Forecasting Stock Returns

Forthcoming in the *Handbook of Economic Forecasting*, Volume 2

David E. Rapach[*]
Saint Louis University
rapachde@slu.edu

Guofu Zhou[†]
Washington University in St. Louis
zhou@wustl.edu

July 1, 2012

Abstract

We survey the literature on stock return forecasting, highlighting the challenges faced by forecasters as well as strategies for improving return forecasts. We focus on U.S. equity premium forecastability and illustrate key issues via an empirical application based on updated data. Some studies argue that, despite extensive in-sample evidence of equity premium predictability, popular predictors from the literature fail to outperform the simple historical average benchmark forecast in out-of-sample tests. Recent studies, however, provide improved forecasting strategies that deliver statistically and economically significant out-of-sample gains relative to the historical average benchmark. These strategies—including economically motivated model restrictions, forecast combination, diffusion indices, and regime shifts—improve forecasting performance by addressing the substantial *model uncertainty* and *parameter instability* surrounding the data-generating process for stock returns. In addition to the U.S. equity premium, we succinctly survey out-of-sample evidence supporting U.S. cross-sectional and international stock return forecastability. The significant evidence of stock return forecastability worldwide has important implications for the development of both asset pricing models and investment management strategies.

JEL classification: C22, C38, C53, C58, E32, G11, G12, G14, G17

Keywords: Equity premium; Economic variables; Technical indicators; Forecast combination; Diffusion index; Regime shifts; Asset pricing model; Asset allocation; Business cycle

---

[*]Contact information for Rapach: Department of Economics, Saint Louis University, 3674 Lindell Boulevard, St. Louis, MO 63108; e-mail: rapachde@slu.edu; phone: 314-977-3601.

[†]Contact information for Zhou: Olin School of Business, Washington University in St. Louis, St. Louis, MO 63130; e-mail: zhou@wustl.edu; phone: 314-935-6384. We are grateful to Graham Elliott and Allan Timmermann (Editors), two anonymous referees, Massimo Guidolin, Lutz Kilian, Michael McCracken, Ľuboš Pástor, and participants at the *Handbook of Economic Forecasting* conference hosted by the Federal Reserve Bank of St. Louis for very helpful comments.

1 Introduction

Forecasting stock returns is a fascinating endeavor with a long history. From the standpoint of practitioners in finance, asset allocation requires real-time forecasts of stock returns, and improved stock return forecasts hold the promise of enhancing investment performance. It is thus not surprising that finance practitioners employ a plethora of variables in an attempt to forecast stock returns. Academics in finance are also keenly interested in stock return forecasts, since the ability to forecast returns has important implications for tests of market efficiency; more generally, understanding the nature of stock return forecastability helps researchers to produce more realistic asset pricing models that better explain the data.

While stock return forecasting is fascinating, it can also be frustrating. Stock returns inherently contain a sizable unpredictable component, so that the best forecasting models can explain only a relatively small part of stock returns. Furthermore, competition among traders implies that once successful forecasting models are discovered, they will be readily adopted by others; the widespread adoption of successful forecasting models can then cause stock prices to move in a manner that eliminates the models' forecasting ability (e.g., Lo, 2004; Timmermann and Granger, 2004; Timmermann, 2008). However, rational asset pricing theory posits that stock return predictability can result from exposure to time-varying aggregate risk, and to the extent that successful forecasting models consistently capture this time-varying aggregate risk premium, they will likely remain successful over time.

Along this line, there is a common misconception that stock return predictability is contrary to market efficiency. The canonical random walk model, used historically by many modelers and popularized to the broader public by Malkiel (1973, 2011), implies that future stock returns are unpredictable on the basis of currently available information. While the random walk model is consistent with market efficiency, so is a predictable return process, insofar as predictability is consistent with exposure to time-varying aggregate risk. According to the well-known Campbell and Shiller (1988a) present-value decomposition, deviations in the dividend-price ratio from its long-term mean signal changes in expected future dividend growth rates and/or expected future stock returns; changes in the latter represent time-varying discount rates and return predictability.[1] In this context, fluctuations in aggregate risk exposure that produce time-varying discount rates and return predictability are entirely compatible with market efficiency. In other words, only when the risk-adjusted time-varying expected return—after further adjusting for transaction costs and other trading frictions (e.g., liquidity and borrowing constraints, research costs)—is nonzero can we say that a market is inefficient.

Theoretically, asset returns are functions of the state variables of the real economy, and the real economy itself displays significant business-cycle fluctuations. If the quantity and price of aggregate risk are linked to economic fluctuations, then we should expect time-varying expected returns and return predictability, even in an efficient market. For instance, if agents become more risk averse during economic contractions when consumption and income levels are depressed, then they will require a higher expected return on stocks near business-cycle troughs to be willing to take on the risk associated with holding stocks; variables that measure and/or predict the state of the economy should thus help to predict returns (e.g., Fama and French, 1989; Campbell and Cochrane, 1999; Cochrane, 2007, 2011). Indeed, an important theme of this chapter is that stock return predictability is closely tied to business-cycle fluctuations.

While rational asset pricing is consistent with return predictability, theory does impose certain bounds on the maximum degree of return predictability that is consistent with asset pricing models. To the extent that return predictability exceeds these bounds, this can be interpreted as evidence for mispricing or market inefficiencies stemming from, for example, information processing limitations and/or the types of psychological influences emphasized in behavioral finance. Since information processing limitations and psychological influences are likely to be exacerbated during rapidly changing economic conditions, return predictability resulting from market inefficiencies is also likely linked to business-cycle fluctuations.

The degree of stock return predictability is ultimately an empirical issue. There is ample in-sample evidence that U.S. aggregate stock market returns are predictable using a variety of

[1] See Koijen and van Nieuwerburgh (2011) for a survey on the relationship between the present-value identity and stock return predictability.

economic variables. While thorny econometric issues complicate statistical inference, such as the well-known Stambaugh (1986, 1999) bias, there is an apparent consensus among financial economists that, on the basis of in-sample tests, stock returns contain a significant predictable component (Campbell, 2000). Bossaerts and Hillion (1999) and Goyal and Welch (2003, 2008), however, find that the predictive ability of a variety of popular economic variables from the literature does not hold up in *out-of-sample* forecasting exercises, a finding eerily reminiscent of Meese and Rogoff (1983) in the context of exchange rate predictability. Under the widely held view that predictive models require out-of-sample validation,[2] this finding casts doubt on the reliability of stock return predictability.

The unreliability of return predictability, however, can result from the econometric methods themselves; as pointed out by Lamoureux and Zhou (1996), many studies of return predictability implicitly use a strong prior on predictability. Moreover, in highlighting the importance of out-of-sample tests for analyzing return predictability, Pesaran and Timmermann (1995) demonstrate the relevance of *model uncertainty* and *parameter instability* for stock return forecasting. Model uncertainty recognizes that a forecaster knows neither the "best" model specification nor its corresponding parameter values. Furthermore, due to parameter instability, the best model can change over time. Given the connection between business-cycle fluctuations and stock return predictability, it is not surprising that model uncertainty and parameter instability are highly relevant for stock return forecasting, as these factors are also germane to macroeconomic forecasting.

The substantial model uncertainty and parameter instability surrounding the data-generating process for stock returns render out-of-sample return predictability challenging to uncover. Fortunately, recent studies provide forecasting strategies that deliver statistically and economically significant out-of-sample gains, including strategies based on:

- *economically motivated model restrictions* (e.g., Campbell and Thompson, 2008; Ferreira and Santa-Clara, 2011);

[2] Campbell (2008, p. 3) succinctly expresses the prevailing view: "The ultimate test of any predictive model is its out-of-sample performance."

- *forecast combination* (e.g., Rapach et al., 2010);
- *diffusion indices* (e.g., Ludvigson and Ng, 2007; Kelly and Pruitt, 2012; Neely et al., 2012);
- *regime shifts* (e.g., Guidolin and Timmermann, 2007; Henkel et al., 2011; Dangl and Halling, 2012).

These forecasting strategies produce out-of-sample gains largely by accommodating model uncertainty and parameter instability. We focus on these strategies in this chapter, discussing their implementation and how they improve stock return forecasts.

We illustrate relevant issues concerning the formation and evaluation of stock return forecasts via an empirical application based on forecasting the U.S. equity premium with updated data. In addition to significantly outperforming the simple historical average benchmark forecast according to the conventional mean squared forecast error (MSFE) statistical criterion, we show that the forecasting approaches cited above generate sizable utility gains from an asset allocation perspective.

Although we present the leading themes of this chapter in the context of forecasting the U.S. equity premium, we also survey the evidence on stock return forecastability in international markets. Ang and Bekaert (2007) provide compelling in-sample evidence of stock return predictability for France, Germany, the United Kingdom, and the United States. We discuss recent studies that furnish out-of-sample evidence of return predictability for these and other industrialized countries. In addition, we survey studies that explore out-of-sample stock return predictability along cross-sectional dimensions, including portfolios sorted by size, book-to-market value, and industry.[3]

While stock return forecasting will always be extremely challenging—and we will likely never explain more than a small part of returns—the take-away message of this chapter is that methods are available for reliably improving stock return forecasts in an economically meaningful manner. Consequently, investors who account for stock return predictability with available forecasting procedures significantly outperform those who treat returns as entirely unpredictable.

[3]While not focusing on forecasting *per se*, Hawawini and Keim (1995) and Subrahmanyam (2010) provide informative surveys on international and cross-sectional stock return predictability.

The remainder of this chapter is organized as follows. Section 2 provides a benchmark by discussing the degree of stock return predictability that we should expect theoretically in a predictive regression framework. Section 3 surveys the literature on U.S. aggregate stock market return forecastability; this section also contains the empirical application based on forecasting the U.S. equity premium. Section 4 discusses the evidence on stock return forecastability in international and cross-sectional contexts. Section 5 concludes.

2 What Level of Predictability Should We Expect?

Stock return predictability is typically examined via the following predictive regression model:

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}, \tag{1}$$

where r_{t+1} is the return on a broad stock market index in excess of the risk-free interest rate (or equity premium or excess stock return) from the end of period t to the end of period $t+1$, x_t is a variable available at the end of t used to predict the equity premium (such as the dividend-price ratio), and ε_{t+1} is a zero-mean disturbance term. Given an asset pricing model, how much predictability should we expect in (1)?

An asset pricing model is almost always uniquely determined by its stochastic discount factor (SDF, or state-price density or pricing kernel), which is a random variable, m_{t+1}, that satisfies

$$E(R_{j,t+1} m_{t+1} | I_t) = 1, \quad j = 1, \ldots, N, \tag{2}$$

where $R_{j,t+1}$ is the gross return on asset j and I_t is the information available at t. A particular asset pricing model entails a specification of m_{t+1} (e.g., Cochrane, 2005).[4] Assume that the risk-free rate is constant; this assumption is not a problem, since the risk-free rate changes little relative to

[4]The SDF corresponds to the representative investor's intertemporal marginal rate of substitution in consumption-based asset pricing models.

stock returns. Ross (2005, p. 56) shows that the regression R^2 in (1) has an elegant upper bound:

$$R^2 \leq R_f^2 \mathrm{Var}(m_t), \tag{3}$$

where R_f is the (gross) risk-free rate and m_t is the SDF for a given asset pricing model. Following Ross (2005) by using an annualized risk-free rate of 3.5%, an annualized standard deviation of 20% for the U.S. aggregate stock market, and an upper bound on market risk aversion equaling five times the observed market risk aversion on asset returns, the R^2 bound in (3) is approximately 8% for monthly returns. This bound, however, is too loose to be binding in applications. For example, in their well-known study, Fama and French (1988) report monthly R^2 statistics of 1% or less for predictive regression models based on dividend-price ratios, and Zhou (2010) reports monthly R^2 statistics of less than 1% for individual predictive regressions based on 10 popular economic variables from the literature.

Incorporating insights from Kan and Zhou (2007), Zhou (2010) refines the Ross (2005) R^2 bound so that it is more relevant for empirical research. The default SDF, which, by construction, also satisfies (2) and thus prices the N risky assets, is given by

$$m_{0,t} = \mu_m + (1_N - \mu_m \mu)' \Sigma^{-1} (R_t - \mu), \tag{4}$$

where R_t is the vector of gross returns on the N risky assets, μ (Σ) is the mean vector (covariance matrix) for R_t, $E(m_t) = \mu_m$, and 1_N is an N-vector of ones.[5] Let $z_t = (z_{1,t}, \ldots, z_{K,t})'$ denote a K-vector of state variables corresponding to a specific asset pricing model and consider the following linear regression model:

$$m_{0,t} = a + b' z_t + e_t, \tag{5}$$

where b is a K-vector of slope coefficients. For a linear regression model, we have $E(e_t) = 0$ and $\mathrm{Cov}(e_t, z_t) = 0$, by construction. Under the slightly stronger assumption that $E(e_t | z_t) = 0$,[6] Zhou

[5] We assume that μ is not proportional to 1_N and that the N risky assets are not redundant (so that Σ is nonsingular).

