency. To facilitate comparison, both series are standardized prior to plotting. The middle panel shows the raw series, and the bottom panel shows smoothed versions of the series, constructed as six-quarter moving averages.

The affinity between stock return volatility and the level of economic activity is striking. Overall, the volatility and GDP series track one another closely. The plots confirm conventional wisdom: When business conditions are poor, stock return volatility tends to be higher. Interestingly, there appears to be time variation in the degree of affinity between volatility and business conditions. The two series co-move tightly over a period beginning in the the mid-1960s and extending through the early 1980s. During other periods, the

---

[5]An unpublished Appendix, available upon request, explores robustness to variations on Eq. (1), including alternative measures based on absolute returns instead of squared returns, measures that attempt to correct for time variation in expected returns, and measures that exploit intraday price data.

[6]The unpublished Appendix shows that most findings are robust to alternatively modeling the level of volatility.

series are more divergent. For example, real business conditions do not respond to the spike in return volatility in October 1987, and the two series appear out-of-phase during most of the 1990s.[7]

The close relation between stock return volatility and real economic conditions illustrates the challenge in identifying macroeconomic and financial variables that can improve long-horizon volatility forecasts. Because lagged volatility captures a rich set of information regarding current economic conditions, successful forecasting variables must capture additional relevant information to be helpful. The remaining sections of the paper consider a set of potentially useful forecasting variables and empirically assess the extent to which these variables improve volatility forecasts relative to a simple benchmark.

3. A set of forecasting variables motivated by theory

This section discusses potential explanations for time variation in stock return volatility and describes the set of forecasting variables used in the empirical analysis.

3.1. Potential explanations for secular variation in stock return volatility

Conceptually, the conditional variance of market returns depends upon the conditional variances of future cash flows, the conditional variances of discount rates, and conditional covariances between these two series. Under a constant discount rate, the conditional variance of the aggregate return depends only on the conditional variances of future aggregate cash flows. This illustrates one channel for time-varying, countercyclical stock return volatility: shocks to fundamentals (dividends) that display exactly these features. Bansal and Yaron (2004) provide a model in this spirit. The model assumes that the volatilities of dividend and consumption growth are countercyclical and, consequently, generates countercyclical stock market volatility.

Mele (2005, 2007) emphasizes that, separately from the fundamentals channel, time variation in discount rates could, in and of itself, generate countercyclical variation in stock return volatility. For this mechanism to operate, expected returns must be convex in a state variable that captures business conditions. Intuitively, during bad times expected returns must be relatively sensitive to fluctuations in the state variable, while during good times expected returns are less sensitive to such fluctuations. When this asymmetry is sufficiently strong, Mele (2007) shows that the price dividend ratio is an increasing, concave function of the state variable. Consequently, return volatility increases on the downside.

---

[7]Plots that employ alternative measures of US economic conditions, such as the Aruoba-Diebold-Scotti Business Conditions Index developed by Aruoba, Diebold, and Scotti (2009), are very similar.

A third explanation for time-varying stock market volatility focuses on the role of learning. In models of this flavor, agents use information contained in public signals to make inferences about unknown aspects of the economy. Timmermann (1993, 1996) considers a setting where the stock price is the sum of expected future dividends with an exogenously specified discount rate and shows that learning effects can increase stock return volatility relative to a benchmark setting with observable dividend growth. Brennan and Xia (2001) extend this finding to a stochastic dynamic general equilibrium model with rational learning. Veronesi (1999) develops a dynamic, rational expectations equilibrium model of asset prices assuming that the dividend growth rate shifts randomly between two unobservable states. In equilibrium, investors' desire to hedge against changes in their level of uncertainty makes asset prices more sensitive to bad news in good times relative to good news in bad times. Consequently, stock return volatility is countercyclical, as in the data.

Finally, the liquidity and credit crisis of 2007–2008 prompted academic interest in the capacity for financial intermediation to amplify shocks to asset markets. Brunnermeier and Pedersen (2009) study the interrelation between traders' funding liquidity and asset market liquidity. In this setting, borrower's balance sheet effects can amplify relatively small initial shocks through a *loss spiral* and a reinforcing *margin spiral*. For leveraged investors, declines in asset values erode net worth much faster than gross worth. Such investors might need to sell assets to maintain leverage, leading to further asset price drops. The margin spiral refers to the fact that margins tend to increase in the wake of large price drops. This exacerbates pressure on leveraged investors to sell off assets, leading to larger subsequent price drops, further increases in margins, and so on. Brunnermeier and Pedersen (2009) show that a vicious cycle can ensue in which multiple equilibria exist. Adrian and Shin (2010) provide supporting empirical evidence for investment banks. In addition to balance sheet effects, lenders' capital limitations, network effects, and bank runs could amplify shocks in financial markets.

3.2. Forecasting variables

The theoretical explanations for time variation in volatility rely on key unobservable concepts, such as investors' uncertainty regarding the true economic state and expected stock returns. In the empirical analysis that follows, I consider a number of observable series plausibly correlated with one or more of these theoretical drivers. The set of forecasting variables includes:

- Changes in bank leverage (*blev*): Following Adrian and Shin (2010), bank leverage is computed as the ratio of total assets to total equity using US Flow of Funds data for securities, brokers, and dealers.

- Consumption-wealth ratio (*cay*): The consumption-wealth ratio (*cay*), proposed by Lettau and Lud-vigson (2001), is the residual obtained from estimating a co-integrating relation between aggregate consumption, wealth, and labor income.

- Commercial paper-to-Treasury spread (*cp*): This variable captures the spread between the three-month commercial paper rate and the rate on three-month Treasury bills.

- Default return spread (*dfr*): The default return spread is the difference between long-term corporate bond and long-term government bond returns.

- Default spread (*dfy*): The default spread is the difference between the yield on BAA-rated corporate bonds and the yield on long term US government bonds.

- Expected return (*exret*): This variable is a regression-based estimate of the expected excess return on the S&P 500 index. The predictive regression for stock returns includes several variables common in the return forecasting literature, such as the default spread, the net payout yield, and the consumption-wealth ratio.

- Current and expected GDP growth (*gdp* and *egdp*): Current economic activity is measured using the annualized growth rate in real, seasonally adjusted GDP. Expectations of future economic growth are based on six- to 12-month GDP growth forecasts from the Livingston Survey. The Livingston Survey captures economists' real time macroeconomic forecasts at a bi-annual frequency. I follow Campbell and Diebold (2009) in constructing expected GDP growth rate using nominal GDP and consumer price index (CPI) forecasts in six and 12 months' time.

- Investment-capital ratio (*ik*): The investment-to-capital ratio proposed by Cochrane (1991) is the ratio of aggregate investment to aggregate capital for the US economy.

- Volatility of growth in industrial production (*ipvol*): This variable is a proxy for the conditional volatil-ity of growth in US industrial production. The construction of this variable follows Engle, Ghysels, and Sohn (2008).

- Net payout (*npy*): Following Boudoukh, Michaely, Richardson, and Roberts (2007), the net payout yield is constructed using monthly data on aggregate market capitalization, dividends, and net equity issuance from the Center for Research in Security Prices (CRSP).

- Volatility of inflation growth (*ppivol*): This variable is a proxy for the conditional volatility of inflation growth based on the producer's price index (PPI). The construction of this variable follows Engle, Ghysels, and Sohn (2008).

- Term spread (*tms*): The term spread is the difference between the long term yield on government bonds and the Treasury bill rate.

When possible, all variables are sampled at both the quarterly and monthly frequency. Several variables are not available at the monthly frequency, including changes in bank leverage, the consumption-wealth ratio, current and expected GDP growth, and the investment-capital ratio. For the empirical analysis using monthly data, I use the growth in industrial production (*ip*) as an alternative measure of economic activity. The unpublished Appendices to the paper list data sources and provide additional details regarding the construction of several variables.

3.3. Discussion

Motivated by Mele (2005, 2007), the set of predictors includes several prominent variables from the literature on return predictability. Return forecasting regressions often feature some measure of corporate payout yield. Although the dividend yield is frequently employed, Lettau and Van Nieuwerburgh (2008) find that this yield exhibits breaks in level. Boudoukh, Michaely, Richardson, and Roberts (2007) show that an alternative yield based on net payout exhibits greater stability over time and better forecasts returns relative to the dividend yield. I use the net payout yield in this study. However, results are robust to alternatively using the dividend yield. The default yield and consumption-wealth ratio are also heavily used in return forecasting regressions. Campbell and Diebold (2009) find that expected GDP growth predicts stock returns and volatility. Finally, the derived variable *exret* provides a direct proxy for unobserved expected stock returns.

The inclusion of the volatilities of inflation and production growth follows Schwert (1989) and Engle, Ghysels, and Sohn (2008). These volatilities provide information regarding the extent of uncertainty surrounding macroeconomic prospects, a feature emphasized by the literature linking learning and stock market volatility. Variables such as the default spread, commercial paper-to-Treasury spread, and default return respond aggressively at the onset of economic crises, when default probabilities for corporate debt increase dramatically. In a similar vein, Estrella and Hardouvelis (1991) find that the term spread forecasts recessions.

Changes in bank leverage help measure liquidity and credit conditions that are crucial when shocks to the economy are amplified via financial intermediation. Adrian and Shin (2010) find a positive relation between asset values and leverage for security brokers and dealers. As bank leverage increases, banks become more susceptible to the possibility of loss spirals and margin spirals. Interest rate spreads could also convey information regarding credit and liquidity conditions. For example, the commercial paper-to-Treasury spread widened in late 2007 and again in late 2008 as the financial crisis intensified.

3.4. Descriptive Statistics

Table 1 presents summary statistics for the full set of forecasting variables over the period 1952–2010. Panels A and B provide statistics for data sampled at the quarterly and monthly frequencies, respectively. The forecasting variables exhibit a wide range of persistence. The first-order autocorrelation ρ_1 is positive for most variables and is greater than 0.8 for roughly half of the variables. Highly persistent forecasting variables could cause econometric problems in forecasting regressions, particularly when shocks to the target and predictor variables are correlated (Stambaugh, 1999). To determine whether standard asymptotic results are likely to provide reasonable guidance, I report results for the unit root test suggested by Phillips and Perron (1998). The test rejects the null hypothesis of a unit root for most variables. This suggests that, while some of the forecasting variables are persistent, this persistence is not so severe so as to require alternative inference frameworks, such as near unit root asymptotics (see, e.g., Campbell and Yogo, 2006; Jansson and Moreira, 2006; and Tourus, Valkanov, and Yan, 2004).

4. In-sample analysis

Consider the following specification for log volatility:

$$LVOL_t = \alpha + \sum_{k=1}^{K} \rho_k LVOL_{t-k} + \beta' X_{t-1} + \varepsilon_t. \tag{2}$$

Under the null hypothesis of no Granger causality, $\beta = 0$. This hypothesis can be tested in a standard regression framework. When X_t is a scalar, the test takes the form of a standard t-test, whereas for vector-valued X_t the null hypothesis can be tested via an F-test.