[6] This assumption does not appear to be too restrictive. A sufficient condition for $E(e_t | z_t) = 0$ is that the returns

(2010) proves the following result, which tightens the Ross (2005) bound:

$$R^2 \leq \rho^2_{z,m_0} R^2_f \mathrm{Var}[m_t(z_t)], \tag{6}$$

where ρ_{z,m_0} is the multiple correlation between the state variables, z_t, and the default SDF, $m_{0,t}$. This R^2 bound thus improves the Ross (2005) bound by a factor of ρ^2_{z,m_0}. Since ρ_{z,m_0} is typically small in practice, the reduction in the R^2 bound is substantial. For example, Kan and Zhou (2007) report ρ_{z,m_0} values ranging from 0.10 to 0.15 using either the aggregate market portfolio or 25 Fama-French size/value-sorted portfolios for R_t in (4) and consumption growth and the surplus consumption ratio as the state variables in (5), where the latter correspond to the state variables for the well-known Campbell and Cochrane (1999) consumption-based asset pricing model with habit formation. For these ρ_{z,m_0} values and Ross's (2005) assumptions, the R^2 bound ranges from 0.08% to 0.18%. These values for the R^2 bound are so low that predictive regressions based on a number of popular predictors—even though they have monthly R^2 statistics of 1% or less—actually violate the theoretical bounds implied by the Campbell and Cochrane (1999) model.

The upshot of this analysis is that, from the perspective of asset pricing models, we should expect only a limited degree of stock return predictability. Indeed, a seemingly "small" monthly R^2 of 1% or less can nevertheless signal "too much" return predictability and the existence of market inefficiencies from the standpoint of existing asset pricing models. Predictive models that claim to explain a large part of stock return fluctuations imply either that existing asset pricing models are grossly incorrect or that massive market inefficiencies exist (along with substantial risk-adjusted abnormal returns); both of these implications appear unlikely. In general, such high return predictability is simply too good to be true and should be viewed with appropriate suspicion. This does not mean that we should throw up our hands and give up on stock return forecasting; instead, it means that the ability to forecast even a seemingly small part of stock return fluctuations on a reasonably consistent basis is no mean feat.

and state variables are jointly elliptically distributed. The t-distribution is a special case of the elliptical distribution, and Tu and Zhou (2004) show that the t-distribution fits return data well.

Furthermore, as shown asymptotically by Inoue and Kilian (2004) and in Monte Carlo simulations by Cochrane (2008), if our ultimate aim is to determine whether we can reject the null hypothesis that the population parameter $\beta \neq 0$ in (1), in-sample tests will typically be more powerful than out-of-sample tests. Intuitively, unlike in-sample tests, out-of-sample tests cannot utilize the entire available sample when estimating the parameters used to generate return predictions; in-sample tests inherently make more observations available for estimating parameters, thereby increasing estimation efficiency and test power.[7] The greater number of available observations for in-sample parameter estimation can also lead to higher in-sample R^2 statistics relative to out-of-sample R^2 statistics (Campbell and Thompson, 2008). The monthly R^2 statistics in the neighborhood of 1% cited previously in the context of the R^2 bound refer to in-sample estimation of predictive regressions. Out-of-sample R^2 statistics will frequently be even lower, reiterating the fundamental notion that we should expect only a limited degree of stock return forecastability.

Although we should expect a limited degree of stock return forecastability, it is important to realize that a little goes a long way. That is, even an apparently small degree of return predictability can translate into substantial utility gains for a risk-averse investor who does not affect market prices (e.g., Kandel and Stambaugh, 1996; Xu, 2004; Campbell and Thompson, 2008). We illustrate this for U.S. equity premium forecasts in Section 3.

3 U.S. Aggregate Stock Market Return Forecastability

This section first provides an overview of the academic literature on aggregate U.S. stock market return forecastability, highlighting recently proposed strategies for significantly improving stock return forecasts. After discussing key issues involving the evaluation of stock return forecasts, it presents the empirical application based on forecasting the U.S. equity premium using updated

[7] If we are interested in testing the null hypothesis that return predictability exists at some point in the sample, Clark and McCracken (2005) show that out-of-sample tests can be more powerful than conventional in-sample tests in the presence of structural breaks near the end of the sample; also see Rossi (2005). We discuss the importance of accounting for structural breaks when forecasting stock returns in Section 3. See the chapter by Barbara Rossi in this volume for a detailed survey of recent advances in forecasting in the presence of structural breaks.

data. The empirical application illustrates a number of relevant issues regarding the formation and evaluation of stock return forecasts.

3.1 Overview of Return Forecastability

Academic investigation of U.S. aggregate stock market return forecastability begins with Cowles's (1933) seminal paper, "Can Stock Market Forecasters Forecast?" Cowles (1933) constructs a portfolio using Dow Jones industrial average (DJIA) market index forecasts made in editorials by William Peter Hamilton during his 26-year (1903–1929) tenure as editor of the *Wall Street Journal*. Hamilton bases his forecasts on Charles Dow's then-popular Dow Theory—a type of technical analysis—of which Hamilton was a strong proponent. The portfolio based on Hamilton's forecasts generates a lower annualized average return than a buy-and-hold portfolio based on the DJIA (12% and 15.5%, respectively). Similarly, Cowles (1933) finds that portfolios based on broad market recommendations from 24 individual financial publications for 1928–1932 fail to outperform a passive investment in the DJIA by 4% on average on an annualized basis. Cowles (1933) also concludes that the performances of the most successful of the 24 individual portfolios do not differ substantially from what would be expected from pure chance.[8] In sum, Cowles (1933) answers the question posed in his title in the negative.

A spate of prominent studies during the 1960s examines the forecasting power of various technical indicators, including popular filter rules, moving averages, and momentum oscillators. Technical indicators seek to provide insights into future returns on the basis of patterns in past prices. Analyzing a variety of broad market indices, Alexander (1961, 1964) reports that portfolios based on certain filter rules earn higher returns than buy-and-hold portfolios that invest in the indices. However, after switching the focus from broad market indices to individual stocks, studies by Cootner (1962), Fama and Blume (1966), and Jensen and Bennington (1970), among others, present evidence that portfolios constructed from filter rules (and technical indicators more gener-

[8] In an early—and labor-intensive—Monte Carlo experiment, Cowles (1933) simulates random recommendations by randomly drawing numbered cards to reach this conclusion.

ally) frequently fail to outperform buy-and-hold portfolios, especially after accounting for transaction costs. These negative findings proved influential in the ascendancy of the efficient market hypothesis—in the form of the random walk model—among academic researchers, as epitomized by Fama's (1970) influential survey and the first edition of Malkiel's (1973) book, *A Random Walk Down Wall Street*.[9]

Starting in the late 1970s, a vast literature compiles evidence that numerous economic variables predict monthly, quarterly, and/or annual U.S. aggregate stock returns in the predictive regression framework given by (1). The most popular predictor in this literature is the dividend-price ratio (Rozeff, 1984; Campbell and Shiller, 1988a, 1998; Fama and French, 1988, 1989; Cochrane, 2008; Lettau and Van Nieuwerburgh, 2008; Pástor and Stambaugh, 2009). Other economic variables that evince predictive ability include the earnings-price ratio (Campbell and Shiller, 1988b, 1998), book-to-market ratio (Kothari and Shanken, 1997; Pontiff and Schall, 1998), nominal interest rates (Fama and Schwert, 1977; Breen et al., 1989; Ang and Bekaert, 2007), interest rate spreads (Campbell, 1987; Fama and French, 1989), inflation (Nelson, 1976; Campbell and Vuolteenaho, 2004), dividend payout ratio (Lamont, 1998), corporate issuing activity (Baker and Wurgler, 2000; Boudoukh et al., 2007), consumption-wealth ratio (Lettau and Ludvigson, 2001), stock market volatility (Guo, 2006), labor income (Santos and Veronesi, 2006), aggregate output (Rangvid, 2006), output gap (Cooper and Priestly, 2009), expected business conditions (Campbell and Diebold, 2009), oil prices (Driesprong et al., 2008), lagged industry portfolio returns (Hong et al., 2007), and accruals (Hirshleifer et al., 2009).[10] Fama's (1991) sequel survey reflects the growing evidence for stock return predictability in predictive regressions. This evidence also spurred the development of general equilibrium asset pricing models that feature rational time-varying expected returns (e.g., Campbell and Cochrane, 1999; Bansal and Yaron, 2004).

The evidence for U.S. aggregate stock return predictability is predominantly in-sample. In-

[9] More recent studies, including Brock et al. (1992), Brown et al. (1998), Lo et al. (2000), and Neely et al. (2012), provide evidence that certain technical indicators provide useful signals for forecasting stock returns. However, using White's (2000) "reality check" bootstrap, Sullivan et al. (1999, 2001) show that data snooping can account for some of the predictive ability of technical indicators. (We discuss data snooping in more detail in Section 3.4.) See Park and Irwin (2005) for a survey of the technical analysis literature.

[10] The cited studies are representative and do not constitute an exhaustive list.

sample tests of return predictability in the context of predictive regressions are complicated by the well-known Stambaugh (1986, 1999) bias. This bias arises when the predictor is highly persistent and the predictor and return innovations are correlated. Importantly, the Stambaugh bias potentially leads to substantial size distortions when testing the null hypothesis of no predictability, $\beta = 0$, using a conventional t-statistic approach. A number of studies develop procedures for improving inference in predictive regressions with persistent predictors, including Amihud and Hurvich (2004), Lewellen (2004), Torous et al. (2005), Campbell and Yogo (2006), Amihud et al. (2009), and Pástor and Stambaugh (2009).[11] Furthermore, the evidence for stock return predictability using valuation ratios frequently appears stronger at longer horizons. Studies investigating long-horizon predictability typically employ overlapping return observations. Overlapping observations induce severe serial correlation in the disturbance term in predictive regressions, potentially creating additional size distortions in conventional tests. Studies that analyze the statistical implications of predictive regressions with overlapping returns and develop econometric procedures for making more reliable inferences include Hodrick (1992), Goetzmann and Jorion (1993), Nelson and Kim (1993), Valkanov (2003), Boudoukh et al. (2008), Britten-Jones et al. (2011), and Hjalmarsson (2012). Despite the formidable econometric difficulties surrounding in-sample predictive regression tests, Campbell (2000, p. 1512) observes that "most financial economists appear to have accepted that aggregate returns do contain an important predictable component."

Bossaerts and Hillion (1999), Goyal and Welch (2003, 2008), Brennan and Xia (2005), and Butler et al. (2005) argue that the in-sample evidence of return predictability is not robust to out-of-sample validation. The study by Goyal and Welch (2008), which won the 2008 Michael Brennan Best Paper Award for the *Review of Financial Studies*, has been especially influential in this regard. Considering a variety of economic variables from the literature, Goyal and Welch (2008) show that out-of-sample equity premium forecasts based on the bivariate predictive regression, (1), fail to consistently outperform the simple historical average benchmark forecast in terms of MSFE. The historical average forecast assumes that $\beta = 0$ in (1) and corresponds to the constant expected

[11] Cavanaugh et al. (1995) is an important precursor to studies of predictive regressions with persistent predictors.

equity premium (or random walk with drift) model, implying that the information in x_t is not useful for predicting the equity premium. A multiple regression forecasting model that includes all potential predictors—the "kitchen sink" forecast—also performs much worse than the historical average forecast; this is not surprising, since it is well known that, due to in-sample overfitting, highly parameterized models typically perform very poorly in out-of-sample forecasting. Overall, Goyal and Welch (2008) conclude that predictive regressions are unstable and that conventional forecasting models fail to outperform the historical average.

Fortunately, a collection of recent studies shows that certain forecasting approaches improve upon conventional predictive regression forecasts and significantly outperform the historical average forecast in out-of-sample tests. These approaches improve forecasting performance by addressing the substantial model uncertainty and parameter instability characterizing the data-generating process for stock returns.

3.1.1 Economically Motivated Model Restrictions

The first approach for improving forecasting performance imposes economically motivated restrictions on predictive regression forecasts of stock returns. Recall the bivariate predictive regression model given by (1),

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1}, \tag{7}$$

where r_{t+1} is now the the log excess stock return and the i subscript indexes one of K potential return predictors ($i = 1, \ldots, K$). An equity premium forecast based on (7) is naturally computed as

$$\hat{r}_{i,t+1} = \hat{\alpha}_{i,t} + \hat{\beta}_{i,t} x_{i,t}, \tag{8}$$

where $\hat{\alpha}_{i,t}$ and $\hat{\beta}_{i,t}$ are ordinary least squares (OLS) estimates of α_i and β_i, respectively, in (7) based on data from the start of the available sample through t. As discussed in Section 2, since out-of-sample forecasts can only utilize data up to the time of forecast formation, these parameter estimates will be less efficient than their in-sample counterparts. The limited available estimation sample—and given that stock returns contain a sizable unpredictable component—means that the forecasting model parameters are potentially very imprecisely estimated, which can lead to poor forecasting performance. In response, Campbell and Thompson (2008) recommend imposing sign restrictions on $\hat{\beta}_{i,t}$ and $\hat{r}_{i,t+1}$ in (8). In particular, theory typically suggests the sign of β_i in (7); if $\hat{\beta}_{i,t}$ has an unexpected sign, then we set $\hat{\beta}_{i,t} = 0$ in (8) when forming the forecast. In addition, risk considerations usually imply a positive expected equity premium, so that we set the forecast equal to zero if $\hat{r}_{i,t+1} < 0$ in (8). Such sign restrictions reduce parameter estimation uncertainty and help to stabilize predictive regression forecasts. Campbell and Thompson (2008) find that, in contrast to unrestricted bivariate predictive regression forecasts, restricted predictive regression forecasts based on a number of economic variables outperform the historical average forecast.