11

4.1. Results

Table 2 presents estimation results at a quarterly horizon over a variety of sample periods. The specification includes two lags of the dependent variable [$K = 2$ in Eq. (2)].[8] For each forecasting variable, the table displays the estimated slope coefficient ($\hat{\beta}$) as well as the increase in R^2 for the regression relative to a benchmark univariate AR(2) model, expressed as a percentage. To facilitate comparisons across different forecasting variables, all variables are standardized prior to regression.[9]

For the 1927–2010 period, the null of no predictability is rejected for several variables, including the commercial paper-to-Treasury spread, the default return, and the default yield. The 1952–2010 period includes several additional forecasting variables that do not extend back to 1927. Among these, the investment-to-capital ratio emerges as a highly significant predictor. The evidence also suggests that the net payout yield and expected return proxy forecast stock return volatility over the 1952–2010 sample period.

The remaining columns of Table 2 partition the full sample into three roughly equal subsamples. The first covers 1927-1951 and includes the Great Depression, World War II, and its aftermath. The second begins in 1952 with passage of the Treasury Accord that gave the Federal Reserve the ability to pursue active monetary policy. The final subperiod begins in 1986 and covers the Great Moderation. The evidence for predictability appears strongest during the 1952–1985 subsample, both in terms of statistical and economic significance. During this period, five of the individual forecasting variables are statistically significant, and increases in R^2 relative to the benchmark univariate model tend to be higher.

Table 2 also reports results for a kitchen sink specification that includes all available forecasting variables. In place of the slope estimate $\hat{\beta}$, the table displays the F-statistic testing the null hypothesis of no Granger causality; that is, the null that (vector-valued) $\beta = 0$. The increase in R^2 relative to the benchmark AR(2) specification is also reported. The null hypothesis of no Granger causality is rejected for all sample periods. Consistent with results for regressions that consider the predictive variables one by one, the evidence for predictability is particularly strong during the 1952–1985 subsample.

Table 3 presents in-sample regression results at a monthly sampling frequency, with six lags of the dependent variable included in the specification [$K = 6$ in Eq. (2)]. Moving from quarterly to monthly sampling constrains the set of forecasting variables, as some predictors are not available monthly. However,

---

[8]The choice of two lags of the dependent variable is supported by information criteria such as AIC and BIC, which select an AR(2) specification among AR alternatives over most subsamples examined. Results reported in the unpublished Appendix illustrate similar findings for regressions that include four lags of the dependent variable.

[9]Because all variables are standardized prior to estimation, the intercept α in Eq. (2) is omitted. Results apply standard errors corrected for heteroskedasticity and serial correlation.

monthly sampling might provide additional power to detect forecasting ability. The results under monthly sampling are consistent with those under quarterly sampling. Strong statistical evidence exists that the commercial paper-to-Treasury spread and the default return Granger cause volatility. For some subsamples, the evidence suggests that the default spread, net payout yield, expected return proxy, and inflation volatility forecast volatility. Consistent with results under quarterly sampling, the null of no Granger causality is strongly rejected for the kitchen sink specification that includes all variables simultaneously.

4.2. Economic significance and summary

Because all variables are standardized prior to analysis, the $\hat{\beta}$ values reported in Tables 2 and 3 reflect the impact on the log volatility forecast, measured in units of a standard deviation, of a 1 standard deviation change in the forecasting variable, ceteris paribus. For example, a coefficient estimate of 0.25 implies that a 1 standard deviation shock to the forecasting variable increases the forecast for log volatility in the subsequent period by one-quarter of a standard deviation. Most of the actual estimates reported in Tables 2 and 3 are smaller than 0.25, with many smaller than 0.1.

The increase in R^2 attributable to adding a given forecasting variable (or variables) to the benchmark model provides another metric to assess the economic significance of forecast improvements. For univariate models (in the sense of including a single macroeconomic or financial predictor), the increase in R^2 is often less than 1%, and is usually less than 3% even in cases with strong statistical evidence for predictability. One exception is the relatively large increase in R^2 of 7.78% achieved by the commercial paper-to-Treasury spread over the 1952–1985 subsample (see Table 2). At a quarterly sampling frequency, the kitchen sink specifications generate (unadjusted) R^2 improvements ranging from 3% to 11.5%, depending on the sample period examined, with the most optimistic results occuring over the 1952–1985 sample period.

Overall, the in-sample results suggest that macroeconomic economic variables do Granger cause stock return volatility at monthly and quarterly horizons. Among the variables considered, the most useful appear to be the commercial paper-to-Treasury spread, the default return, the default spread, the expected return proxy, and the investment-to-capital ratio. Two empirical facts temper the optimistic tone of these results. First, the economic significance of the predictive power afforded by these variables is relatively small. Second, the evidence for predictability is particularly strong from 1952 to 1985 and less emphatic during the Great Moderation that follows. These points raise the question of whether in-sample predictive power extends to an out-of-sample research design.

5. Out-of-sample analysis

The extensive literature on stock return predictability warns that models that appear superior from an in-sample perspective could perform poorly out-of-sample. In particular, Goyal and Welch (2008) show that many return forecasting models perform poorly from an out-of-sample perspective and appear to break down during the Great Moderation. This section applies similar scrutiny to volatility forecasting models that include macroeconomic and financial predictors. The forecasting models are of the form Eq. (2). As benchmarks, I consider simple univariate AR models, i.e., models of the form Eq. (2) with $\beta = 0$. Under quarterly sampling the benchmark model is an AR(2) specification, and under monthly sampling the benchmark is an AR(6) specification.

Out-of-sample tests split the data into two subsets, the first R observations constituting an estimation sample and the final P observations constituting a holdout sample used to estimate the MSPE associated with each model. Under the out-of-sample approach, forecasts for the holdout sample are constructed using past information, as though the econometrician were constructing real time forecasts.[10]

The question of interest is whether models that include additional macroeconomic and financial predictors are superior to the benchmark from an out-of-sample perspective. Unfortunately, the notion of a superior model is somewhat vague, with multiple potential interpretations. One interpretation emphasizes properties of the data generating process: Does an out-of-sample analysis support the notion that macroeconomic variables Granger cause stock return volatility? An alternative interpretation focuses on a normative stance: Do volatility models that incorporate macroeconomic variables improve the accuracy of out-of-sample forecasts?

To illustrate the distinction between these two interpretations, suppose the true model is

$$LVOL_t = \alpha + \rho_1 LVOL_{t-1} + \rho_2 LVOL_{t-2} + \beta X_{t-1} + \varepsilon_t, \tag{3}$$

where X_t represents some forecasting variable. So long as $\beta \neq 0$, X_t Granger causes volatility. In this case, the population MSPE of forecasts associated with Eq. (3) must be lower than the population MSPE for the benchmark univariate AR(2) model.

In the realistic case in which model parameters must be estimated using historical data, forecasts pro-

---

[10]Implementations vary in terms of how the estimation sample evolves through time. In some cases, parameter values are estimated once using the initial R observations (fixed sample), in others a moving window of size R is used to produce forecasts (rolling), and in some cases an initial window of size R is expanded as new data become available (recursive).

duced under model Eq. (3) could, in fact, under-perform forecasts based on the benchmark, even when the latter is misspecified. This is due to the well-known bias-variance trade-off under mean square error loss. Intuitively, when $\beta \neq 0$, there is both a benefit and a cost to adding X_{t-1} to the forecasting model. The benefit is a reduction in conditional forecast bias. The cost is higher forecast variability related to the need to estimate the additional parameter β. The latter could outweigh the former. In such a case, the benchmark model is superior in terms of mean square forecast error, while, at the same time, the alternative model is superior in the sense that it is correctly specified.

For further intuition, suppose that a target variable Y_t follows an independent and identically distributed normal distribution with unknown mean μ and standard deviation σ^2. A researcher poses the following alternative models for Y_t:

Model 1 $Y_t = 0$.

Model 2 $Y_t = \mu + \varepsilon_t$, $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$.

The misspecified Model 1 assumes that Y_t is zero in every period, while Model 2 is correctly specified. If the relevant empirical question is whether Model 1 or Model 2 provides a more appropriate description of the DGP for Y_t, the answer is obvious. It is trivial to reject Model 1 in favor of Model 2 (a sample size as small as one is sufficient, because the probability that $Y_t = 0$ is zero).

Now suppose the empirical focus centers on which model delivers more accurate forecasts for a given sample size N, assuming squared error loss. There is no need to estimate model parameters to produce forecasts under Model 1. The MSPE for this model is $\sigma^2 + \mu^2$, where the term μ^2 captures a cost associated with forecast bias and the term σ^2 reflects unforecastable noise in Y_t. For Model 2, assume that the unknown parameter μ is estimated using the sample mean of Y_t. Forecasts under Model 2 are consequently unbiased, but estimation error associated with μ generates an additional component in MSPE. The MSPE for this forecast is $\sigma^2 + \frac{\sigma^2}{N}$, where the second term captures the cost due to parameter estimation. Comparing the two MSPEs, the obviously misspecified Model 1 produces more accurate forecasts whenever $\mu^2 < \frac{\sigma^2}{N}$. All else equal, this is more likely to occur when $|\mu|$ is smaller, σ^2 is larger, and the sample size N is smaller. In the empirical exercise of interest in this paper, the models are more complicated, and the random variables involved are persistent. Still, the intuition from this example remains useful. Even if a variable Granger causes volatility, such that $\beta \neq 0$ in Eq. (3), the benchmark univariate model could yield more accurate forecasts. This is more likely to occur when β is relatively small, and for smaller sample sizes.

5.1. Out-of-sample forecasting tests

This section describes out-of-sample tests for Granger causality and superior predictive ability.

5.1.1. Testing for Granger causality

Let $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ represent the (population) forecast errors for the benchmark and augmented models, respectively. Under the null of no Granger causality, the benchmark and alternative models have equal population MSPE, i.e., $E(\varepsilon_{1,t}^2) = E(\varepsilon_{2,t}^2)$. In most applications, inference is complicated by the fact that forecasts are based on estimates of population parameters. West (1996) provides a general treatment of asymptotically valid inference when forecasts are constructed based on estimated parameters and states conditions under which standard t-tests of the null are asymptotically standard normal. Unfortunately, the cases of interest in this paper violate the conditions for asymptotic normality, because the forecasting models of interest are nested.

Clark and West (2007) suggest a test for equal MSPE for nested models that is approximately standard normal. The approach of Clark and West (2007) rests on the following idea. Under the null hypothesis, the more parsimonious model (the benchmark model) generates the data and the MSPE for this model is expected to be smaller than the MSPE for the larger model. Clark and West (2007) adjust the out-of-sample estimate of the MSPE difference to account for additional noise associated with the larger model's forecast.