Campbell and Thompson (2008), Campbell (2008), and Ferreira and Santa-Clara (2011) consider other types of restrictions on stock return forecasts involving valuation ratios. We focus on Ferreira and Santa-Clara's (2011) sum-of-the-parts method. By definition, the gross return on a broad market index is

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = CG_{t+1} + DY_{t+1}, \tag{9}$$

where P_t is the stock price, D_t is the dividend, $CG_{t+1} = P_{t+1}/P_t$ is the gross capital gain, and $DY_{t+1} = D_{t+1}/P_t$ is the dividend yield. The gross capital gain can be expressed as

$$CG_{t+1} = \frac{(P_{t+1}/E_{t+1})}{(P_t/E_t)} \frac{E_{t+1}}{E_t} = \frac{M_{t+1}}{M_t} \frac{E_{t+1}}{E_t} = GM_{t+1} GE_{t+1}, \tag{10}$$

where E_t denotes earnings, $M_t = P_t/E_t$ is the price-earnings multiple, and $GM_{t+1} = M_{t+1}/M_t$ ($GE_{t+1} = E_{t+1}/E_t$) is the gross growth rate of the price-earnings multiple (earnings). Using (10), the dividend yield can be written as

$$DY_{t+1} = \frac{D_{t+1}}{P_{t+1}} \frac{P_{t+1}}{P_t} = DP_{t+1} GM_{t+1} GE_{t+1}, \tag{11}$$

where $DP_t = D_t/P_t$ is the dividend-price ratio. Based on (10) and (11), the gross return in (9) becomes

$$R_{t+1} = GM_{t+1}GE_{t+1}(1+DP_{t+1}), \tag{12}$$

which, for the log return, can be expressed as

$$\log(R_{t+1}) = gm_{t+1} + ge_{t+1} + dp_{t+1}, \tag{13}$$

where gm_{t+1} (ge_{t+1}) is the log growth rate of the price-earnings multiple (earnings), and dp_t is the log of one plus the dividend-price ratio.

Ferreira and Santa-Clara (2011) use (13) as the basis for a stock return forecast. Since price-earnings multiples and dividend-price ratios are highly persistent and nearly random walks, reasonable forecasts of gm_{t+1} and dp_{t+1} based on information through t are zero and dp_t, respectively. Earnings growth is nearly entirely unpredictable, apart from a low-frequency component (van Binsbergen and Koijen, 2010), so that Ferreira and Santa-Clara (2011) employ a 20-year moving average of log earnings growth through t, $\overline{ge}_t^{20}$, as a forecast of ge_{t+1}. Their sum-of-the-parts equity premium forecast is then given by

$$\hat{r}_{t+1}^{SOP} = \overline{ge}_t^{20} + dp_t - r_{f,t+1}, \tag{14}$$

where $r_{f,t+1}$ is the log risk-free rate, which is known at the end of t.[12] Comparing (14) to (8), it is evident that the sum-of-the-parts forecast is a predictive regression forecast that restricts the slope coefficient to unity for $x_{i,t} = dp_t$ and sets the intercept to $\overline{ge}_t^{20} - r_{f,t+1}$.[13] Ferreira and Santa-Clara (2011) show that their sum-of-the-parts forecast significantly outperforms the historical av-

[12] Ferriera and Santa-Clara (2011) report complete results for the log return, but they note that their results are similar for the log excess return.

[13] While Ferreira and Santa-Clara (2011) focus on log returns, we can compute a simple (instead of log) return forecast as follows. Let $gr_{m,t+1}$ ($gr_{e,t+1}$) denote the net growth rate of the price-earnings multiple (earnings). Using these definitions, (12) becomes $R_{t+1} = (1+gr_{m,t+1})(1+gr_{e,t+1})(1+DP_{t+1})$. Multiplying out the right-hand-side and treating all cross-product terms as approximately zero, we have $R_{t+1} = 1 + gr_{m,t+1} + gr_{e,t+1} + DP_{t+1}$. Treating the price-earnings multiple and dividend-price ratio as approximately random walks, a simple excess return forecast is given by $\overline{gr}_{e,t}^{20} + DP_t$ minus the risk-free rate, where $\overline{gr}_{e,t}^{20}$ is a 20-year moving average of earnings growth through t. This forecast is analogous to (14). We use this simple excess return forecast for the asset allocation exercise in our empirical application in Section 3.3.

erage forecast. Furthermore, Monte Carlo simulations indicate that the sum-of-the-parts forecast improves upon conventional predictive regression forecasts by substantially reducing estimation error.

Bayesian approaches provide another avenue for incorporating economically reasonable restrictions via prior views. Along this line, a number of important studies, including Kandel and Stambaugh (1996), Barberis (2000), Wachter and Warusawitharana (2009), and Pástor and Stambaugh (2012), employ Bayesian methods to examine the implications of prior views on return predictability and estimation risk for optimal portfolio choice. An important result emerging from these studies is that return predictability significantly affects asset allocation, even for investors with relatively skeptical priors beliefs on the existence of return predictability.

3.1.2 Forecast Combination

Rapach et al. (2010) consider another approach for improving equity premium forecasts based on forecast combination. Since Bates and Granger's (1969) seminal paper, it has been know that combining forecasts across models often produces a forecast that performs better than the best individual model. As emphasized by Timmermann (2006), forecast combination can be viewed as a diversification strategy that improves forecasting performance in the same manner that asset diversification improves portfolio performance. Intuitively, from the standpoint of equity premium forecasting, particular forecasting models capture different aspects of business conditions; furthermore, the predictive power of individual models can vary over time, so that a given model provides informative signals during certain periods but predominantly false signals during others. If the individual forecasts are relatively weakly correlated, a combination of the individual forecasts should be less volatile, thereby stabilizing the individual forecasts, reducing forecasting risk, and improving forecasting performance in environments with substantial model uncertainly and parameter instability (e.g., Hendry and Clements, 2004; Clements and Hendry, 2006; Timmermann, 2006).

A combination (or pooled) forecast takes the form of a weighted average of the individual forecasts given by (8) for $i = 1, \ldots, K$:

$$\hat{r}_{t+1}^{POOL} = \sum_{i=1}^{K} \omega_{i,t} \hat{r}_{i,t+1}, \tag{15}$$

where $\{\omega_{i,t}\}_{i=1}^{K}$ are the combining weights based on information available through t and $\sum_{i=1}^{K} \omega_{i,t} = 1$. Simple combining schemes frequently perform surprisingly well. The simplest scheme sets $\omega_{i,t} = 1/K$ for all i to give the mean combination forecast. This is analogous to a "naïve" portfolio rule that places equal weight on each asset. The advantage of simple rules is that they do not require the estimation of combining weights. Similar to the situation discussed previously with respect to parameter estimation, it is typically difficult to precisely estimate weights for more elaborate combining schemes.

Nevertheless, it can be beneficial to "tilt" the combining weights toward certain individual forecasts, although it is advisable to hew relatively closely to equal weights. In line with this idea, Rapach et al. (2010) compute a discount MSFE (DMSFE) combination forecast that computes weights based on the forecasting performance of individual models over a holdout out-of-sample period (Stock and Watson, 2004):

$$\omega_{i,t} = \phi_{i,t}^{-1} / \sum_{k=1}^{K} \phi_{k,t}^{-1}, \tag{16}$$

where

$$\phi_{i,t} = \sum_{s=m}^{t-1} \theta^{t-1-s} (r_{s+1} - \hat{r}_{i,s+1})^2, \tag{17}$$

$m+1$ delineates the start of the holdout out-of-sample period, and θ is a discount factor. The DMSFE forecast thus attaches greater weight to individual predictive regression forecasts with lower MSFE (better forecasting performance) over the holdout out-of-sample period. When $\theta = 1$, there is no discounting, so that all observations are treated equally when computing MSFE over the holdout out-of-sample period, while $\theta < 1$ allows for greater emphasis on recent forecasting performance.[14] Rapach et al. (2010) show that simple and DMSFE combination forecasts of the

[14]When the individual forecasts are uncorrelated, $\theta = 1$ corresponds to the optimal combination forecast derived by Bates and Granger (1969). More generally, when individual forecasts are correlated, the optimal weights depend on the correlations. Given the difficulties in precisely estimating the parameters for optimal weights, theoretically optimal

quarterly U.S. equity premium consistently outperform the historical average.

It is curious that the simple combination forecast performs much better than the kitchen sink forecast, since both approaches entail the estimation of K slope coefficients. Rapach et al. (2010) show that the simple combination forecast can be interpreted as a "shrinkage" forecast that circumvents the in-sample overfitting problem plaguing highly parameterized forecasting models. Consider the multiple predictive regression model underlying the kitchen sink forecast, which, for transparency, we express in deviation form:

$$r_{t+1} - \bar{r} = \sum_{i=1}^{K} \beta_i^{KS}(x_{i,t} - \bar{x}_i) + \varepsilon_{t+1}, \tag{18}$$

where $\bar{r}$ and $\bar{x}_i$ are the sample means based on data available at the time of forecast formation for r_t and $x_{i,t}$, respectively. The kitchen sink forecast is then given by

$$\hat{r}_{t+1} = \bar{r} + \sum_{i=1}^{K} \hat{\beta}_i^{KS}(x_{i,t} - \bar{x}_i), \tag{19}$$

where $\hat{\beta}_i^{KS}$ is the OLS estimate of β_i^{KS} in the multiple regression, (18), using data available at the time of forecast formation. The simple combination forecast can be expressed as

$$\hat{r}_{t+1} = \bar{r} + (1/K)\sum_{i=1}^{K} \hat{\beta}_i(x_{i,t} - \bar{x}_i), \tag{20}$$

where $\hat{\beta}_i$ is the OLS slope coefficient estimate for the bivariate regression of r_{t+1} on $(x_{i,t} - \bar{x})$ based on data available at the time of forecast formation. Comparing (20) to (19), we see that the simple combination forecast replaces $\hat{\beta}_i^{KS}$ in (19) with $(1/K)\hat{\beta}_i$. This stabilizes the forecast via two channels: (1) reducing estimation variability by substituting the bivariate regression slope coefficient estimates for the multiple regression estimates; (2) shrinking the forecast toward the historical average forecast by premultiplying each slope coefficient by $1/K$. Stabilization permits the combination forecast to incorporate information from a host of economic variables while avoiding

weights frequently do not perform well in practice. See Timmermann (2006) for a detailed treatment of theoretically optimal combining weights and their practical limitations.

in-sample overfitting.

In a similar spirit to forecast combination, Cremers (2002) uses Bayesian model averaging to incorporate information from a multitude of potential predictors in a predictive regression framework. In essence, Bayesian model averaging provides an alternative procedure for shrinking the slope coefficients in (19), where the degree of shrinkage is now governed by the posterior probabilities that each of the predictors appears in the model (as well as the prior distributions for the slope coefficients). The slope coefficients for predictors with "low" posterior inclusion probabilities receive greater shrinkage toward zero, thereby stabilizing the forecast and preventing overfitting. For the 1969–1998 forecast evaluation period, Cremers (2002) finds that monthly U.S. equity premium forecasts based on Bayesian model averaging slightly outperform the historical average benchmark, while they substantially outperform forecasts based on models selected via conventional information criteria such as the AIC and SIC.

3.1.3 Diffusion Indices

Ludvigson and Ng (2007), Kelly and Pruitt (2012), and Neely et al. (2012) adopt a diffusion index approach to improve equity premium forecasting. Diffusion indices provide a means for conveniently tracking the key comovements in a large number of potential return predictors. The diffusion index approach assumes a latent factor model structure for the potential predictors:

$$x_{i,t} = \lambda_i' f_t + e_{i,t} \quad (i = 1, \ldots, K), \tag{21}$$

where f_t is a q-vector of latent factors, λ_i is a q-vector of factor loadings, and $e_{i,t}$ is a zero-mean disturbance term. A strict factor model assumes that the disturbance terms are contemporaneously and serially uncorrelated, while an "approximate" factor model permits a limited degree of contemporaneous and/or serial correlation in $e_{i,t}$ (e.g., Bai, 2003). Under (21), comovements in the predictors are primarily governed by fluctuations in the relatively small number of factors ($q \ll K$).