Let $\widehat{LVOL}_{1,t+1}$ and $\widehat{LVOL}_{2,t+1}$ denote one-step ahead forecasts of $LVOL_{t+1}$ for the benchmark and augmented models, respectively. The corresponding sample MSPEs are $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, with

$$\hat{\sigma}_i^2 \equiv P^{-1} \sum (LVOL_{t+1} - \widehat{LVOL}_{i,t+1})^2 \tag{4}$$

for $i = 1, 2$ and where P denotes the number of out-of-sample forecast observations. Clark and West (2007) propose a test of the null of equal MSPE based on the statistic:

$$\text{Clark and West (CW)} = \hat{\sigma}_1^2 - \hat{\sigma}_2^2 + \underbrace{P^{-1} \sum (\widehat{LVOL}_{1,t+1} - \widehat{LVOL}_{2,t+1})^2}_{\text{Adjustment}}. \tag{5}$$

The final term in Eq. (5) captures the adjustment for additional noise associated with the larger model's forecast. This adjustment involves the average squared difference between forecasts generated by the smaller and larger models. When the forecasts from the larger model are highly volatile relative to forecasts based on the benchmark (parsimonious) model, the additional noise related to parameter estimation is large, and the adjustment term in Eq. (5) is correspondingly large. Although their proposed adjusted-MSPE statistic

16

is not asymptotically normal, Clark and West (2007) argue that standard normal critical values lead to tests with actual sizes close to, but slightly smaller than, nominal size for reasonably large samples. The test is one-sided, because under the alternative $\sigma_2^2 < \sigma_1^2$.[11]

5.1.2. Testing for superior predictive ability

The alternative question of whether macroeconomic and financial variables improve out-of-sample forecasts is nuanced, as this question involves the entire forecasting method and not simply the forecasting model. A forecasting method is a broad concept that encompasses not only a set of model specifications, but also the detailed procedure used to obtain forecasts. This procedure includes the method(s) used to estimate unknown parameters, the choice of estimation window, and so forth.[12]

Given two alternative forecasting methods, Giacomini and White (2006) propose a test for superior predictive ability. Let $\mathscr{G}_t$ represent some set of information available at time t. Under the null hypothesis, the two models possess equal forecasting ability. Adapted to the setting of interest in this paper, the null is

$$H_0 : E(\hat{\sigma}_1^2 - \hat{\sigma}_2^2 | \mathscr{G}_t) = 0. \tag{6}$$

The null hypothesis (6) differs from that considered by West (1996), Clark and West (2007), and related studies in an important way. The null hypothesis involves expectations of the estimated MSPE, not the population MSPE, of the two forecasting models. Consequently, the test explicitly captures the impact of parameter estimation uncertainty on forecast performance. Because the test involves forecast methods instead of forecast models, virtually any forecast can be accommodated, including forecasts formed as data-driven combinations of underlying forecasts. The Clark and West testing framework, by comparison, assumes that forecasts are based on linear parametric models estimated using least squares.

The general framework developed by Giacomini and White (2006) permits conditional comparison of forecast performance. Taking $\mathscr{G}_t = \{\emptyset, \Omega\}$, the trivial information set, results in an unconditional test of equal predictive ability. The Giacomini and White (2006) test then takes the form

$$\text{Giacomini and White (GW)} = \frac{\hat{\sigma}_1^2 - \hat{\sigma}_2^2}{\hat{\sigma}_P / \sqrt{P}}, \tag{7}$$

---

[11]Computationally, the test can be executed by regressing the adjusted out-of-sample squared forecast error differences on a constant and examining the associated t-statistic.

[12]For example, two forecasts produced from the same model, one based on a rolling estimation window and the other produced using an expanding window, represent different forecasting methods.

where $\hat{\sigma}_P$ is a heteroskedasticity and autocorrelation consistent (HAC) estimator of the asymptotic variance $\sigma_P^2 = \text{var}\left[\sqrt{P}(\hat{\sigma}_1^2 - \hat{\sigma}_2^2)\right]$. By contrast with the Clark and West test, the Giacomini and White test of equal (unconditional) predictive ability is a two-sided test. The test statistic Eq. (7) is equivalent to the test statistic proposed by Diebold and Mariano (1995), and the asymptotic results in Giacomini and White (2006) provide a rigorous justification for this test when forecast parameters are estimated.

5.2. Design specifics

Forecasting models are estimated using either a rolling or recursive procedure with an initial sample of 20 years of data (80 quarters or 240 months). The analysis explores several alternative out-of-sample periods. The first extends from 1947 through 2010. While this period affords a long horizon over which to assess forecasting ability, several forecasting variables are unavailable. An alternative out-of-sample period covers 1972–2010 and includes additional forecasting variables such as the investment-to-capital ratio and expected GDP growth. The final two out-of-sample periods cover 1972–2000 and 1982–2010. These periods focus on the sensitivity of results to inclusion of the 1970s in the out-of-sample period. Goyal and Welch (2008) find that the apparent ability of some variables to forecast stock returns is heavily influenced by the oil shock of 1973–1975. The 1972–2000 and 1982–2010 periods contain the same number of out-of-sample observations, but the former period includes the turbulent 1970s, while the latter does not.

In addition to univariate models that include only a single macroeconomic predictor, and a kitchen sink model that includes all predictors, I evaluate several forecast combinations. Rapach, Strauss, and Zhou (2010) show that simple forecast combination methods improve the out-of-sample performance of stock return forecasts based on multiple predictors. The combined forecasts take the form

$$\widehat{LVOL}_{t+1} = \sum_{n=1}^{N} \hat{\omega}_{n,t}\widehat{LVOL}_{n,t+1}, \tag{8}$$

where $\hat{\omega}_{n,t}$ represents the combining weight on the nth individual forecast at time t. The notation emphasizes that combining weights are permitted to be data-driven.

The first two combined forecasts are computed as the mean and median across the N individual forecasts. The third combined forecast is a trimmed mean that sets $\hat{\omega}_{n,t} = 0$ for the largest and smallest volatility forecast each period, with the remaining $N - 2$ forecasts equally weighted. The final combined forecast (labeled "MSPE") follows Stock and Watson (2004) and adjusts combining weights based on historical out-of-sample performance. Specifically, the combining weights evolve according to

18

$$\hat{\omega}_{n,t}^{MSPE} = \frac{\phi_{n,t}^{-1}}{\sum_{j=1}^{N} \phi_{j,t}^{-1}}, \tag{9}$$

where

$$\phi_{n,t} \equiv \sum_{p=R}^{t-1} \left(LVOL_{p+1} - \widehat{LVOL}_{n,p+1} \right)^2. \tag{10}$$

5.3. Results

Tables 4 and 5 present out-of-sample results using a rolling estimation scheme for quarterly and monthly sampling frequencies, respectively. For each model and each out-of-sample period, the table presents as CW the adjusted MSPE test statistic Eq. (5) of Clark and West (2007). Asterisks indicate rejections of the null of no Granger causality at conventional levels.[13]

To convey the economic significance of differences in forecast performance, I follow Campbell and Thompson (2008), Goyal and Welch (2008), and Rapach, Strauss, and Zhou (2010) and consider the out-of-sample R^2 statistic, defined as

$$R_{OOS}^2 = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}, \tag{11}$$

where $\hat{\sigma}^2$ represents the out-of-sample MSPE for the model of interest and $\hat{\sigma}_0^2$ represents the out-of-sample MSPE based on the historical average. A measure of the economic significance of forecast improvement relative to the univariate benchmark is

$$\Delta R_{OOS}^2 \equiv \left(1 - \frac{\hat{\sigma}_2^2}{\hat{\sigma}_0^2} \right) - \left(1 - \frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2} \right) = \frac{\hat{\sigma}_1^2 - \hat{\sigma}_2^2}{\hat{\sigma}_0^2}, \tag{12}$$

expressed as a percentage. Intuitively, this is the out-of-sample analog of the ΔR^2 statistic reported in Tables 2 and 3 for in-sample regressions. Asterisks following the ΔR_{OOS}^2 statistic indicate rejections of the null of equal predictive ability based on the Giacomini and White test [Eq. (7)].

The Clark and West test results largely corroborate in-sample findings. Under quarterly sampling (Table 4), the commercial paper-to-Treasury spread, default return, and inflation volatility Granger cause stock return volatility over the 1947–2010 out-of-sample period. For the 1972–2010 out-of-sample period, these same variables, along with bank leverage and the investment-to-capital ratio, Granger cause volatility. Comparing results for 1972–2000 with those for 1982–2010 shows that evidence of forecasting power for stock

---

[13]To enhance readability, the statistic is scaled by a factor of one thousand. I report Clark and West (2007) test statistics for the combined forecasts. However, these forecasts are not technically covered by the asymptotic results in Clark and West (2007), which require that forecasts be generated from linear models estimated by OLS.

return volatility is sensitive to inclusion of the turbulent 1970s in the sample. As an example, the null of no Granger causality is rejected for the kitchen sink model over the 1972–2000 out-of-sample period, but not for the 1982–2010 period of equal length.

The evidence for superior predictive ability, as captured by ΔR^2_{OOS} values and the Giacomini and White (2006) test results, is weaker. Forecasts from univariate models (in the sense of including a single lagged macroeconomic predictor) generally yield ΔR^2_{OOS} values that are negative, or small. In most cases, the null of equal predictive ability cannot be rejected. For the univariate models, the only instances of rejections correspond to instances in which the benchmark produces superior forecasts. The kitchen sink model yields ΔR^2_{OOS} values that are negative. With the exception of the 1972–2000 period, the null of equal predictive ability is rejected in favor of the univariate benchmark model. Consistent with the equity premium analysis in Rapach, Strauss, and Zhou (2010), forecast combinations deliver improved out-of-sample performance. For all out-of-sample periods except 1982–2010, the combined forecasts outperform the benchmark. In several cases, primarily associated with the 1972–2000 period, the forecast improvements are statistically significant.

The results reported in Table 4 starkly illustrate the distinction between the two out-of-sample tests. In many instances in which the Clark and West test rejects, the Giacomini and White test fails to reject, and vice versa. As an illustrative example, consider the performance of the kitchen sink specification over the 1972–2000 out-of-sample period. The Clark and West test statistic is positive and significant, yet the ΔR^2_{OOS} value is negative, implying that the benchmark model performs better. The large, positive Clark and West test statistic is attributable to a large upward adjustment term in Eq. (5). This adjustment is based on the variability of forecasts under the more heavily parameterized model relative to the benchmark. Forecasts under the kitchen sink model are substantially more variable than those under the benchmark, so the adjustment term is relatively large. Intuitively, although the kitchen sink model under-performs the benchmark out-of-sample, it does not under-perform it by enough to be consistent with the null hypothesis of no Granger causality. Hence, the Clark and West test rejects.

Table 5 presents out-of-sample results at the quarterly frequency under a recursive, as opposed to rolling, estimation scheme. Most major findings continue to obtain. Specifically, the Clark and West test indicates that several of the individual macroeconomic variables, including the commercial paper-to-Treasury spread, default return, and investment-to-capital ratio, Granger cause stock return volatility. The Giacomini and White test, however, rarely rejects the null of equal predictive ability for the univariate models, and the few

20

rejections are in favor of the benchmark. Finally, the combined forecasting models again outperform the benchmark, with forecast improvements that are often statistically significant, except over the 1982–2010 subperiod.[14]

The contrast between results under rolling and recursive schemes is more substantial for the kitchen sink specification. Under a rolling scheme (Table 4), the kitchen sink model significantly under-performs the benchmark, with the difference being both statistically and economically significant in some subsamples. By contrast, the kitchen sink model often outperforms the benchmark under the recursive scheme (Table 5), although the differences tend not to be significant. To understand this pattern of results, recall that, under the rolling scheme, the estimation sample is limited to 20 years of historical data throughout the out-of-sample period. Under the recursive scheme, the estimation sample grows continually. The increasing sample sizes under the recursive scheme deliver less volatile forecasts and superior out-of-sample performance.[15] While this intuition holds for univariate models as well as the kitchen sink model, effects tend to be more pronounced in the latter case, as the model is relatively heavily parameterized.