For either the strict or approximate factor model, the latent factors can be consistently estimated by principal components (e.g., Bai, 2003; Stock and Watson, 2006). Estimates of the latent factors then serve as regressors in the following predictive regression model:

$$r_{t+1} = \alpha_{DI} + \beta_{DI}' f_t + \varepsilon_{t+1}, \tag{22}$$

where β_{DI}' is a q-vector of slope coefficients. The basic intuition behind (22) is the following. All of the K predictors, $x_{i,t}$ $(i = 1, \ldots, K)$, potentially contain relevant information for forecasting r_{t+1}; however, as previously discussed, individual predictors can also provide noisy signals. Rather than using the $x_{i,t}$ variables directly in a predictive regression, we use the factor structure in (21) to identify the important common fluctuations in the potential predictors—as represented by f_t—thereby filtering out the noise in the individual predictors—as captured by $e_{i,t}$. The factor structure thus generates a more reliable signal from a large number of predictors to employ in a predictive regression.

An equity premium forecast based on (22) is given by

$$\hat{r}_{t+1}^{DI} = \hat{\alpha}_{DI,t} + \hat{\beta}_{DI,t}' \hat{f}_{t,t}, \tag{23}$$

where $\hat{f}_{t,t}$ is the principal component estimate of f_t based on data available through t and $\hat{\alpha}_{DI,t}$ and $\hat{\beta}_{DI,t}$ are OLS estimates of α_{DI} and β_{DI}, respectively, in (22) from regressing $\{r_s\}_{s=2}^{t}$ on a constant and $\{\hat{f}_{s,t}\}_{s=1}^{t-1}$. Implementation of this approach requires the specification of q, the number of latent factors. Bai and Ng (2002) and Onatski (2010) provide procedures for consistently selecting q, and these procedures can be applied to data available through t as a first step in computing the diffusion index forecast, (23). From a forecasting perspective, it is advisable to keep q relatively small, again to avoid an overparameterized forecasting model.

Ludvigson and Ng (2007) analyze diffusion index models along the lines of (22) for quarterly data and factors extracted from 209 macroeconomic and 172 financial variables. In addition to detecting significant in-sample predictive power for the estimated factors, Ludvigson and Ng (2007) find that diffusion index forecasts of the quarterly U.S. equity premium substantially outperform the historical average forecast.

Neely et al. (2012) use a diffusion index approach to forecast the monthly U.S. equity premium. They extract factors from a set of 14 economic variables from the literature and 14 technical indicators computed from moving average, momentum, and volume-based rules. Selecting the number of factors using the Onatski (2010) ED algorithm, Neely et al. (2012) show that the diffusion index forecast significantly outperforms the historical average, as well as the 28 predictive regression forecasts based on the individual economic variables and technical indicators. At present, their diffusion index approach based on both economic variables and technical indicators appears to provide the best monthly U.S. equity premium forecast.

An interesting extension of the diffusion index approach relies on "targeted" predictors (Bai and Ng, 2008).[15] From a forecasting standpoint, a potential drawback to the diffusion index model is that the estimated factors are designed to explain the covariation among the individual predictors themselves, without explicitly taking into account the relationship between the predictors and the targeted variable that we want to forecast. Kelly and Pruitt (2011) develop a three-pass regression filter (3PRF) to estimate the factors that are the most relevant for forecasting the target. In an application of the 3PRF approach, Kelly and Pruitt (2012) use factors extracted from an array of disaggregated valuation ratios to generate out-of-sample U.S. equity premium forecasts that significantly outperform the historical average forecast.

3.1.4 Regime Shifts

A fourth approach for improving equity premium forecastability centers on regime shifts. This approach recognizes that the data-generating process for stock returns is subject to parameter instability, in line with the results of Paye and Timmermann (2006) and Rapach and Wohar (2006a), who find significant evidence of parameter instability in predictive regression models of U.S. aggregate stock returns using the structural break tests of Bai and Perron (1998, 2003), Hansen

[15] See the chapter by Serena Ng in this volume for a detailed analysis of targeted predictors.

(2000), and Elliott and Müller (2006).[16]

One strategy for modeling breaks follows the pioneering work of Hamilton (1989) by assuming that parameters take on different values as the economy switches between a relatively small number of latent states.[17] In the context of stock return predictability, it is natural to expect such states, corresponding, for example, to bull/bear markets or business-cycle expansions/recessions. Consider the following Markov-switching predictive regression model:

$$r_{t+1} = \alpha_{S_{t+1}} + \beta'_{S_{t+1}} x_t + \sigma_{S_{t+1}} u_{t+1}, \tag{24}$$

where S_{t+1} is a first-order Markov-switching process representing the state of the economy, x_t is a vector of predictors, and u_{t+1} is a zero-mean variate with unit variance. S_{t+1} can take integer values between 1 and m, corresponding to the state of the economy, where the transition between states is governed by an $m \times m$ matrix with typical element,

$$p_{ij} = \Pr(S_t = j | S_{t-1} = i) \ \ (i, j = 1, \ldots, m). \tag{25}$$

Since the state of the economy is unobservable, (24) cannot be estimated using conventional regression techniques. Hamilton (1989) develops a nonlinear iterative filter that can be used to estimate the parameters of Markov-switching models via maximum likelihood and make inferences regarding the state of the economy.

Conditional on the parameter estimates in (24), a forecast of r_{t+1} for $m = 2$ is given by

$$\hat{r}^{MS}_{t+1} = \Pr(S_{t+1} = 1 | I_t)(\hat{\alpha}_{1,t} + \hat{\beta}'_{1,t} x_t) + \Pr(S_{t+1} = 2 | I_t)(\hat{\alpha}_{2,t} + \hat{\beta}'_{2,t} x_t), \tag{26}$$

where $\Pr(S_{t+1} = j | I_t)$ is the probability that $S_{t+1} = j$ given information available through t (which is produced by the estimation algorithm) and $\hat{\alpha}_{j,t}$ and $\hat{\beta}_{j,t}$ are the estimates of α_j and β_j, re-

[16]Paye and Timmermann (2006) also find significance evidence of structural breaks in predictive regression models of aggregate returns for other G-7 countries.

[17]Ang and Timmermann (2012) provide an insightful survey of regime switching in financial markets.

spectively, in (24) using data available through t for $j = 1,2$. Intuitively, (26) diversifies across forecasts from the two possible regimes. In periods where it is difficult to determine next period's state, $\Pr(S_{t+1} = 1|I_t) \approx \Pr(S_{t+1} = 2|I_t) \approx 0.5$, so that approximately equal weights are placed on the two regime forecasts; if there is strong evidence based on data through t that $S_{t+1} = 1$ ($S_{t+1} = 2$), then much more weight is placed on the first (second) regime forecast. In this way, the Markov-switching forecast accommodates structural breaks in model parameters while accounting for the uncertainty inherent in identifying the state of the economy.

Guidolin and Timmermann (2007) estimate a multivariate four-regime Markov-switching model for U.S. stock and bond returns via maximum likelihood, where the dividend yield serves as a predictor.[18] Characterizing the four states as "crash," "slow growth," "bull," and "recovery," they present statistical evidence favoring a four-regime structure. Most relevant from our perspective, they also find that real-time asset allocation decisions guided by Markov-switching model forecasts of stock and bond returns yield substantial utility gains relative to asset allocation decisions based on constant expected excess return forecasts.[19]

Henkel et al. (2011) estimate (24) as part of a two-regime Markov-switching vector autogression process that includes the dividend-price ratio, short-term nominal interest rate, term spread, and default spread in x_t. They estimate their model via Bayesian methods and find that the two states correspond closely to NBER-dated business-cycle expansions and recessions, with in-sample return predictability highly concentrated during recessions. They also show that monthly U.S. equity premium forecasts based on (26) outperform the historical average benchmark in terms of MSFE, and, like the in-sample results, out-of-sample return predictability is concentrated during economic downturns. Overall, the results in Henkel et al. (2011) suggest that the historical average forecast is sufficient during "normal" times, while economic variables provide useful signals for forecasting returns during contractionary episodes.

Markov-switching models appear to be most popular type of nonlinear model for forecast-

[18] An equation similar to (24) with $m = 4$ is part of Guidolin and Timmermann's (2007) multivariate model.

[19] Ang and Bekaert (2002) find that regime switching has important implications for optimal portfolio choice for an investor allocating across equities in the United States, the United Kingdom, and Germany.

ing stock returns. Other well-known nonlinear specifications include threshold models, where regimes are defined by the values taken by observable transition variables relative to threshold parameters, and neural nets; see Franses and van Dijk (2000), Teräsvirta (2006), White (2006), and González-Rivera and Lee (2009) for informative surveys of nonlinear forecasting models. Guidolin et al. (2009) analyze the circumstances under which nonlinear forecasting models of stock returns are likely to be useful.[20]

Instead of parameters switching among a relatively small number of states via a Markov-switching process, time-varying parameter (TVP) models allow for parameters to evolve from period to period, so that each period can be viewed as a new regime. Dangl and Halling (2012) specify the following TVP predictive regression model:

$$r_{t+1} = \alpha_t + \beta_t' x_t + \varepsilon_{t+1}, \tag{27}$$

$$\begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix}, \tag{28}$$

where

$$\varepsilon_t \sim N(0, \sigma^2), \tag{29}$$

$$w_t \sim N(0, W_t), \tag{30}$$

and $w_t = (w_{1,t}, w_{2,t}')'$. According to (28), the intercept and slope coefficients in the predictive regression, (27), evolve as (driftless) random walks. After placing restrictions on W_t to limit the parameter space, the TVP model given by (27)–(30) can be estimated using the Kalman filter and maximum likelihood.

[20]Non- or semiparametric modeling represents another approach for approximating general functional forms for the relationship between expected returns and predictors. See Chen and Hong (2010) for recent promising results on equity premium forecasting using a nonparametric approach. Aït-Sahalia and Brandt (2001) employ semiparametric methods to directly relate predictors of stock return moments to optimal portfolio weights.

An equity premium forecast based on the TVP model is given by

$$\hat{r}_{t+1}^{TVP} = \hat{\alpha}_{t,t} + \hat{\beta}_{t,t}' x_t, \tag{31}$$

where $\hat{\alpha}_{t,t}$ and $\hat{\beta}_{t,t}$ are estimates of α_t and β_t, respectively, in (27) based on data available through t. The TVP forecast in (31) allows for forecasting model coefficients to evolve in a very general way, so that the TVP forecast can respond to changes in economic structure resulting from a myriad of factors, including changes in technology, institutions, and policy. Dangl and Halling (2012) employ Bayesian methods to estimate (27)–(30) for the monthly U.S. equity premium. They find that forecasts based on (31) significantly outperform the historical average, and, similarly to Henkel et al. (2011), the out-of-sample gains are concentrated during recessions.[21]

Pettenuzzo and Timmermann (2011) adopt a Bayesian approach that allows for a few, large breaks in predictive regression model coefficients, rather than assuming that the coefficients constantly evolve according to (30). Petteunuzzo and Timmermann (2011) are primarily interested in the implications of infrequent structural breaks for optimal portfolio choice. Indeed, they show that structural breaks in predictive regressions have important implications for investors with long horizons and that ignoring breaks can results in sizable welfare losses.

3.2 Forecast Evaluation

MSFE is the most popular metric for evaluating forecast accuracy, and it is not surprising that MSFE is routinely reported in studies of stock return forecastability. This raises two important issues. The first relates to statistical tests of equal MSFE when comparing forecasts from nested models, while the second concerns the adequacy of the MSFE criterion itself.

To fix ideas, suppose that a sample of T observations for r_t and $x_{i,t}$ is available. We divide the total sample into an initial in-sample estimation period comprised of the first n_1 observations and an out-of-sample period comprised of the last $n_2 = T - n_1$ observations. One-step-ahead equity

[21] Since they consider a large number of potential predictors, Dangl and Halling (2012) also employ Bayesian model averaging along the line of Cremers (2002) in constructing out-of-sample forecasts for their TVP model structure.

premium forecasts are computed over these last n_2 observations using (8). The MSFE for the predictive regression forecast over the forecast evaluation period is given by

$$\text{MSFE}_i = (1/n_2)\sum_{s=1}^{n_2}(r_{n_1+s} - \hat{r}_{i,n_1+s})^2\,. \tag{32}$$

To analyze out-of-sample stock return predictability, the accuracy of the predictive regression forecast is compared to the historical average benchmark forecast, which assumes constant expected excess returns,

$$\bar{r}_{t+1} = (1/t)\sum_{s=1}^{t} r_t, \tag{33}$$

and its MSFE is given by

$$\text{MSFE}_0 = (1/n_2)\sum_{s=1}^{n_2}(r_{n_1+s} - \bar{r}_{n_1+s})^2\,. \tag{34}$$

The out-of-sample R^2 (Campbell and Thompson, 2008) is a convenient statistic for comparing MSFEs. It is analogous to the conventional in-sample R^2 and measures the proportional reduction in MSFE for the predictive regression forecast relative to the historical average:

$$R^2_{OS} = 1 - (\text{MSFE}_i/\text{MSFE}_0). \tag{35}$$

Obviously, when $R^2_{OS} > 0$, the predictive regression forecast is more accurate than the historical average in terms of MSFE ($\text{MSFE}_i < \text{MSFE}_0$).

While R^2_{OS} measures the improvement in MSFE for the predictive regression forecast visá-vis the historical average, we are also interested in determining whether the improvement is statistically significant; that is, we are interested in testing H_0: $\text{MSFE}_0 \leq \text{MSFE}_i$ against H_A: $\text{MSFE}_0 > \text{MSFE}_i$, corresponding to H_0: $R^2_{OS} \leq 0$ against H_A: $R^2_{OS} > 0$.[22] The well-known Diebold and Mariano (1995) and West (1996) statistic for testing the null of equal MSFE (or equal predic-

[22] Corradi and Swanson (2006), West (2006), and the chapter by Todd Clark and Michael McCracken in this volume provide instructive surveys of statistical tests of relative forecast accuracy.

tive ability) is given by

$$DMW_i = n_2^{0.5} \bar{d}_i \hat{S}_{d_i,d_i}^{-0.5}, \tag{36}$$

where

$$\begin{aligned}
\bar{d}_i &= (1/n_2) \sum_{s=1}^{n_2} \hat{d}_{i,n_1+s}, &(37)\\
\hat{d}_{i,n_1+s} &= \hat{u}_{0,n_1+s}^2 - \hat{u}_{i,n_1+s}^2, &(38)\\
\hat{u}_{0,n_1+s} &= r_{n_1+s} - \bar{r}_{n_1+s}, &(39)\\
\hat{u}_{i,n_1+s} &= r_{n_1+s} - \hat{r}_{i,n_1+s}, &(40)\\
\hat{S}_{d_i,d_i} &= (1/n_2) \sum_{s=1}^{n_2} \left(\hat{d}_{i,n_1+s} - \bar{d}_i\right)^2. &(41)
\end{aligned}$$

The DMW_i statistic is equivalent to the t-statistic corresponding to the constant for a regression of $\hat{d}_{i,n_1+s}$ on a constant for $s = 1, \ldots, n_2$. When comparing forecasts from nonnested models, Diebold and Mariano (1995) and West (1996) show that DMW_i has a standard normal asymptotic distribution, so that it is straightforward to test H_0: $\text{MSFE}_0 \leq \text{MSFE}_i$ against H_A: $\text{MSFE}_0 > \text{MSFE}_i$ by comparing the sample statistic to 1.282, 1.645, and 2.326 for the 10%, 5%, and 1% significance levels, respectively.

Clark and McCracken (2001) and McCracken (2007), however, show that DMW_i has a nonstandard asymptotic distribution when comparing forecasts from nested models. In the context of predictive regressions, out-of-sample tests of stock return predictability entail a comparison of nested forecasts, since the predictive regression model, (7), reduces to the constant expected excess return model when $\beta_i = 0$. For nested forecast comparisons, the asymptotic distribution of DMW_i is a function of Brownian motion and depends on two parameters: (1) $\pi = n_2/n_1$; (2) the dimension of the set of predictors, x_t (which is one for a bivariate predictive regression). Clark and McCracken (2001) and McCracken (2007) provide tabulated critical values for a variety of parameter values that are relevant in applied research. A stark feature of the asymptotic critical values is that they frequently shift markedly to the left relative to standard normal critical values.

For example, consider a bivariate predictive regression model and $\pi = 2$, which corresponds to reserving the first third of the total sample for the initial in-sample period. From Table 1 in McCracken (2007), the 10%, 5%, and 1% critical values are 0.281, 0.610, and 1.238, respectively, which are well below their standard normal counterparts. The implication is that tests of equal predictive ability based on conventional critical values can often be severely undersized, leading to tests with very low power to detect out-of-sample return predictability. In short, it is crucial to use appropriate critical values when testing for stock return forecastability; otherwise, statistically significant evidence of out-sample return predictability can easily be missed.

Clark and West (2007) adjust DMW_i to produce a modified statistic, *MSFE-adjusted*, for comparing nested model forecasts that has an asymptotic distribution well approximated by the standard normal. The *MSFE-adjusted* statistic also performs well in finite-sample simulations. Clark and West (2007) thus provide a very convenient method for assessing statistical significance when comparing nested forecasts that obviates the need to look up a new set of critical values for each application. The *MSFE-adjusted* statistic is straightforward to compute by first defining

$$\tilde{d}_{i,n_1+s} = \hat{u}^2_{0,n_1+s} - \left[\hat{u}^2_{i,n_1+s} - (\bar{r}_{n_1+s} - \hat{r}_{i,n_1+s})^2\right], \tag{42}$$

and then regressing $\tilde{d}_{i,n_1+s}$ on a constant for $s = 1, \ldots, n_2$; *MSFE-adjusted* is the t-statistic corresponding to the constant. Recent studies of stock return forecastability that employ the *MSFE-adjusted* statistic include Rapach et al. (2010), Kong et al. (2011), Dangl and Halling (2012), and Neely et al. (2012).[23]

While MSFE is overwhelmingly the most popular measure of forecast accuracy, it is not necessarily the most relevant metric for assessing stock return forecasts. Leitch and Tanner (1991) consider why many firms purchase professional forecasts of economic and financial variables that frequently fail to outperform forecasts from simple time-series models in terms of MSFE. They

[23]The Diebold and Mariano (1995) and West (1996) DMW_i and Clark and West (2007) *MSFE-adjusted* statistics are tests of population-level predictability. There is a subtle but important distinction between population-level and finite-sample predictability. Section 3 of the chapter by Todd Clark and Michael McCracken in this volume discusses this distinction and recently proposed procedures for testing finite-sample predictability that have potentially important implications for analyzing stock return forecastability.

argue that forecast profitability is a more relevant metric for assessing forecasts, helping to explain the value of professional forecasts to firms.[24] Comparing professional and simple time-series model forecasts of interest rates, for which profitability is readily measured, Leitch and Tanner (1991) find that professional forecasts are often more profitable. Furthermore, there is a weak relationship between MSFE and forecast profitability. Among the conventional forecast error measures that they analyze, only the direction-of-change metric appears significantly correlated with forecast profitability. Henrisksson and Merton (1981), Cumby and Modest (1987), and Pesaran and Timmermann (1992) provide statistical tests of directional forecasting ability, and academic researchers often employ these tests when analyzing stock return forecasts (e.g., Breen et al, 1989; Pesaran and Timmermann, 2002, 2004; Marquering and Verbeek, 2004).

In line with the conclusions of Leitch and Tanner (1991), academic researchers also frequently analyze stock return forecasts with profit- or utility-based metrics, which provide more direct measures of the value of forecasts to economic agents. In these exercises, stock return forecasts serve as inputs for *ad hoc* trading rules or asset allocation decisions derived from expected utility maximization problems. A leading utility-based metric for analyzing U.S. equity premium forecasts is the average utility gain for a mean-variance investor. Consider a mean-variance investor with relative risk aversion γ who allocates her portfolio between stocks and risk-free bills based on the predictive regression forecast, (8), of the equity premium.[25] At the end of t, the investor allocates the following share of her portfolio to equities during $t+1$:

$$a_{i,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{i,t+1}}{\hat{\sigma}^2_{t+1}}\right), \tag{43}$$

where $\hat{\sigma}^2_{t+1}$ is a forecast of the variance of stock returns.[26] Over the forecast evaluation period, the

[24] Also see Granger and Machina (2006) and Batchelor (2011).

[25] For asset allocation exercises, we use simple (instead of log) returns, so that the portfolio return is given by the sum of the individual portfolio weights multiplied by the asset returns.

[26] Under the assumption of constant return volatility, the variance of stock returns can be estimated using the sample variance computed from a recursive window of historical returns. To allow for a time-varying variance, a rolling window or some type of GARCH model can be used. Campbell and Thompson (2008) estimate $\hat{\sigma}^2_{t+1}$ using the sample variance computed from a five-year rolling window of historical returns. See Andersen et al. (2006) for an extensive survey of return volatility forecasting. Note that a general expected utility maximization problem for an investor

investor realizes the average utility,

$$\hat{v}_i = \hat{\mu}_i - 0.5\gamma\hat{\sigma}_i^2, \tag{44}$$

where $\hat{\mu}_i$ ($\hat{\sigma}_i^2$) is the sample mean (variance) of the portfolio formed on the basis of $\hat{r}_{i,t+1}$ and $\hat{\sigma}_{t+1}^2$ over the forecast evaluation period. If the investor instead relies on the historical average forecast of the equity premium (using the same variance forecast), she allocates the portfolio share,

$$a_{0,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\bar{r}_{t+1}}{\hat{\sigma}_{t+1}^2}\right), \tag{45}$$

to equity during $t+1$ and, over the forecast evaluation period, realizes the average utility,

$$\hat{v}_0 = \hat{\mu}_0 - 0.5\gamma\hat{\sigma}_0^2, \tag{46}$$

where $\hat{\mu}_0$ ($\hat{\sigma}_0^2$) is the sample mean (variance) of the portfolio formed on the basis of $\bar{r}_{t+1}$ and $\hat{\sigma}_{t+1}^2$ over the forecast evaluation period. The difference between (44) and (46) represents the utility gain accruing to using the predictive regression forecast of the equity premium in place of the historical average forecast in the asset allocation decision. This utility gain, or certainty equivalent return, can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the information in the predictive regression forecast relative to the information in the historical average forecast alone. Marquering and Verbeek (2004), Campbell and Thompson (2008), Cooper and Priestly (2009), Rapach et al. (2010), Ferreira and Santa-Clara (2011), Dangl and Halling (2012), and Neely et al. (2012) all detect sizable utility gains for mean-variance investors who rely on equity premium forecasts based on economic variables relative to the historical average forecast.

Kandel and Stambaugh (1996), Guidolin and Timmermann (2007), Pettenuzzo and Timmermann (2011), and Cenesizoglu and Timmermann (2011) measure utility gains for investors with

requires a forecast of the entire conditional distribution of future returns, including conditional mean and volatility forecasts.

power utility defined over wealth who optimally allocate among assets, and they also find significant economic gains accruing to equity premium forecasts based on economic variables. Similarly to Leitch and Tanner (1991), Cenesizoglu and Timmermann (2011) find a weak relationship between MSFE and utility gains. Hong et al. (2007) examine whether equity premium forecasts based on lagged industry portfolio returns possess market timing ability. They consider a portfolio that allocates all of the portfolio to equities (Treasury bills) if the equity premium forecast is positive (negative). For investors with power utility, portfolios formed using equity premium forecasts based on lagged industry returns generate sizable utility gains relative to portfolios relying on equity premium forecasts that ignore lagged industry returns.

3.3 Empirical Application

3.3.1 Monthly U.S. Equity Premium Forecastability

We next consider an application based on forecasting the monthly U.S. equity premium that illustrates many of the concepts and methodologies discussed in Sections 3.1 and 3.2. We use updated data from Goyal and Welch (2008) spanning 1926:12–2010:12.[27] The equity premium is the log return on the S&P 500 (including dividends) minus the log return on a risk-free bill.[28] Fourteen popular economic variables serve as candidate predictors:

1. **Log dividend-price ratio** [$\log(DP)$]: log of a 12-month moving sum of dividends paid on the S&P 500 index minus the log of stock prices (S&P 500 index).

2. **Log dividend yield** [$\log(DY)$]: log of a 12-month moving sum of dividends minus the log of lagged stock prices.

[27]We thank Amit Goyal for kindly providing the data on his web page at http://www.hec.unil.ch/agoyal/. The data and their sources are described in detail in Goyal and Welch (2008). Other than the economic variables compiled by Goyal and Welch (2008), Han et al. (2012) and Neely et al. (2012) recently show that technical indicators are valuable as predictors in predictive regression forecasts of stock returns, a new finding in academic research.

[28]When performing asset allocation exercises, we measure the equity premium as the simple aggregate market return minus the simple risk-free rate.

3. **Log earnings-price ratio** [$\log(EP)$]: log of a 12-month moving sum of earnings on the S&P 500 index minus the log of stock prices.

4. **Log dividend-payout ratio** [$\log(DE)$]: log of a 12-month moving sum of dividends minus the log of a 12-month moving sum of earnings.

5. **Stock variance** ($SVAR$): monthly sum of squared daily returns on the S&P 500 index.

6. **Book-to-market ratio** (BM): book-to-market value ratio for the DJIA.

7. **Net equity expansion** ($NTIS$): ratio of a 12-month moving sum of net equity issues by NYSE-listed stocks to the total end-of-year market capitalization of NYSE stocks.

8. **Treasury bill rate** (TBL): interest rate on a three-month Treasury bill (secondary market).

9. **Long-term yield** (LTY): long-term government bond yield.

10. **Long-term return** (LTR): return on long-term government bonds.

11. **Term spread** (TMS): long-term yield minus the Treasury bill rate.

12. **Default yield spread** (DFY): difference between BAA- and AAA-rated corporate bond yields.

13. **Default return spread** (DFR): long-term corporate bond return minus the long-term government bond return.

14. **Inflation** ($INFL$): calculated from the CPI (all urban consumers); we use $x_{i,t-1}$ in (7) for inflation to account for the delay in CPI releases.