Table 6 presents out-of-sample results at a monthly frequency based on a rolling estimation window. These results exhibit the same main features as results at the quarterly frequency. Among the variables available at a monthly frequency, strong evidence exists that the commercial paper-to-Treasury spread and default return Granger cause stock return volatility. The Clark and West test also rejects the null of no Granger causality for the kitchen sink model. At the same time, the Giacomini and White test rarely rejects the null of equal predictive ability. As with previous results, forecast combinations yield statistically significant improvements over the benchmark, although the forecasting gains are rather small. An unpublished Appendix contains monthly results using a recursive estimation scheme, as well as additional results exploring robustness to alternative proxies for volatility, attempts to forecast the level (as opposed to logarithm) of volatility, and the specification of the benchmark model. In all cases the findings are qualitatively similar.

5.4. Monte Carlo analysis

To shed additional light on the pattern of results obtained in the out-of-sample study, I conduct a Monte Carlo simulation experiment. Data are simulated from the following bivariate data generating process:

---

[14]The asymptotic results in Giacomini and White (2006) require that the estimation sample remain bounded. Strictly speaking, this condition is violated under a recursive scheme in which the sample grows without bound. With this caveat in place, I report inference results in Table 5 for comparison with those in Table 4.

[15]This reasoning suggests that recursive schemes are always preferable to rolling schemes. In practice, it is less clear which scheme is preferable, because forecasts could better adapt to unmodeled structural change under a rolling scheme.

$$Y_t = -0.61 + 0.52Y_{t-1} + 0.22Y_{t-2} + \beta X_{t-1} + \varepsilon_t,$$

$$X_t = 0.27 + 0.60X_{t-1} + v_t,$$

$$var(\varepsilon_t) = 0.092, \ var(v_t) = 0.160, \ \text{and} \ cov(\varepsilon_t, v_t) = 0.026, \tag{13}$$

where ε_t and v_t are independent, bivariate normal shocks with the indicated covariance structure. The calibration is based on estimates of a restricted VAR(2) for log volatility and the commercial paper-to-Treasury spread (cp) at the quarterly frequency. In one set of simulations, the parameter β in Eq. (13) is set to the corresponding in-sample estimate.[16] A second set of simulations imposes the null of no Granger causality by setting $\beta = 0$.

The simulation analysis focuses on a rolling estimation scheme and considers estimation sample sizes R ranging from 20 (five years of quarterly data) to one thousand (250 years of quarterly data). Most sample sizes cluster in the empirically relevant range of five to 40 years of historical data. Two choices for the out-of-sample size P are entertained: $P = 120$ (30 years) and $P = 240$ (60 years). These roughly correspond to the out-of-sample periods featured in the empirical results. For each set of simulated data, one-step-ahead predictions are computed for $t = R + 1$ through $t = R + P$ for both a benchmark AR(2) specification and a forecasting regression that also includes X_{t-1}. The resulting out-of-sample squared forecast error differences are used to construct the CW and GW test statistics. Of central interest are the probabilities that these tests reject their respective null hypotheses at the 10% level. The Monte Carlo exercise approximates these probabilities via the proportion of rejections over ten thousand simulated samples.

Panel A of Table 7 presents results for the case $\beta > 0$, and Panel B presents results for the case $\beta = 0$ (no Granger causality). For reference, the far-right column of the table presents the true ΔR^2_{OOS}, i.e., the value of ΔR^2_{OOS} for $P = \infty$. This is computed via simulation for each value of R. To understand the pattern of ΔR^2_{OOS} values, consider the simulation where $\beta > 0$ (Panel A), so that X_t Granger causes Y_t. For sufficiently large R, estimation uncertainty regarding β is minimal and the model that includes X_{t-1} has superior predictive ability (lower MSPE). For very small sample sizes, estimation error associated with β is sufficiently severe such that the misspecified, but simpler, benchmark model achieves superior predictive ability. Consistent

---

[16]This estimate implies an increase in population R^2 relative to an AR(2) benchmark of slightly over 1%. This is consistent with the typical increase in forecasting power associated with statistically significant predictors (see Table 3).

with this reasoning, the ΔR^2_{OOS} values reported in Panel A range from around -3.5% for $R = 20$ to just under 1% for $R = 1,000$. It is notable that $\Delta R^2_{OOS} = 0$ (equal predictive ability) for a sample size in the neighborhood of 60 observations (15 years). In Panel B, ΔR^2_{OOS} is always negative, because in this case X_{t-1} is an extraneous forecasting variable. At the smallest sample size, ΔR^2_{OOS} is approximately -5%. As expected, ΔR^2_{OOS} approaches zero as the estimation sample becomes very large.

Panel A addresses the case in which X_t Granger causes Y_t, so that the relevant null for the CW test is false. In the calibrated simulations, the CW test exhibits reasonable power, particularly for the longer out-of-sample evaluation period ($P = 240$). Power for the CW test is strictly increasing in the estimation sample size R. For $R = 80$, which corresponds to the sample size in the actual empirical analysis, the power of the test is approximately 40% when $P = 120$ and 60% when $P = 240$.

Turning to the GW test, the null hypothesis of equal predictive ability is virtually true when $R = 60$, in the sense that $\Delta R^2_{OOS} \approx 0$. Rejections for this case reflect the size of the test, and it appears that the test is slightly undersized. For most sample sizes, the difference in predictive ability between the two models, as measured by ΔR^2_{OOS}, is small. Consequently, it is not surprising that the power of the GW test is close to its theoretical size of 10%. In short, the GW test has little power to distinguish between the two models when the true difference in MSPE is so slight. Only for extreme sample sizes (large or small) does the GW test exhibit much power, and even in these cases power is well below 50%. Finally, when the CW test rejects, it is typically the case that the GW test fails to reject. For example, the CW test rejects around 60% of the time when $R = 80$ and $P = 240$, while the event "CW test rejects and the GW test fails to reject" occurs nearly as often (56% of the time).

Under the null of no Granger causality (results presented in Panel B), the rejection proportions reported for the CW test characterize size. However, the null hypothesis of equal predictive ability is false, so the rejection probabilities for the GW test indicate power. The CW test appears to be relatively well-sized. For the smallest sample sizes the test is slightly oversized, and for larger sample sizes the test is slightly undersized, consistent with simulation results reported by Clark and West (2007). The power of the GW test is decreasing in R, which reflects that fact that as R increases the deviation from the null of equal predictive ability decreases. For smaller sample sizes, the GW test has noticeably higher power relative to similar sample sizes in Panel A. This is due to the fact that the magnitude of the ΔR^2_{OOS} values tends to be larger, i.e., the deviation from the null of equal predictive ability is larger.

The Monte Carlo results are consistent with the paper's empirical findings. First, the simulations suggest

that, for univariate models, the GW test is unlikely to have much power to discern between the benchmark and alternative models. When the GW test does have power, it is likely to be in the direction of rejecting the null in favor of the benchmark model. Under the same conditions, the CW test has reasonable power to detect Granger causality out-of-sample. Situations in which the CW test rejects the null, while the GW test fails to reject, occur frequently in the simulations as in the actual data.

5.5. Out-of-sample forecast performance and the business cycle

Exploring connections between the business cycle and relative forecast performance may shed light on the sources of predictive power contained in the macroeconomic and financial variables. Let $\Delta SE_{i,t}$ represent the difference between the squared forecast error for the benchmark univariate model (indexed as UV) and the squared forecast error for the ith forecasting model:

$$\Delta SE_{i,t} \equiv \left(LVOL_t - \widehat{LVOL}_{UV,t} \right)^2 - \left(LVOL_t - \widehat{LVOL}_{i,t} \right)^2 . \tag{14}$$

Fig. 2 displays time series plots of cumulative $\Delta SE_{i,t}$ for the out-of-sample period 1947–2010.[17] Periods when the plot line slopes upward represent periods in which the corresponding forecast model outperforms the benchmark, while downward-sloping segments indicate periods when the benchmark forecast is more accurate. Vertical lines indicate business cycle peaks, i.e., the point at which an economic expansion transitions to a recession based on NBER business cycle dating. Under the null of no Granger causality (the relevant null for the Clark and West test), one would expect these plots to trend steadily downward, as additional estimation error associated with the more heavily parameterized model increases the cumulative squared error relative to the benchmark.

Fig. 2 shows that the timing of out-of-sample predictive power for several models concentrates near the onset of recession periods. For example, the commercial paper-to-Treasury spread forecasts well following a series of recessions in the 1960s and 1970s. The default return delivers significant out-of-sample improvements following the financial crisis of 2008. Inflation volatility forecasts well following recessions in the 1970s and following the financial crisis of 2008. Other variables, such as the term spread and the volatility of industrial production, seem to perform poorly independently of economic conditions, consistent with the null of no Granger causality.

---

[17]The figure is based on quarterly sampling and a recursive estimation window. This corresponds to the left-most set of results in Table 5. Plots based on monthly data, or rolling estimation windows, are qualitatively similar.

To supplement the graphical analysis presented in Fig. 2, I also examine regressions of $\Delta SE_{i,t}$ on the growth rate in quarterly GDP (gdp):

$$\Delta SE_{i,t} = \alpha + \beta \; gdp_t + \varepsilon_t. \tag{15}$$

This regression explores whether differences in forecast performance vary over the business cycle. Table 8 reports the estimated coefficient $\hat{\beta}$, the t-statistic associated with this coefficient, and the R^2-value for the regression (as a percentage). The variables $\Delta SE_{i,t}$ and gdp_t are standardized prior to regression. Hence, the regression intercept α is omitted from the specification.

A negative estimate of β indicates a countercyclical pattern in forecasting power (relative to the benchmark). Consistent with the plots of cumulative differences in squared forecast error, the relative performance for several forecasts appears to be countercyclical. The evidence is particularly strong for the 1972–2010 subsample, where the null that forecast performance is unrelated to the business cycle is rejected for a number of predictors including the commercial paper-to-Treasury spread, the default spread, and the net payout yield. The last variable is somewhat anomalous, as its forecasting power appears to be procyclical. Both the kitchen sink and combined forecasts exhibit countercyclical performance over both sample periods. This suggests that, on the whole, forecast improvements relative to the benchmark are countercyclical.

The evidence in Fig. 2 and Table 8 suggests an economic interpretation for more successful forecasting variables: These variables appear to capture a second factor beyond current economic conditions and related to the onset of recessions. For example, the commercial paper-to-Treasury spread and the default return could capture information regarding liquidity and credit conditions around the onset of recessions, when financial intermediation amplifies negative shocks as discussed in Brunnermeier and Pedersen (2009). Alternatively, or in addition, uncertainty about future economic growth rates could be high during these periods. The linkage between forecast performance and recessions appears to be complex. Different variables exhibit different patterns of forecast performance around recessions. For example, the default return responds aggressively to the financial crisis of 2008, but not to earlier recessions in the 1960s and 1970s. The opposite holds for the commercial paper-to-Treasury spread. A two-state characterization of the economy (expansion versus recession) could be an oversimplification. Linkages between macroeconomic variables and stock market returns might evolve with underlying shifts in the economic and policy environment. This provides a further motivation for combining forecasts based on different variables, as the extent to which such variables characterize the true economic state could vary with time.

6. Conclusion

This study examines forecasting regressions for stock return volatility that attempt to exploit information contained in macroeconomic and financial variables. Motivated by theoretical literature, I identify a set of candidate predictors and test the ability of these variables to improve volatility forecasts. Out-of-sample tests for Granger causality confirm in-sample results indicating that several variables, including the commercial paper-to-Treasury spread, default return, default spread, and the investment-to-capital ratio do help forecast volatility in a population sense. The Giacomini and White (2006) test for superior (unconditional) predictive ability, however, rarely indicates a statistical difference in the mean square prediction error between these forecasts and the univariate benchmark. Simple forecast combination schemes do statistically outperform the benchmark, although not by a large margin. Finally, the forecasting ability possessed by the macroeconomic variables concentrates around the onset of recessions, consistent with the notion that these variables capture information regarding factors that drive volatility, including macroeconomic uncertainty, time-varying expected stock returns, and credit conditions.

Macroeconomic and financial data are more abundant than ever before. Because volatility co-varies with business conditions, a tendency exists to suspect that incorporating macroeconomic information should greatly improve longer horizon volatility forecasts. The relatively comprehensive analysis in this paper shows that only modest forecasting gains are possible. The explanation has not changed since Schwert (1989): Volatility co-moves tightly with the business cycle, and lagged volatility itself contains a wealth of information about business conditions. While some evidence shows that interest rate spreads and other variables help predict volatility increases around the onset of recessions, leveraging these variables into large out-of-sample forecast improvements is difficult.

The paper relies on a realized variance paradigm, in which volatility is treated as observed, and focuses on simple, linear forecasting models. A multitude of alternative approaches exist for modeling and forecasting volatility. These include generalized autoregressive conditional heteroskedasticity (GARCH) and related parametric specifications, stochastic volatility models, regime switching models, approaches based on implied volatility, and so on. Along these lines, several recent, related papers propose sophisticated approaches for incorporating macroeconomic information into volatility forecasts. Engle, Ghysels, and Sohn (2008) develop a GARCH-MIDAS (mixed data sampling) modeling framework that decomposes variation in volatility into a short run and a secular component. Macroeconomic factors can be incorporated through a MIDAS polynomial (see, e.g., Ghysels, Santa-Clara, and Valkanov, 2005), while the short-run volatility component adheres to a standard GARCH process. David and Veronesi (2009) propose and estimate a struc-

tural model in which investors learn about regular and unusual states for earnings growth and inflation. In their framework, the econometrician (or forecaster) is not assumed to possess the same information set as market participants but extracts this information using prices. The present paper complements these studies by suggesting other useful observable signals that might be incorporated into more sophisticated modeling approaches. Investigating the relative merits of different modeling approaches represents an important topic for future research.

References

Adrian, T., Shin, H. S., 2010. Liquidity and leverage. Journal of Financial Intermediation 19, 418–437.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 529–626.

Aruoba, S., Diebold, F. X., Scotti, C., 2009. Real-time measurement of business conditions. Journal of Business and Economic Statistics 27, 417–427.

Bansal, R., Yaron, A., 2004. Risks for the long run: a potential resolution of asset pricing puzzles. Journal of Finance 59, 1481–1509.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B 64, 253–280.

Boudoukh, J., Michaely, R., Richardson, M., Roberts, M. R., 2007. On the importance of measuring payout yield: Implications for empirical asset pricing. Journal of Finance 62, 877–915.

Breen, W., Glosten, L., Jagannathan, R., 1989. Economic significance of predictable variations in excess stock returns. Journal of Finance 44, 1177–1189.

Brennan, M. J., Xia, Y., 2001. Stock price volatility and equity premium. Journal of Monetary Economics 47, 249–283.

Brunnermeier, M. K., Pedersen, L. H., 2009. Market liquidity and funding liquidity. Review of Financial Studies 22, 2201–2238.

Campbell, J., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.

Campbell, J. Y., Thompson, S. B., 2008. Predicting excess stock returns out of sample: can anything beat the historial average? Review of Financial Studies 21, 1509–1531.

Campbell, J. Y., Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics 81, 27–60.

Campbell, S. D., Diebold, F. X., 2009. Stock returns and expected business conditions: half a century of evidence. Journal of Business and Economic Statistics 27, 266–278.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Cochrane, J. H., 1991. Production-based asset pricing and the link between stock returns and economic fluctuations. Journal of Finance 46, 207–234.

David, A., Veronesi, P., 2009. What ties return volatilities to price valuations and fundamentals? Unpublished working paper.

Diebold, F. X., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–265.

Engle, R. F., Ghysels, E., Sohn, B., 2008. On the economic sources of stock market volatility. Unpublished working paper.

Engle, R. F., Rangel, J. G., 2008. The spline-garch model for low-frequency volatility and its global macroeconomic causes. Review of Financial Studies 21, 1187–1222.

Estrella, A., Hardouvelis, G., 1991. The term structure as a predictor of real economic activity. Journal of Finance 46, 555–576.

French, K. R., Schwert, G., Stambaugh, R. F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Ghysels, E. P., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return trade-off after all. Journal of Financial Economics 76, 509–548.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578.

Glosten, L. R., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess retrun on stocks. Journal of Finance 48, 1779–1801.

Goyal, A., Welch, I., 2008. A comprehensive look at the empirical performance of the equity premium prediction. Review of Financial Studies 21, 1455–1508.

Harvey, C. R., 2001. The specification of conditional expectations. Journal of Empirical Finance 8, 573–637.

Henkel, S., Martin, J., Nardari, F., 2011. Time-varying short-horizon predictability. Journal of Financial Economics 99, 560–580.

Jansson, M., Moreira, M., 2006. Optimal inference in regression models with nearly integrated regressors. Econometrica 74, 681–714.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth and expected stock returns. Journal of Finance 56, 815–849.

Lettau, M., Ludvigson, S., 2010. Measuring and modeling variation in the risk-return trade-off. Handbook of Financial Econometrics 1, 617–690.

Lettau, M., Van Nieuwerburgh, S., 2008. Reconciling the return predictability evidence. Review of Financial Studies 21, 1607–1652.

Ludvigson, S., Ng, S., 2007. The empirical risk-return trade-off: a factor analysis approach. Journal of Financial Economics 83, 171–222.

MacKinnon, J. G., 1994. Approximate asymptotic distribution functions for unit-root and cointegration tests. Journal of Business and Economic Statistics 12, 167–176.

Marquering, W., Verbeek, M., 2004. The economic value of predicting stock index returns and volatility. Journal of Financial and Quantitative Analysis 39, 407–429.

Mele, A., 2005. Understanding stock market volatility. London School of Economics Financial Market Groups Review 67, 10–15.

Mele, A., 2007. Asymmetric stock market volatility and the cyclical behavior of expected returns. Journal of Financial Economics 86, 446–478.

Phillips, P. C., Perron, P., 1998. Testing for a unit root in time series regression. Biometrika 75, 335–346.

Rapach, D. E., Strauss, J. K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862.

Schwert, G., 1989. Why does stock market volatility change over time? Journal of Finance 44, 1115–1153.

Shanken, J. A., 1990. Intertemporal asset pricing: an empirical investigation. Journal of Econometrics 45, 99–120.

Spiegel, M., 2008. Forecasting the equity premium: Where we stand today. Review of Financial Studies 21, 1453–1454.

Stambaugh, R. F., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.

Stock, J. H., Watson, M. W., 2004. Combination forecasts of optimal growth in a seven-country data set. Journal of Forecasting 23, 405–430.

Taylor, S. J., 1986. Modeling Financial Time Series. John Wiley and Sons, Ltd.

Timmermann, A., 1993. How learning in financial markets generates excess volatility and predictability in stock prices. Quarterly Journal of Economics 108, 1135–1145.

Timmermann, A., 1996. Excess volatility and return predictability of stock returns in autoregressive dividend models with learning. Review of Economic Studies 63, 523–577.

Tourus, W., Valkanov, R., Yan, S., 2004. On predicting returns with nearly integrated explanatory variables. Journal of Business 77, 937–966.

Veronesi, P., 1999. Stock market overreaction to bad news in good times: a rational expectations equilibrium model. Review of Financial Studies 12, 975–1007.

West, K. D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.

Whitelaw, R. F., 1994. Time variations and covariations in the expectation and volatility of stock market returns. Journal of Finance 49, 515–541.