We first compute bivariate predictive regression forecasts of the equity premium based on (8) for each predictor. We use 1926:12–1956:12 as the initial in-sample estimation period, so that we compute out-of-sample forecasts for 1957:01–2010:12 (648 observations). The forecasts employ a recursive (or expanding) estimation window, meaning that the estimation sample always starts in 1926:12 and additional observations are used as they become available. Forecasting model parameters are also frequently estimated with a rolling window, which drops earlier observations as additional observations become available. Rolling estimation windows are typically justified by appealing to structural breaks, although a rolling window generally will not be an optimal estimation window in the presence of breaks. Pesaran and Timmermann (2007) and Clark and McCracken (2009) show that, from an MSFE perspective, it can be optimal to employ pre-break data when estimating forecasting models, a manifestation of the classic bias-efficiency tradeoff. More generally, they demonstrate that the optimal window size is a complicated function of the timing and size of breaks. Since these parameters are difficult to estimate precisely, recursive estimation windows frequently perform better in terms of MSFE than rolling windows or windows selected on the basis of structural break tests when forecasting stock returns.[29]

The 1957:01–2010:12 forecast evaluation period covers most of the postwar era, including the oil price shocks of the 1970s; the deep recession associated with the Volcker disinflation in the early 1980s; the long expansions of the 1960s, 1980s, and 1990s; and the recent Global Financial Crisis and concomitant Great Recession. The selection of the forecast evaluation period is always somewhat arbitrary. Hansen and Timmermann (2012) recently develop an out-of-sample test of forecasting ability that is robust to the in-sample/out-of-sample split.[30]

Table 1 reports results for out-of-sample horse races pitting the individual bivariate predictive regression forecasts against the historical average. Panels A and B of Table 1 give results for unrestricted predictive regression forecasts and predictive regression forecasts that implement the Campbell and Thompson (2008) sign restrictions described in Section 3.1.1, respectively.[31] In addition to the full 1957:01–2010:12 forecast evaluation period, we present results computed

[29] We confirmed this in our application. Rossi and Inoue (2012) develop out-of-sample tests of forecasting ability that are robust to the estimation window size.

[30] Section 5 of the chapter by Todd Clark and Michael McCracken in this volume examines issues relating to the choice of in-sample/out-of-sample split when evaluating forecasts. In Figures 1 and 3, we employ a graphical device from Goyal and Welch (2003, 2008) to assess the consistency of out-of-sample predictive ability.

[31] Specifically, the Campbell and Thompson (2008) restrictions entail setting the slope coefficient used to generate the bivariate predictive regression forecast to zero if the sign of the estimated slope coefficient differs from its theoretically expected sign, and a nonnegativity constraint is imposed by setting the forecast to zero if the predictive regression forecast is negative.

separately during NBER-dated business-cycle expansions and recessions.[32]

TABLE 1 HERE

The R^2_{OS} statistics in the second column of Table 1, Panel A succinctly convey the message of Goyal and Welch (2008): individual predictive regression forecasts frequently fail to beat the historical average benchmark in terms of MSFE. Indeed, 12 of the 14 R^2_{OS} statistics are negative in the second column of Panel A, indicating that the predictive regression forecast has a higher MSFE than the historical average. For the two predictors with a positive R^2_{OS} ($SVAR$ and TMS), the Clark and West (2007) p-values reported in the brackets are greater than 0.10, so that these economic variables do not display statistically significant out-of-sample predictive ability at conventional levels.[33]

Figure 1 further illustrates the generally poor performance of the conventional predictive regression forecasts. The black line in each panel depicts the cumulative difference in squared forecast errors for the historical average forecast vis-á-vis the predictive regression forecast:

$$\text{CDSFE}_{i,\tau} = \sum_{s=1}^{\tau} (r_{n_1+s} - \bar{r}_{n_1+s})^2 - \sum_{s=1}^{\tau} (r_{n_1+s} - \hat{r}_{i,n_1+s})^2, \tag{47}$$

for $\tau = 1, \ldots, n_2$. Goyal and Welch (2003, 2008) recommend this highly informative graphical device to assess the ability of a predictive regression forecast to consistently outperform the historical average. The figure can be conveniently used to determine if the predictive regression forecast has a lower MSFE than the historical average for any period by simply comparing the height of the curve at the beginning and end points of the segment corresponding to the period of interest: if the curve is higher (lower) at the end of the segment relative to the beginning, then the predictive regression forecast has a lower (higher) MSFE than the historical average during the period. A

[32]The NBER business-cycle peak and trough dates defining expansions and recessions are available at http://www.nber.org/cycles.html. The U.S. economy was in recession for approximately 17% of the months during 1957:01–2010:12.

[33]Interestingly, three of the economic variables with negative R^2_{OS} statistics in Table 1—log(DP), log(DY), and TBL—actually have p-values less than or equal to 0.10, so that we reject the null that $R^2_{OS} \leq 0$ in favor or $R^2_{OS} > 0$ at conventional levels, even though the sample R^2_{OS} is negative. This is a manifestation of the Clark and McCracken (2001) and McCracken (2007) result discussed in Section 3.2.

predictive regression forecast that always outperforms the historical average will thus have a slope that is positive everywhere. Of course, realistically, this ideal will not be reached in practice, but the closer the curve is to this ideal the better. The black lines in Figure 1 all fall far short of the ideal. All of the lines are predominantly negatively sloped or flat, with relatively short and infrequent positively sloped segments. In fact, numerous segments are steeply and negatively sloped, indicating that the historical average substantially outperforms the predictive regression forecasts during these periods. At best, the predictive regression forecasts provide episodic—or " elusive" (Timmermann, 2008)—evidence of out-of-sample stock return predictability.[34] In sum, Figure 1, along with the R^2_{OS} statistics in the second column of Table 1, Panel A, support a bearish view of predictive regression forecasts of the U.S. equity premium.

FIGURE 1 HERE

Closer inspection of Figure 1 reveals a pattern to the episodic nature of equity premium forecastability. In a number of instances, the curves are relatively steeply and positively sloped during and around NBER-dated recessions (depicted by the vertical bars in Figure 1), indicating that out-of-sample stock return predictability is largely a recessionary phenomenon. Additional evidence of this is provided in the fourth and sixth columns of Table 1, Panel A, which report R^2_{OS} statistics computed separately during expansions and recessions, respectively. For seven of the predictors, the R^2_{OS} statistics move from being negative (and typically below -1%) during expansions to 1% or above during recessions. Furthermore, five of these R^2_{OS} statistics are significant at conventional levels during recessions according to the Clark and West (2007) p-values, despite the decreased number of available observations. The difference in relative out-of-sample forecasting performance across business-cycle phases is particularly evident for $\log(DP)$ and $\log(DY)$, where the R^2_{OS} statistics go from -1.24% and -2.28% during expansions to 2.41% and 3.56%, respectively, during recessions.

[34] Giacomini and Rossi (2010) provide a measure of relative local forecasting performance and tests of the stability of forecasting gains (also see Giacomini and Rossi, 2009). Their tests, however, require rolling estimation windows when comparing nested forecasts, while we focus on recursive estimation windows in our application.

Up to this point, we have analyzed equity premium forecasts in terms of MSFE. As discussed in Section 3.2, however, MSFE is not necessarily the most relevant metric for assessing stock return forecasts. The third column of Table 1, Panel A reports average utility gains (in annualized percent return) for a mean-variance investor with relative risk coefficient of five who allocates among equities and risk-free bills using predictive regression forecasts in place of the historical average.[35] Relative to the R^2_{OS} statistics in the second column, the predictive regression forecasts appear significantly more valuable according to the average utility gains, with 10 of the 14 economic variables offering positive gains. The annualized gain is above 0.5% for seven of the economic variables, meaning that the investor would be willing to pay more than 50 basis points to have access to the information in the predictive regression forecasts compared to the historical average forecast. The utility gains are greater than 100 basis points for $\log(DY)$, TBL, LTY, and TMS. Similarly to the pattern in the R^2_{OS} statistics, the average utility gains are typically higher during recessions than expansions (see the fifth and seventh columns of Table 1, Panel A). In fact, the differences in forecasting performance as measured by the utility gains are more pronounced than for the R^2_{OS} statistics. For 12 of the 14 predictors, the average utility gain is higher during recessions than expansions, and the differences are especially large for $\log(DP)$ and $\log(DY)$, where the average utility gains increase from -1.47% and -1.98% during expansions to 11.87% and 16.17%, respectively, during recessions. Overall, the average utility gains in Table 1, Panel A provide stronger support for stock return forecastability, highlighting the need to supplement standard statistical criteria with more direct value-based measures when analyzing out-of-sample stock return predictability.[36]

Panel B of Table 1 presents R^2_{OS} statistics and average utility gains for predictive regression forecasts that impose the Campbell and Thompson (2008) sign restrictions. Comparing the second column of Panel B to Panel A, we see that the theoretically motivated restrictions generally

[35] The results are qualitatively similar for other reasonable risk aversion coefficient values. We follow Campbell and Thompson (2008) and estimate the variance of stock returns using the sample variance computed from a five-year rolling window of historical returns.

[36] An important outstanding issue is assessing the statistical significance of average utility gains; see McCracken and Valente (2011) for insightful initial theoretical results.

improve the predictive regression forecasts in terms of MSFE. Eleven of the 14 R^2_{OS} statistics in Panel B are greater than their Panel A counterparts, and the R^2_{OS} statistics turn from negative to positive for $\log(DP)$, $\log(DY)$, TBL, and LTR as we move from Panel A to B. The R^2_{OS} statistics in the fourth and sixth columns of Panel B help to explain the overall increase in forecast accuracy corresponding to the restrictions. The R^2_{OS} statistics in the sixth column of Panels A and B are reasonably similar; if anything, there is a tendency for the restrictions to decrease the R^2_{OS} statistics during recessions. In contrast, the R^2_{OS} statistics become substantially less negative during expansions for a number of predictors, especially $\log(DP)$ and $\log(DY)$ (see the fourth column of Panels A and B). The restrictions thus appear to improve overall forecasting performance by ameliorating the underperformance of the predictive regression forecasts during expansions. The results in Panel B indicate that the restrictions have relatively little effect on forecasting performance as measured by the average utility gains.

Figure 2, which graphs the individual bivariate predictive regression forecasts, sheds additional light on the role of sign restrictions. The sign restrictions on the estimated slope coefficients for the predictive regression forecasting models are rarely binding.[37] Instead, as shown in Figure 2, the predictive regression forecasts are negative for many periods for a number of predictors, so that it is the sign restrictions on the predictive regression forecasts themselves that are relevant. For example, the predictive regression forecasts based on valuation ratios—$\log(DP)$, $\log(DY)$, $\log(EP)$, and *BM*—are often negative during the mid-to-late 1990s. The corresponding lines in Figure 1 are predominantly steeply and negatively sloped during this period, indicating that the predictive regression forecasts perform very poorly relative to the historical average. The gray lines in Figure 1 correspond to differences in cumulated squared forecast errors for the historical average forecast relative to the restricted predictive regression forecasts. The gray lines for the valuation ratios lie above the black lines starting in the mid 1990s, so that the restrictions are particularly useful in improving forecast performance around this time. In general, the sign restrictions stabilize the predictive regression forecasts by truncating them from below, helping to avoid more implausible

[37] The major exception is *SVAR*, where theory suggests $\beta_i > 0$, but the estimated slope coefficient is always negative. This is evident from the R^2_{OS} of exactly zero for *SVAR* in Panel B.

equity premium forecasts.

FIGURE 2 HERE

Panel A of Table 2 reports R^2_{OS} statistics and average utility gains for monthly equity premium forecasts based on multiple economic variables, with no sign restrictions placed on the forecasts. Figure 3 graphs the cumulative differences in squared forecast errors for the historical average forecast relative to forecasts based on multiple economic variables, while Figure 4 depicts the forecasts themselves.

The first forecast in Table 2, the kitchen sink, corresponds to a multiple predictive regression model that includes all 14 economic variables as regressors. Confirming the results in Goyal and Welch (2008) and Rapach et al. (2010), the kitchen sink forecast performs very poorly according to the MSFE metric, with an R^2_{OS} of -8.43% over the full forecast evaluation period. The line in Figure 3, Panel A is nearly always negative sloped, showing that the kitchen sink forecast consistently underperforms the historical average in terms of MSFE. Panel A of Figure 4 indicates that the kitchen sink forecast is highly volatile, more so than any of the individual bivariate predictive regression forecasts (note the difference in the vertical axis scales in Figure 2 and Figure 4, Panel A). The monthly kitchen sink forecast reaches nearly 4%—implying an annualized expected equity premium of nearly 48%—and falls below -4%—implying an annualized expected equity premium near -50%. These extreme values are highly implausible, clearly demonstrating the in-sample overfitting problem that causes highly parameterized models to produce large forecast errors; such errors are stringently penalized by the MSFE criterion. Despite the very poor performance of the kitchen sink forecast in terms of MSFE, it does deliver a positive overall average utility gain, although the gain is less than 25 basis points on an annualized basis. This again illustrates how different evaluation criteria can lead to different conclusions regarding stock return forecasting performance.

TABLE 2 HERE

FIGURE 3 HERE

FIGURE 4 HERE

In the spirit of Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999), the next forecast in Table 2 selects the forecasting model using the SIC, from among the 2^K possible specifications for the $K = 14$ potential predictors, based on data available at the time of forecast formation.[38] The idea is to use the SIC to guard against in-sample overfitting, since the SIC penalizes models with more parameters. While the R^2_{OS} increases for the full forecast evaluation period for the SIC relative to the kitchen sink forecast, it is still well below zero (-5.61%), and Figure 3, Panel B indicates that it is consistently outperformed by the historical average. Panel B of Figure 4 shows that, while the SIC forecast is less volatile than the kitchen sink forecast (as expected), it remains quite volatile, so that the SIC forecast still appears to be plagued by in-sample overfitting. The SIC forecast also fails to outperform the historical average in terms of average utility gain.

The last four forecasts considered in Table 2 employ three of the recently proposed approaches for improving equity premium forecasts reviewed in Section 3.1. The first two of these forecasts are combination forecasts based on (15), which we implement in two ways: (1) a simple combining scheme, $\omega_{i,t} = 1/K$ for $i = 1, \ldots, K$ (POOL-AVG); (2) combining weights that depend on recent forecasting performance, (16) for $\theta = 0.75$ (POOL-DMSFE).[39] The next forecast is the diffusion index forecast given by (23), where we use the first principal component extracted from the 14 economic variables. The final forecast is the sum-of-the-parts forecast in (14).

The results in Table 2, Panel A demonstrate the usefulness of recently proposed forecasting strategies. The POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the parts forecasts all deliver positive R^2_{OS} statistics for the full 1957:01–2010:12 forecast evaluation period, and each of the corresponding Clark and West (2007) p-values indicates significance at the 5% level.[40] The R^2_{OS} statistics for these four forecasts range from 0.44% (POOL-AVG) to 0.93% (sum-of-the-parts),

[38] We obtain similar results for other information criteria, such as the AIC.

[39] We use 1947:01–1956:12 as the initial holdout out-of-sample period when computing the POOL-DMSFE forecast.

[40] Strictly speaking, Clark and West (2007) analyze the *MSFE-adjusted* statistic for forecasts generated from linear predictive regressions estimated via OLS. The POOL-AVG, POOL-DMSFE, and sum-of-the-parts forecasts do not exactly conform to this structure, so that we use the *MSFE-adjusted* statistic as an approximation to statistical significance.

all of which are larger than any of the R_{OS}^2 statistics for the individual bivariate predictive regression forecasts in the second column of Table 1. In marked contrast to Figure 1 and Panels A and B of Figure 3, these four forecasts produce out-of-sample gains on a quite consistent basis over time, as demonstrated in Panels C–F of Figure 3. The POOL-AVG and POOL-DMSFE forecasts, in particular, deliver very consistent gains, with slopes in Figure 3 that are nearly always positive. Although the lines are briefly negatively sloped during the late 1990s, the deterioration in performance during this time is much milder than that exhibited by a number of the individual bivariate predictive regression forecasts in Figure 1 (not to mention the kitchen sink and SIC forecasts in Panels A and B of Figure 3). While it generates a higher R_{OS}^2 for the full forecast evaluation period, the diffusion index forecast performs more erratically than the combination forecasts, with a much sharper dropoff during the late 1990s. The sum-of-the-parts forecast generates a higher R_{OS}^2 for the full evaluation period than the combination and diffusion index forecasts, while providing less (more) consistent gains than the combination forecasts (diffusion index forecast). The POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the-parts forecasts also generate sizable average utility gains for the 1957:01–2010:12 forecast evaluation period, ranging from 125 to 247 basis points on an annualized basis. These four forecasts thus perform well according to both MSFE and direct utility-based criteria.

Panels C–F of Figure 4 provide insight into the success of the POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the-parts forecasts. Relative to many of the individual bivariate predictive regression forecasts in Figure 2 and the kitchen sink and SIC forecasts in Panels A and B of Figure 4, these four forecasts tend to be considerably more stable (again, note the differences in vertical axis scales). Stabilization is accomplished by shrinkage for the combination forecasts (Section 3.1.2), by filtering noise from predictors for the diffusion index forecast (Section 3.1.3), and by reducing estimation error for the sum-of-the-parts forecast (Section 3.1.1). Such stabilization is necessary due to the substantial model uncertainty and parameter instability surrounding stock return forecasting. Successful stock return forecasting strategies incorporate information from multiple economic variables—information ignored by the historical average forecast—but in a manner that accommodates model uncertainty and parameter instability, thereby producing return forecasts that are economically plausible.

Supporting the economic plausibility of the POOL-AVG, POOL-DMFSE, diffusion index, and sum-of-the parts forecasts, the behavior of these forecasts appears linked to business-cycle fluctuations in Figure 4. In particular, during the mid-1970s and 1990–1991 recessions, and especially the severe recession of the early 1980s, the forecasts descend to distinct local minima near business-cycle peaks and then increase sharply during the course of recessions, reaching distinct local maxima near business-cycle troughs.[41] This countercyclical pattern in expected returns is very much in line with the time-varying risk aversion explanation of return predictability in Fama and French (1989), Campbell and Cochrane (1999), and Cochrane (2007, 2011), among others: economic agents have relatively low risk aversion at the end of economic expansions—and therefore require a lower equity risk premium—but agents' risk aversion increases during contractions as income and consumption levels fall—necessitating a higher equity risk premium. While the behavior of the diffusion index forecasts also conforms to this pattern during the recent Great Recession, the behavior of the POOL-AVG, POOL-DMSFE, and sum-of-the-parts forecasts diverges somewhat from this pattern during this period, as the latter three forecasts typically decline over the course of the recession and subsequently increase sharply after the cyclical trough in 2009:06. This could be due to the severe disruptions in financial markets and unprecedented policy interventions associated with the Global Financial Crisis and very weak economic recovery from the Great Recession.

The last four columns of Table 2, Panel A show that the out-of-sample gains, in terms of both the R^2_{OS} statistics and average utility gains, are concentrated during recessions for the POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the-parts forecasts. The R^2_{OS} statistics are positive for three of these four forecasts during expansions (the exception is the diffusion index), but none of these is significant at conventional levels. The R^2_{OS} statistics range from 1.10% (POOL-AVG) to 4.15% (diffusion index) during recessions, and all of these statistics are significant at the 1%

[41] Some of the bivariate predictive regression forecasts in Figure 2 also exhibit this pattern, in particular, $\log(DP)$, $\log(DY)$, and TMS.

level, despite the reduced number of available observations. The average utility gain during recessions is a very substantial 16.42% for the diffusion index. The economically plausible behavior of the POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the-parts forecasts over the business cycle accords with the substantial out-of-sample gains associated with these forecasts during recessions.[42]

Panel B of Table 2 reports results for forecasts based on multiple economic variables with nonnegativity restrictions, where we set the forecast to zero if a given method produces a negative forecast. By stabilizing the relatively volatile kitchen sink and SIC forecasts, the nonnegativity restrictions substantially increase the R^2_{OS} statistics for these forecasts, although they remain well below zero. The nonnegativity constraints are never binding for the POOL-AVG and POOL-DMSFE forecasts (see Figure 4, Panels C and D), so that the R^2_{OS} statistics are identical for these forecasts across Panels A and B. For the diffusion index and sum-of-the-parts forecasts, the nonnegativity restrictions only lead to slight increases in the R^2_{OS} statistics. Overall, nonnegativity restrictions have limited impact on the performance of the POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the-parts forecasts, presumably due to the stabilizing nature of these strategies.[43]

While positive, the R^2_{OS} statistics for the POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the-parts forecasts in the second column of Table 2 seem small at first blush. As discussed in Section 2, however, theory predicts that the predictable component in the monthly equity premium will be small. Moreover, the monthly R^2_{OS} statistics for these four forecasts, while below 1%, still represent excessive stock return predictability from the standpoint of leading asset pricing models, making them economically relevant.

The average utility gains in Table 2 point to substantial economic value for the POOL-AVG, POOL-DMSE, diffusion index, and sum-of-the-parts forecasts. Nevertheless, an important word

[42] Of course, we identify enhanced return predictability during recessions *ex post*, since the NBER dates business-cycle peaks and troughs retrospectively. As discussed in Section 3.1.4, Markov-switching models provide a natural framework for switching between forecasting models according to estimated probabilities of the state of the economy, helping to exploit enhanced return predictability during recessions in real time. Another possibility is to rely on a real-time index of business conditions, such as Aruoba et al. (2009), to guide switching between forecasts over the business cycle.

[43] Since the asset allocation exercise places a lower bound of zero on the equity portfolio weight, the average utility gains are identical across Panels A and B of Table 2.

of caution is in order. The mean-variance investor for whom we compute utility gains is assumed to have a constant relative risk aversion coefficient over time. To the extent that return predictability is generated by time-varying risk aversion over the business cycle, the utility gains computed under the assumption of constant risk aversion only obtain for a "small" (i.e., nonrepresentative) investor who does not affect market prices. In essence, a nonrepresentative investor can exploit the return predictability created by the time-varying risk aversion of the representative investor.[44]

3.3.2 U.S. Equity Premium Forecastability at Longer Horizons

Since the literature on in-sample return predictability frequently analyzes predictability at longer horizons, we also compute R^2_{OS} statistics and average utility gains for quarterly and annual (nonoverlapping) forecasts.[45] Results for quarterly equity premium forecasts based on individual bivariate prediction regressions are reported in Table 3. The R^2_{OS} statistics for the unrestricted predictive regression forecasts in Panel A are reasonably similar to the corresponding statistics for the monthly forecasts in Table 1. The average utility gains in Table 3, Panel A are often more sizable than those in Table 1, Panel A, so that the economic significance of out-of-sample return predictability appears stronger at we move from a monthly to quarterly horizon. Following the pattern in Table 1, return predictability at the quarterly horizon is concentrated during business-cycle recessions for the valuation ratios and TMS. When we impose Campbell and Thompson (2008) restrictions in Table 3, Panel B, $\log(DP)$ and $\log(DY)$ both have positive and significant R^2_{OS} statistics, and the R^2_{OS} for $\log(DY)$ is well above 1%.

TABLE 3 HERE

Table 4 reports results for quarterly forecasts based on multiple economic variables. Panel A (B) reports results for unrestricted forecasts (forecasts with nonnegativity restrictions imposed). The R^2_{OS} statistics and average utility gains for the quarterly kitchen sink and SIC forecasts in both

[44]This brings to mind Warren Buffett's well-known quote: "We simply attempt to be fearful when others are greedy and to be greedy only when others are fearful."

[45]Quarterly and annual data are also available from Amit Goyal's web page at http://www.hec.unil.ch/agoyal/.

panels of Table 4 deteriorate relative to the corresponding monthly values in Table 2. In contrast, the R^2_{OS} statistics and average utility gains increase for the POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the-parts forecasts as we move from a monthly to quarterly horizon (and the R^2_{OS} statistics remain significant). In both panels of Table 4, the R^2_{OS} statistics for the POOL-DMSFE, diffusion index, and sum-of-the-parts forecasts are now all greater than 1%, and the latter two are above 2%. Similarly to Table 2, imposing nonnegativity restrictions on the quarterly forecasts substantially improves the R^2_{OS} statistics for the kitchen sink and SIC forecasts, although the R^2_{OS} statistics both remain well below zero, while the restrictions have little effect on the POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the-parts forecasts.

TABLE 4 HERE

Table 5 reports results for individual bivariate prediction regression forecasts (Panel A), as well as forecasts based on multiple economic variables (Panel B), at an annual horizon. Among the unrestricted individual forecasts, only $\log(DP)$ has a positive R^2_{OS} (1.76%). Campbell and Thompson (2008) restrictions generally raise the R^2_{OS} statistics for the individual forecasts—the R^2_{OS} increases to 3.51% for $\log(DP)$—although the majority remain negative. Nearly all of the the average utility gains for the individual forecasts increase as we move from Table 3 to Table 5, so that out-of-sample return predictability again becomes more economically significant as the forecasting horizon lengthens.

TABLE 5 HERE

The R^2_{OS} statistics for the kitchen sink and SIC forecasts are substantially lower in Table 5 vis-á-vis Table 4, indicating that the accuracy of these forecasts continues to deteriorate as the horizon lengthens. The kitchen sink and SIC forecasts also produce negative average utility gains at an annual horizon. In line with the previous pattern, the R^2_{OS} statistics for the POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the parts forecasts increase as we move from a quarterly horizon in Table 4 to an annual horizon in Table 5. The R^2_{OS} statistics for both the diffusion index and sum-of-the parts forecasts are now well above 5% (and the R^2_{OS} for the former is above 7%). The average utility gains fall somewhat for the POOL-AVG, POOL-DMSFE, and sum-of-the parts forecasts in Table 5 relative to Table 4, while the average utility gain increases to a very sizable 4.89% for the diffusion index forecast in Table 5.