Panel A: Log volatility for the S&P 500 index

Panel B: Volatility and economic conditions: raw series

Panel C: Volatility and economic conditions: smoothed series
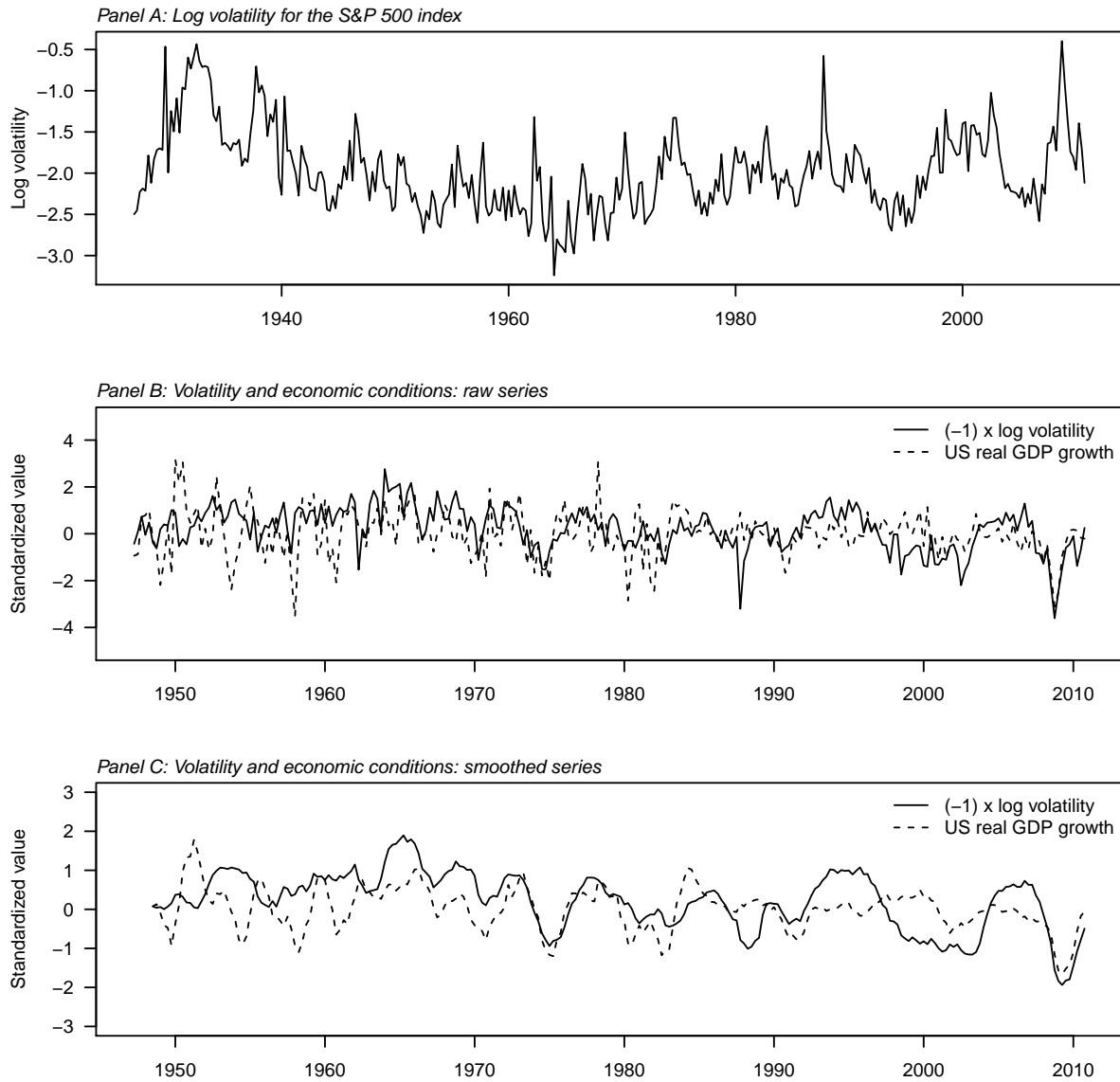
Fig. 1. This figures illustrates the relation between stock return volatility and the business cycle. The top panel presents a time series plot of log realized volatility on the Standard & Poor's (S&P) 500 index at the quarterly frequency for the period 1927–2010. The bottom two panels illustrate the covariation between stock return volatility and the business cycle. These panels contain quarterly time series plots of the opposite of standardized log realized volatility on the S&P 500 [i.e., (-1) x standardized log volatility], along with standardized US real gross domestic product (GDP) growth over the period 1947–2010. Panel B presents raw time series, and Panel C presents smoothed series constructed as a six quarter moving average.
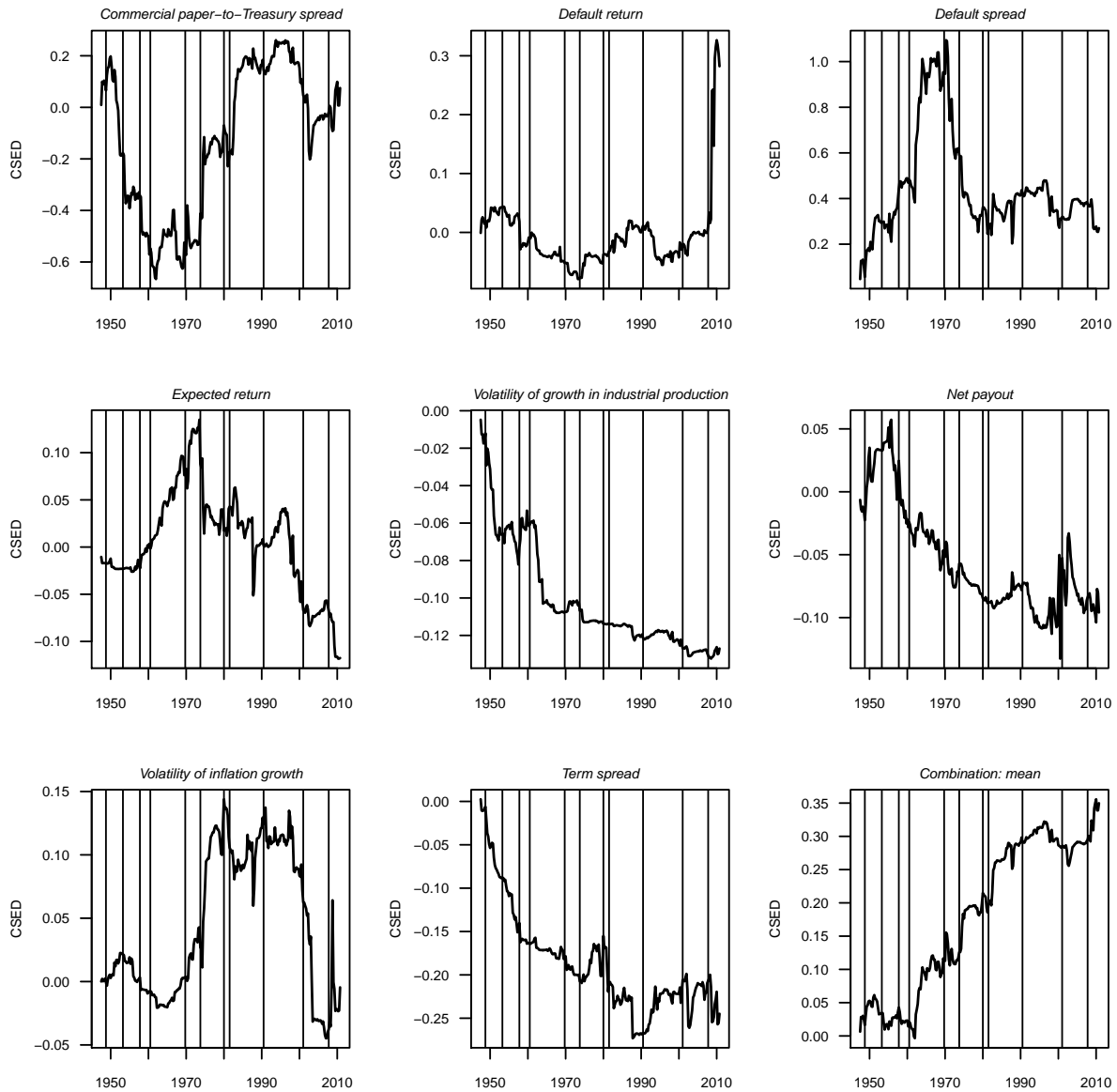
Fig. 2. This figures shows the relation between out-of-sample forecast performance and the business cycle. The figure shows time series plots of the cumulative squared error difference (CSED) between forecasts of log volatility based on the indicated macroeconomic variable(s) and a benchmark univariate AR(2) forecasting model. Vertical lines indicate business cycle peaks, i.e., the point at which an economic expansion transitions to a recession, based on National Bureau of Economic Research business cycle dating. The sample period is quarterly 1947Q3–2010Q4.

Table 1

Forecasting variables: descriptive statistics

The table presents descriptive statistics for the forecasting variables considered in the paper. The mean, standard deviation, skewness, and kurtosis are reported for each variable, as well the first- and second-order sample autocorrelations (ρ_1 and ρ_2). The final two columns of the table report the Z_t test statistic for the Phillips and Perron unit root test and the associated MacKinnon approximate p-value (see Phillips and Perron, 1998; and MacKinnon, 1994). Panel A presents statistics for variables sampled at a quarterly frequency over the period 1952Q2–2010Q4 and Panel B presents statistics for variables sampled at a monthly frequency over the period 1952.2–2010.12. CP = commercial paper; GDP = gross domestic product.

Symbol	Name	Mean	Standard deviation	Skewness	Kurtosis	ρ_1	ρ_2	Phillips and Perron test Z_t	p-value
Panel A: Quarterly sampling frequency									
blev	Changes in bank leverage	0.0072	0.1344	-0.68	4.88	-0.19	0.13	-18.95	0.00
cay	Consumption-wealth ratio	0.0001	0.0193	0.09	2.52	0.92	0.86	-2.75	0.07
cp	CP-to-Treasury spread	0.6461	0.4920	2.25	10.55	0.60	0.45	-8.04	0.00
dfr	Default return	-0.0002	0.0995	0.03	15.80	-0.02	0.06	-16.99	0.00
dfy	Default yield	0.0158	0.0072	1.41	7.02	0.85	0.73	-4.23	0.00
egdp	Expected GDP growth	2.5364	1.4331	-0.66	5.37	0.86	0.72	-3.87	0.00
exret	Expected return	0.0199	0.0201	0.90	3.78	0.78	0.67	-5.40	0.00
gdp	GDP growth	3.0439	3.8156	-0.38	4.34	0.37	0.19	-9.99	0.00
ik	Investment-capital ratio	0.0358	0.0036	0.27	2.43	0.96	0.89	-2.57	0.10
ipvol	Industrial production volatility	0.0045	0.0046	2.25	8.86	0.26	0.11	-12.71	0.00
npy	Net payout yield	-2.1916	0.2064	-1.63	7.23	0.94	0.87	-2.59	0.10
ppivol	Inflation volatility	0.0036	0.0046	4.36	33.26	0.42	0.28	-10.39	0.00
tms	Term spread	0.0160	0.0143	-0.11	3.00	0.83	0.69	-4.61	0.00
Panel B: Monthly sampling frequency									
cp	CP-to-Treasury spread	0.6147	0.4646	2.42	13.61	0.86	0.74	-7.40	0.00
dfr	Default return	-0.0091	0.2236	1.64	37.78	-0.12	-0.03	-30.61	0.00
dfy	Default yield	0.0157	0.0072	1.30	6.10	0.93	0.87	-4.92	0.00
exret	Expected return	0.0052	0.0040	1.19	5.87	0.89	0.82	-6.52	0.00
ip	Growth in industrial production	0.0024	0.0095	0.27	9.43	0.39	0.24	-18.35	0.00
ipvol	Industrial production volatility	0.0062	0.0064	3.31	23.01	0.24	0.14	-23.20	0.00
npy	Net payout	-2.1905	0.2069	-1.67	7.34	0.98	0.96	-2.75	0.07
ppivol	Inflation volatility	0.0044	0.0056	3.79	24.67	0.37	0.39	-24.17	0.00
tms	Term spread	0.0162	0.0142	-0.05	2.85	0.95	0.90	-4.19	0.00

Table 2

In-sample forecasting regressions with quarterly sampling

The table presents results for in-sample predictive regressions for volatility using various macroeconomic and financial forecasting variables. The table reports results from regressions of the form

$$LVOL_t = \alpha + \rho_1 LVOL_{t-1} + \rho_2 LVOL_{t-2} + \beta' X_{t-1} + \varepsilon_t,$$

where $LVOL_t$ is the natural logarithm of return volatility, measured as in Eq. (1), and X_{t-1} represents a (possibly vector-valued) forecasting variable. For each forecasting variable, the table presents the estimated slope coefficient $\hat{\beta}$ on the forecasting variable and the increase in R^2 value relative to a benchmark univariate AR(2) regression (the above model with $\beta = 0$). The table also reports for a kitchen sink regression that includes the full set of forecasting variables. In this case, the table displays the F-statistic testing the null hypothesis that (vector-valued) $\beta = 0$, along with the increase in R^2 value relative to a benchmark univariate AR(2) regression. Results are reported over various estimation samples as indicated above column headers. Dashes indicate estimation samples in which the corresponding forecasting variable is not available. All variables are standardized prior to estimation and, consequently, the intercept α is omitted from the specification. ***, **, and * designate statistical significance at the 1%, 5%, and 10% level, respectively, based on standard errors robust to the presence of heteroskedasticity and serial correlation. CP = commercial paper; GDP = gross domestic product.

Symbol	Name	1927Q2–2010Q4		1952Q2–2010Q4		1927Q2–1951Q4		1952Q2–1985Q4		1986Q1–2010Q4	
		$\hat{\beta}$	ΔR^2	$\hat{\beta}$	ΔR^2	$\hat{\beta}$	ΔR^2	$\hat{\beta}$	ΔR^2	$\hat{\beta}$	ΔR^2
blev	Changes in bank leverage	–	–	-0.03	0.09	–	–	-0.14**	1.98	0.08	0.60
cay	Consumption-wealth ratio	–	–	-0.05	0.26	–	–	-0.09	0.75	-0.09	0.77
cp	CP-to-Treasury spread	0.12***	1.36	0.11**	1.11	0.23**	2.77	0.31***	7.78	0.07	0.45
dfr	Default return	-0.08**	0.58	-0.13***	1.54	0.04	0.20	-0.12*	1.45	-0.14**	1.59
dfy	Default yield	0.17***	1.22	0.07	0.26	0.33***	3.46	0.11	0.79	0.04	0.08
egdp	Expected GDP growth	–	–	-0.02	0.06	–	–	0.00	0.00	-0.03	0.10
exret	Expected return	0.00	0.00	-0.08**	0.56	0.00	0.00	-0.08	0.57	-0.13**	1.65
gdp	GDP growth	–	–	0.00	0.00	–	–	0.00	0.00	-0.02	0.04
ik	Investment-capital ratio	–	–	0.12***	1.42	–	–	0.12**	1.48	0.15**	2.27
ipvol	Industrial production volatility	0.01	0.01	-0.02	0.05	-0.03	0.08	-0.04	0.14	0.09	0.76
npy	Net payout	-0.03	0.12	-0.11***	1.08	-0.06	0.34	0.00	0.00	-0.08	0.68
ppivol	Inflation Volatility	0.04	0.12	0.09	0.64	-0.03	0.09	0.18***	2.87	0.02	0.03
tms	Term spread	-0.03	0.08	-0.01	0.02	-0.05	0.21	-0.06	0.31	-0.08***	0.56
		F	ΔR^2	F	ΔR^2	F	ΔR^2	F	ΔR^2	F	ΔR^2
sink	Kitchen sink	5.29***	3.32	4.21***	6.76	2.34**	5.35	8.16***	11.56	7.63***	7.00

Table 3

In-sample forecasting regressions with monthly sampling

The table presents results for in-sample predictive regressions for volatility using various macroeconomic and financial forecasting variables. The table reports results from regressions of the form

$$LVOL_t = \alpha + \sum_{i=1}^{6} \rho_i LVOL_{t-i} + \beta' X_{t-1} + \varepsilon_t,$$

where $LVOL_t$ is the natural logarithm of return volatility, measured as in Eq. (1), and X_{t-1} represents a (possibly vector-valued) forecasting variable. For each forecasting variable, the table presents the estimated slope coefficient $\hat{\beta}$ on the forecasting variable and the increase in R^2 value relative to a benchmark univariate AR(6) regression (the above model with $\beta = 0$). The table also reports for a kitchen sink regression that includes the full set of forecasting variables. In this case, the table displays the F-statistic testing the null hypothesis that (vector-valued) $\beta = 0$, along with the increase in R^2 value relative to a benchmark univariate AR(6) regression. Results are reported over various estimation samples as indicated above column headers. Dashes indicate estimation samples in which the corresponding forecasting variable is not available. All variables are standardized prior to estimation and, consequently, the intercept α is omitted from the specification. ***, **, and * designate statistical significance at the 1%, 5%, and 10% level, respectively, based on standard errors robust to the presence of heteroskedasticity and serial correlation. CP = commercial paper.

Symbol	Name	1927.2–2010.12		1952.2–2010.12		1927.2–1951.12		1952.1–1985.12		1986.1–2010.12	
		$\hat{\beta}$	ΔR^2	$\hat{\beta}$	ΔR^2	$\hat{\beta}$	ΔR^2	$\hat{\beta}$	ΔR^2	$\hat{\beta}$	ΔR^2
cp	CP-to-Treasury spread	0.07***	0.47	0.06**	0.29	0.17***	1.40	0.16***	2.01	0.03	0.05
dfr	Default return	-0.05***	0.20	-0.06**	0.35	-0.02	0.06	-0.01	0.00	-0.11***	1.22
dfy	Default yield	0.10***	0.40	0.04	0.07	0.20***	1.27	0.03	0.05	0.09	0.32
exret	Expected return	-0.01	0.01	-0.04*	0.17	-0.01	0.01	-0.06**	0.29	-0.02	0.04
ip	Growth in industrial production	-0.01	0.02	-0.01	0.01	-0.03	0.10	0.01	0.01	-0.08	0.53
ipvol	Industrial production volatility	0.00	0.00	0.00	0.00	-0.02	0.04	-0.03	0.07	0.10	0.96
npy	Net payout	-0.02	0.03	-0.06**	0.29	-0.03	0.09	0.00	0.00	-0.05	0.22
ppivol	Inflation volatility	0.03	0.11	0.05*	0.26	0.01	0.02	0.07**	0.47	0.04	0.15
tms	Term spread	-0.02	0.02	-0.01	0.01	-0.01	0.01	-0.02	0.05	-0.05	0.27
		F	ΔR^2	F	ΔR^2	F	ΔR^2	F	ΔR^2	F	ΔR^2
sink	Kitchen sink	4.06***	1.22	6.29***	1.61	2.97***	2.26	4.12***	2.67	2.08**	3.07

Table 4

Out-of-sample results: quarterly sampling and rolling estimation

The table presents out-of-sample results for volatility forecasts that incorporate various macroeconomic and financial variables. The forecasting regressions take the form

$$LVOL_t = \alpha + \rho_1 LVOL_{t-1} + \rho_2 LVOL_{t-2} + \beta' X_{t-1} + \varepsilon_t,$$

where $LVOL_t$ is the natural logarithm of return volatility, measured as in Eq. (1), and X_{t-1} represents a (possibly vector-valued) forecasting variable. For each forecasting model, the table presents CW, the adjusted difference in mean square prediction error (MSPE) (multiplied by one thousand) that forms the basis of the out-of-sample test for equal MSPE suggested by Clark and West (2007). ***, **, and * appear alongside CW test statistics to indicate rejections of the null hypothesis of no Granger causality at significance levels of 1%, 5%, and 10%, respectively. The table also reports ΔR^2_{OOS}, the increase in out-of-sample R^2 relative to a benchmark univariate AR(2) forecasting model. ***, **, or * accompany ΔR^2_{OOS} when the Giacomini and White (2006) test for equal predictive ability rejects at the 1%, 5%, or 10% level, respectively. Forecasts are produced using a rolling estimation procedure with an estimation window of 80 quarters (20 years). See Section 5 of the paper for additional discussion of the Clark and West (2007) and Giacomini and White (2006) tests. Dashes indicate estimation samples in which the corresponding forecasting variable is not available. CP = commercial paper; GDP = gross domestic product.

Symbol	Name	1947Q3–2010Q4, N=254		1972Q3–2010Q4, N=154		1982Q3–2010Q4, N=114		1972Q3–2000Q4, N=114	
		CW	ΔR^2_{OOS}	CW	ΔR^2_{OOS}	CW	ΔR^2_{OOS}	CW	ΔR^2_{OOS}
blev	Changes in bank leverage	–	–	2.43*	0.13	-1.39	-1.52*	2.55	-0.01
cay	Consumption-wealth ratio	–	–	3.09	-0.70	-0.95	-1.22	4.72**	1.96
cp	CP-to-Treasury spread	7.77**	-0.71	15.29**	1.99	1.99	-2.45	21.86***	4.37
dfr	Default return	2.93**	0.37	3.98*	0.32	3.97	0.05	2.63	0.14
dfy	Default spread	-0.26	-1.21	-0.95	-1.54	0.25	-0.24	-0.27	-0.45
egdp	Expected GDP growth	–	–	-1.68	-1.30	1.03	0.36	-1.34	-1.29
exret	Expected return	-0.33	-1.22	3.32*	-0.15	1.78	-0.23	2.27	0.62
gdp	GDP growth	–	–	-0.51	-0.64	-0.86	-0.73	0.34	-0.26
ik	Investment-capital ratio	–	–	6.07**	1.30	5.69*	0.74	6.22**	2.02
ipvol	Industrial production volatility	-1.25	-0.83	-2.04	-1.41	-1.43	-0.90	-0.59	-0.51
npy	Net payout	-1.59	-1.54*	-1.91	-2.12	-0.15	-0.79	-1.82	-2.12
ppivol	Inflation volatility	4.65**	0.16	6.83*	-0.06	-2.79	-2.97***	10.32**	0.88
tms	Term spread	2.64	-0.66	1.43	-1.22	-1.77	-1.46	3.05	-0.46
sink	Kitchen sink	5.71	-7.89**	5.58	-15.79**	-2.34	-6.90**	19.69**	-7.94
Combined forecasts									
Mean		1.82**	0.62	2.72***	1.25*	0.41	0.10	3.84***	2.16**
Median		0.77	0.26	0.98**	0.48	0.04	-0.03	1.31***	0.75**
Trim-mean		1.17*	0.41	1.75**	0.81	0.12	-0.02	2.46***	1.39**
MSPE		2.05**	0.73	3.66***	1.69*	0.91	0.34	5.08***	2.84***

Table 5

Out-of-sample results: quarterly sampling and recursive estimation

The table presents out-of-sample results for volatility forecasts that incorporate various macroeconomic and financial variables. The forecasting regressions take the form

$$LVOL_t = \alpha + \rho_1 LVOL_{t-1} + \rho_2 LVOL_{t-2} + \beta' X_{t-1} + \varepsilon_t,$$

where $LVOL_t$ is the natural logarithm of return volatility, measured as in Eq. (1), and X_{t-1} represents a (possibly vector-valued) forecasting variable. For each forecasting model, the table presents CW, the adjusted difference in mean square prediction error (MSPE) (multiplied by one thousand) that forms the basis of the out-of-sample test for equal MSPE suggested by Clark and West (2007). ***, ** and * appear alongside CW test statistics to indicate rejections of the null hypothesis of no Granger causality at significance levels of 1%, 5%, and 10%, respectively. The table also reports ΔR^2_{OOS}, the increase in out-of-sample R^2 relative to a benchmark univariate AR(2) forecasting model. ***, ** or * accompany ΔR^2_{OOS} when the Giacomini and White (2006) test for equal predictive ability rejects at the 1%, 5%, or 10% level, respectively. Forecasts are produced using a recursive estimation procedure with an initial estimation window of 80 quarters (20 years). See Section 5 of the paper for additional discussion of the Clark and West (2007) and Giacomini and White (2006) tests. Dashes indicate estimation samples in which the corresponding forecasting variable is not available. CP = commercial paper; GDP = gross domestic product.

Symbol	Name	1947Q3–2010Q4, N=254		1972Q3–2010Q4, N=154		1982Q3–2010Q4, N=114		1972Q3–2000Q4, N=114	
		CW	ΔR^2_{OOS}	CW	ΔR^2_{OOS}	CW	ΔR^2_{OOS}	CW	ΔR^2_{OOS}
blev	Changes in bank leverage	–	–	1.03	-0.66	-2.62	-1.90**	2.58	-0.30
cay	Consumption-wealth ratio	–	–	0.66	-0.06	-0.69	-0.60	0.38	-0.32
cp	CP-to-Treasury spread	5.90**	0.13	11.35**	0.60	-0.01	-2.39	16.87**	2.56
dfr	Default return	1.92*	0.48	6.77**	1.21	6.81*	0.76	3.28*	0.38
dfy	Default spread	6.02**	0.46	0.60	0.04	0.71	0.04	0.29	-0.10
egdp	Expected GDP growth	–	–	-0.26	-0.14	0.15	0.06	-0.32	-0.20
exret	Expected return	-0.17	-0.20	2.01**	0.49	1.69	0.26	1.22	0.36
gdp	GDP growth	–	–	-0.55	-0.30	-0.27	-0.12	-0.56	-0.39
ik	Investment-capital ratio	–	–	5.82***	1.56	5.13**	0.95	7.50***	3.11**
ipvol	Industrial production volatility	-0.47	-0.21***	-0.49	-0.32	-0.70	-0.40	0.62	0.27
npy	Net payout	-0.04	-0.16	3.19	0.22	5.45*	0.94	2.56	0.30
ppivol	Inflation volatility	0.42	-0.01	7.39**	-0.50	0.10	-2.15*	8.84*	-0.05
tms	Term spread	-0.70	-0.41*	0.43	-0.71	-2.33	-1.20*	1.26	-0.75
sink	Kitchen sink	11.91***	0.77	21.08***	1.25	12.59***	0.59	23.06**	0.90
	Combined forecasts								
	Mean	1.61***	0.59**	2.92***	1.20**	1.03**	0.36	3.42***	1.72*
	Median	-0.03	-0.03	0.66*	0.27	0.10	0.01	0.94**	0.48*
	Trim-mean	0.66**	0.24*	1.79***	0.74*	0.44	0.13	2.21***	1.12*
	MSPE	1.86***	0.69***	3.77***	1.53**	1.56**	0.57*	4.55***	2.27*

37

Table 6

Out-of-sample results: monthly sampling and rolling estimation

The table presents out-of-sample results for volatility forecasts that incorporate various macroeconomic and financial variables. The forecasting regressions take the form

$$LVOL_t = \alpha + \sum_{i=1}^{6} \rho_i LVOL_{t-i} + \beta' X_{t-1} + \varepsilon_t,$$

where $LVOL_t$ is the natural logarithm of return volatility, measured as in Eq. (1), and X_{t-1} represents a (possibly vector-valued) forecasting variable. For each forecasting model, the table presents CW, the adjusted difference in mean square prediction error (MSPE) (multiplied by one thousand) that forms the basis of the out-of-sample test for equal MSPE suggested by Clark and West (2007). ***, **, and * appear alongside CW test statistics to indicate rejections of the null hypothesis of no Granger causality at significance levels of 1%, 5%, and 10%, respectively. The table also reports ΔR^2_{OOS}, the increase in out-of-sample R^2 relative to a benchmark univariate AR(6) forecasting model. ***, **, or * accompany ΔR^2_{OOS} when the Giacomini and White (2006) test for equal predictive ability rejects at the 1%, 5%, or 10% level, respectively. Forecasts are produced using a rolling estimation procedure with an estimation window of 240 months (20 years). See Section 5 of the paper for additional discussion of the Clark and West (2007) and Giacomini and White (2006) tests. CP = commercial paper.

Symbol	Name	1947.3–2010.12, N=766		1972.3–2010.12, N=466		1982.3–2010.12, N=346		1972.3–2000.12, N=346	
		CW	ΔR^2_{OOS}	CW	ΔR^2_{OOS}	CW	ΔR^2_{OOS}	CW	ΔR^2_{OOS}
cp	CP-to-Treasury spread	2.69**	-0.25	5.05***	0.39	0.71	-0.70	7.70***	1.21
dfr	Default return	2.35**	-0.36	4.12**	-0.48	2.95***	0.47	0.07	-0.02
dfy	Default spread	0.66	-0.20	0.72	-0.25	0.09	-0.04	-0.29	-0.23
exret	Expected return	0.61	-0.22	0.62	-0.50	0.43	-0.17	1.19	-0.48
ip	Growth in industrial production	-0.10	-0.15	0.22	-0.04	0.15	0.00	-0.28	-0.18**
ipvol	Industrial production volatility	0.79	0.17	1.43	0.37	0.75	0.24	0.12	0.00
npy	Net payout	-0.45	-0.45	-0.56	-0.65	-0.19	-0.33	-0.34	-0.52
ppivol	Inflation volatility	0.70	-0.07	0.91	-0.15	0.52	-0.54	1.48*	0.08
tms	Term spread	1.02*	-0.13	0.44	-0.24	-0.20	-0.26	0.87*	0.12
sink	Kitchen sink	5.61***	-1.44	10.42***	-0.77	4.33**	-0.55	9.02***	0.46
Combined forecasts									
	Mean	0.92***	0.33**	1.44***	0.56**	0.58**	0.22	1.17***	0.53*
	Median	0.09	0.02	0.20	0.07	0.45***	0.19**	0.22	0.10
	Trim-mean	0.52**	0.19	0.88**	0.35*	0.48**	0.20	0.62**	0.28*
	MSPE	0.95***	0.34**	1.61***	0.63*	0.65**	0.25*	1.38***	0.62*

Table 7

Out-of-sample tests for predictive ability: simulation analysis

The table reports the percentage of simulations in which the Clark and West (CW) and Giacomini and White (GW) tests reject their respective null hypotheses at the 10% level. Data are simulated from the process described in Eq. (13). Forecasts of Y_t from a benchmark AR(2) model are compared against those from an alternative model that also includes the lagged value of the simulated variable X_t. Panel A corresponds to simulations under the alternative, corresponding to $\beta = 0.09$ in Eq. (13). Panel B corresponds to the case of no Granger causality, i.e., $\beta = 0$ in Eq. (13). Results are presented for a range of estimation sample sizes (R). The out-of-sample period is alternatively set to $P = 120$ (30 years of quarterly data) or $P = 240$ (60 years of quarterly data). Columns labeled CW present the percentage of simulations in which the CW test rejects, and columns labeled GW present the percentage of simulations in which the GW test rejects (both at the 10% level). Columns labeled $CW \cap (\neg GW)$ present the percentage of simulations in which the CW test rejects and the GW test fails to reject. For reference, the table also presents a simulation-based estimate of ΔR^2_{OOS}, a measure of the difference in predictive ability between the two forecasting models. The simulation size is ten thousand.

| | P = 120 | | | P = 240 | | | |
R	CW	GW	$CW \cap (\neg GW)$	CW	GW	$CW \cap (\neg GW)$	ΔR^2_{OOS}
Panel A: Simulation under the alternative							
20	32.16	26.19	31.93	44.34	37.32	43.33	-3.50
30	33.85	15.58	33.19	47.59	19.08	47.41	-1.45
40	36.23	11.03	35.03	51.33	11.64	50.80	-0.65
60	39.59	8.81	37.13	55.92	7.25	54.05	0.01
80	41.82	8.89	37.68	59.91	6.96	56.22	0.31
120	47.31	10.01	40.83	64.63	9.17	57.48	0.56
160	49.28	11.74	39.88	67.56	11.14	57.91	0.67
400	56.90	16.59	41.29	77.29	21.60	56.11	0.86
1,000	60.04	19.73	41.14	81.18	27.35	54.06	0.94
Panel B: Simulation under the null of no Granger causality							
20	11.68	44.03	11.66	14.05	70.04	13.46	-5.44
30	9.60	33.15	9.55	10.78	52.54	10.74	-2.97
40	8.33	25.18	8.30	9.27	42.71	9.27	-2.03
60	7.21	19.96	7.08	7.54	30.20	7.54	-1.20
80	6.89	17.24	6.64	7.00	24.06	6.98	-0.83
120	6.79	14.55	6.32	6.22	18.98	6.12	-0.53
160	6.69	13.44	6.03	5.91	16.73	5.81	-0.39
400	7.41	13.18	5.53	6.31	13.51	5.41	-0.14
1,000	8.42	12.01	5.71	7.13	12.54	5.26	-0.05

Table 8

Forecast performance and the business cycle

This table reports results for regressions of squared forecast error differences on a measure of business conditions at the quarterly frequency. The regressions are of the form

$$\Delta SE_{i,t} = \alpha + \beta \, gdp_t + \varepsilon_t.$$

The dependent variable $\Delta SE_{i,t}$ is the time series of differences in squared forecast error between a forecast produced using a benchmark univariate AR(2) model for log volatility and a forecast produced using using an AR(2) model augmented with the corresponding predictive variable(s). The regressor gdp is the growth in quarterly gross domestic product (GDP). The table reports the estimated coefficient $\hat{\beta}$ as well as the t-statistic associated with this coefficient based on standard errors corrected for heteroskedasticity and serial correlation. The table also presents the R^2 value for the regression (as a percentage). Variables are standardized prior to running the regression and, consequently, the intercept α is omitted. In the left-hand portion of the table, the underlying forecasts are based on an expanding estimation sample beginning in 1927Q3 and covering the evaluation period 1947Q3–2010Q4. In the right-hand portion of the table, the underlying forecasts are based on an expanding estimation sample beginning in 1952Q3 and covering the evaluation period 1972Q3–2010Q4. Dashes indicate estimation samples in which the corresponding forecasting variable is not available. ***, ** and * designate statistical significance at the 1%, 5%, and 10% level, respectively. CP = commercial paper; GDP = gross domestic product.

Symbol	Name	Estimation 1927Q3 onward, evaluation 1947Q3–2010Q4			Estimation 1952Q3 onward, evaluation 1972Q3–2010Q4		
		$\hat{\beta}$	t-statistic	R^2	$\hat{\beta}$	t-statistic	R^2
blev	Changes in bank leverage	–	–	–	0.14	1.18	1.41
cay	Consumption-wealth ratio	–	–	–	-0.03	-0.28	0.05
cp	CP-to-Treasury spread	-0.02	-0.20	0.03	-0.28*	-1.97	5.87
dfr	Default return	-0.23	-1.07	3.41	-0.34	-1.31	8.48
dfy	Default yield	-0.05	-0.56	0.19	-0.23*	-1.78	3.82
egdp	Expected GDP growth	–	–	–	-0.01	-0.11	0.02
exret	Expected return	-0.08	-0.77	0.39	-0.07	-0.56	0.32
gdp	GDP growth	–	–	–	0.19	1.41	2.60
ik	Investment-capital ratio	–	–	–	0.11	1.15	0.97
ipvol	Industrial production volatility	-0.08	-1.47	1.26	0.37	1.26	9.94
npy	Net payout	0.06	0.91	0.25	0.31**	2.33	6.93
ppivol	Inflation volatility	-0.14	-0.82	1.34	-0.20	-1.06	2.94
tms	Term spread	0.13	1.18	1.12	0.00	-0.02	0.00
sink	Kitchen sink	-0.17*	-1.75	2.14	-0.33**	-2.44	7.96
Combination (mean)		-0.18*	-1.81	2.32	-0.24*	-1.77	4.36