Focusing on the POOL-AVG, POOL-DMSFE, diffusion index, and sum-of-the-parts forecasts, larger R^2_{OS} statistics—and, hence, greater out-of-sample return predictability according to the MSFE criterion—are available at longer horizons. This increase in out-of-sample return predictability accords with in-sample results from the literature. The theoretical bounds on return predictability discussed in Section 2, however, also increase as the horizon lengthens, so that new economic information is not necessarily available at longer horizons.

3.4 Data Snooping

Data-snooping concerns naturally arise when considering a large number of potential predictors of stock returns. Lo and MacKinlay (1990b), Foster et al. (1997), and Ferson et al. (2003) analyze data snooping in the context of in-sample tests of stock return predictability. White (2000) develops a "reality check" bootstrap to control for data snooping when testing whether any forecast from a group of competing forecasts significantly outperforms a benchmark forecast. White's (2000) procedure is based on a test statistic that is a function of the maximum of the average loss differentials between the benchmark forecast and each competing forecast. A p-value for the maximum statistic is computed via a nonparametric stationary bootstrap that resamples from the original time series of loss differentials.[46] As emphasized by Clark and McCracken (2012), however, the asymptotic properties of the nonparametric stationary bootstrap do not generally apply when all of the competing forecasting models nest the benchmark model. This is relevant for testing out-of-sample stock return predictability. It is often the case that all of the competing forecasts essentially nest the benchmark, since the benchmark typically corresponds to a constant expected return model that excludes the information in the predictors that appear in the competing forecasts, similarly to

[46] Hansen (2005) develops a more powerful refinement of the White (2000) reality check.

our application in Section 3.3.

Inoue and Kilian (2004) derive the asymptotic distribution of the maximum Diebold and Mariano (1995) and West (1996) statistic,

$$\max DMW = \max_{i=1,\ldots,K} DMW_i, \tag{48}$$

where DMW_i is given by (36), under the null hypothesis that $\beta_i = 0$ in (7) for all $i = 1,\ldots,K$. This provides an asymptotic framework for controlling for data snooping when testing for out-of-sample stock return predictability. While Inoue and Kilian (2004) focus on asymptotic results, they observe that potential finite-sample size distortions can be handled in practice by bootstrap methods. Using annual data for 1927–1999, Rapach and Wohar (2006b) use a data-snooping-robust bootstrap procedure to test whether nine popular economic variables evince significant out-of-sample predictive ability for S&P 500 and CRSP equal-weighted aggregate returns for the 1964–1999 forecast evaluation period. Rapach and Wohar (2006b) compute bootstrapped critical values for

$$\max MSE\text{-}F = \max_{i=1,\ldots,K} MSE\text{-}F_i, \tag{49}$$

where $MSE\text{-}F_i = n_2^{0.5}\bar{d}_i/\text{MSFE}_i$ is a more powerful version of DMW_i from Clark and McCracken (2001). To implement the bootstrap, a pseudo sample of stock return observations matching the original sample size is generated under the null of no predictability by estimating (7) with $\beta_i = 0$ and resampling (with replacement) from the fitted residuals. Autoregressive (AR) processes are estimated for each of the predictors, $x_{i,t}$ ($i = 1,\ldots,K$), and a pseudo sample of predictor observations is built up by resampling from the fitted AR process residuals. Importantly, the residuals for the return and predictor processes are drawn in tandem, thereby preserving the contemporaneous correlations in the original data. The maximum statistic is computed for the pseudo sample and stored. Repeating this process many times, bootstrapped critical values are computed from the empirical distribution of maximum statistics. Rapach and Wohar (2006b) find that maximum statistics are significant at conventional levels according to the bootstrapped critical values, signaling that out-of-sample stock return predictability is reasonably robust to data snooping.

Clark and McCracken (2012) recently prove that a wild fixed-regressor bootstrap delivers asymptotically valid critical values for the $\max DMW$ and $\max MSE\text{-}F$ statistics when comparing multiple forecasts that nest a benchmark. The wild fixed-regressor bootstrap accommodates conditional heteroskedasticity and performs well in finite-sample simulations. The bootstrap is straightforward to implement and provides a theoretically justified bootstrap procedure for controlling for data snooping in tests of out-of-sample stock return predictability.

To implement the wild fixed-regressor bootstrap, a general model that includes all possible predictors is estimated. A pseudo sample of return observations is generated under the null of no predictability by setting $r_t^* = \hat{\alpha}_{0,T} + \eta_t \hat{\varepsilon}_t$, where r_t^* is the pseudo observation for r_t, $\hat{\alpha}_{0,T}$ is the sample mean of r_t, η_t is a draw from the standard normal distribution, and $\hat{\varepsilon}_t$ is the fitted residual from the general model. Simulating the disturbance term using $\eta_t \hat{\varepsilon}_t$ makes this a wild bootstrap. The predictors, $x_{i,t}$ $(i = 1, \ldots, K)$, from the original sample also serve as the observations for the predictors in the pseudo sample, making this a fixed-regressor bootstrap. By generating a large number of pseudo samples for r_t and storing the maximum statistics for each pseudo sample, an empirical distribution of maximum statistics is built up that can be used to compute critical values or a p-value for the maximum statistic corresponding to the original sample.

As mentioned in Section 3.1.3, Neely et al. (2012) calculate monthly U.S. equity premium forecasts using individual bivariate predictive regression models based on 14 economic variables and 14 technical indicators, as well as a diffusion index forecast, (23), that employs principal components to extract a small number of factors from the 28 predictors. Since they consider a large number of forecasting models, each of which nests the historical average benchmark, Neely et al. (2012) use the Clark and McCracken (2012) wild fixed-regressor bootstrap to assess the significance of the $\max MSE\text{-}F$ statistic. The bootstrapped p-value is 4.71%, so that the significant evidence of equity premium predictability is not readily ascribed to data snooping.

4 Stock Return Forecastability Along Other Dimensions

4.1 International Stock Returns

While the literature on aggregate stock return predictability focuses on U.S. returns, a number of studies investigate return predictability for other countries. Papers examining stock return predictability for countries outside of the United States include Cutler et al. (1991), Harvey (1991), Bekaert and Hodrick (1992), Campbell and Hamao (1992), Ferson and Harvey (1993), Solnik (1993), Rapach et al. (2005), Ang and Bekaert (2007), Cooper and Priestley (2009), Della Corte et al. (2010), Hjalmarsson (2010), Kellard et al. (2010), and Henkel et al. (2011). All of these studies estimate in-sample predictive regressions for individual country returns using a variety of domestic and/or U.S. economic variables as predictors. The consensus from these studies is that stock returns are predictable worldwide. In an examination of lead-lag relationships in monthly international stock returns—in the spirit of studies investigating such relationships in portfolios of individual U.S. stocks sorted on size, analyst coverage, volume, and/or industry (e.g., Lo and MacKinlay, 1990a; Brennan et al., 1993; Chordia and Swaminathan, 2000; Hou, 2007)—Rapach et al. (2012) find that lagged U.S returns predict non-U.S. returns, but that the reverse generally does not hold.

Fewer studies consider out-of-sample tests of international stock return predictability. Among the previously cited studies, Solnik (1993), Rapach et al. (2005), Cooper and Priestly (2009), Della Corte et al. (2010), Hjalmarrson (2010), Kellard et al. (2010), Henkel et al. (2011), and Rapach et al. (2012) conduct out-of-sample tests. Solnik (1993) forms predictive regression forecasts of monthly aggregate stock returns for eight developed countries using domestic dividend yields, short- and long-term nominal interest rates, and a January dummy as predictors. He finds that a dynamic trading strategy that allocates across stocks in the eight countries based on the predictive regression forecasts significantly outperforms a strategy that assumes constant expected returns.

Rapach et al. (2005) investigate out-of-sample stock return predictability for 12 developed countries using up to 10 domestic economic variables as predictors for each country. Nominal interest rates exhibit the most consistent forecasting ability across the 12 countries. Cooper and Priestley (2009) test out-of-sample monthly stock return predictability for the United States and the other G-7 countries using predictive regression forecasts based on domestic output gaps. They detect significant evidence of return forecastability using the R^2_{OS} and $MSE\text{-}F_i$ statistics, and they find positive average utility gains for a mean-variance investor with relative risk aversion coefficient of three.

Della Corte et al. (2010) analyze the forecasting power of Lettau and Ludvigson's (2001) consumption-wealth ratio in long spans of annual data for stock returns in the United States, the United Kingdom, Japan, and France. After accounting for "look-ahead" bias in the estimation of the long-run consumption-wealth relationship (Brennan and Xia, 2005; Hahn and Lee, 2006), they fail to find significant evidence of out-of-sample return predictability using the R^2_{OS} and $MSE\text{-}F_i$ statistics. According to a variety of performance measures, however, portfolios formed from return forecasts that incorporate information from the the consumption-wealth ratio substantially outperform portfolios that ignore the consumption-wealth ratio. This again illustrates the importance of supplementing conventional statistical measures of forecast accuracy with direct profit- and/or utility-based metrics. Again highlighting the value of economically motivated model restrictions, Della Corte et al. (2010) also find that imposing restrictions on predictive regression forecasts á la Campbell and Thompson (2008) improves forecasting performance.

Hjalmarsson (2010) analyzes bivariate predictive regression forecasts of monthly stock returns for a large number of primarily developed countries, where the individual bivariate predictive regressions use the dividend-price ratio, earnings-price ratio, short-term nominal interest rate, and term spread as predictors. Overall, Hjalmarrson (2010) finds that the interest rate variables, especially the term spread, display the most out-of-sample predictive ability. Interestingly, Hjalmarsson (2010) shows that forecasts generated from pooled predictive regressions that impose slope homogeneity restrictions typically produce higher R^2_{OS} statistics than conventional predictive regression forecasts. While the slope homogeneity restrictions are unlikely to be literally true, their imposition permits more efficient parameter estimation that can lead to improved forecasting performance in terms of MSFE. This is another example of the usefulness of sensible model restrictions for forecasting stock returns.

Kellard et al. (2010) compare stock return predictability in the United States and United Kingdom on the basis of dividend-price ratios. They find that the dividend-price ratio exhibits stronger out-of-sample forecasting ability in terms of MSFE in the United Kingdom vis-á-vis the United States, and they attribute the difference to the higher proportion of dividend-paying firms in the United Kingdom. As discussed in Section 3.1.4, Henkel et al. (2011) document strong evidence of out-of-sample stock return predictability in the United States using a regime-switching predictive regression based on popular economic variables from the literature. However, they find weaker evidence that regime switching produces out-of-sample gains in other G-7 countries.

Finally, Rapach et al. (2012) find that lagged U.S. stock returns have substantial out-of-sample predictive power for returns in non-U.S. developed countries. Bivariate predictive regression forecasts based on lagged U.S. returns deliver monthly R^2_{OS} statistics of up to nearly 4% for non-U.S. returns over the 1985:01–2010:12 out-of-sample period, and lagged U.S. returns produce especially sizable out-of-sample gains during the recent Global Financial Crisis. While lagged U.S. returns evince predictive power for non-U.S. returns, Rapach et al. (2012) also find that lagged non-U.S. returns display little predictive ability for U.S. returns, pointing to a leading role for the United States in the international equity market.

4.2 Cross-Sectional Stock Returns

In addition to U.S. aggregate stock returns, an ample literature examines return predictability for component portfolios of the aggregate market, including portfolios sorted by market capitalization, book-to-market value, and industry. Ferson and Harvey (1991, 1999), Ferson and Korajczyk (1995), and Kirby (1998), among others, estimate in-sample predictive regressions for component portfolios based on the same types of predictors used in studies of aggregate market return predictability.

Analysis of out-of-sample return predictability for component portfolios, however, is relatively rare. Avramov (2002) adopts a Bayesian approach to investigate return predictability for six size/value-sorted portfolios based on 14 popular economic variables from the literature. Avramov (2002) employs Bayesian model averaging to account for model uncertainty, and he finds that Bayesian model averaging forecasts of monthly component portfolio returns outperform forecasts based on the constant expected component portfolio return assumption in terms of MSFE. In addition, asset allocation exercises reveal sizable utility gains for investors who incorporate return predictability, although it is crucial to account for model uncertainty in the asset allocation decision to realize these gains. Subsequently, Avramov (2004) also uses a Bayesian approach to investigate optimal asset allocation across the 25 Fama-French size/value-sorted portfolios and industry portfolios, respectively, with a smaller set of five economic variables serving as return predictors. For a wide range of prior beliefs, allowing for time variation in monthly expected component portfolio returns produces substantial out-of-sample asset allocation gains relative to assuming constant expected component portfolio returns.

Kong et al. (2011) compute combination forecasts of monthly returns for the 25 Fama-French portfolios from individual bivariate predictive regression forecasts based on the 14 economic variables from Section 3.3 and lagged size/value-sorted portfoilo returns. R^2_{OS} statistics are positive for all 25 of the combination forecasts of component portfolio returns, and 22 of the R^2_{OS} statistics are significant at the 5% level according to the Clark and West (2007) test. Furthermore, out-of-sample component return predictability is notably stronger during business-cycle recessions vis-á-vis expansions, similarly to the situation for U.S. aggregate market returns. Return forecastability is also substantially stronger for portfolios comprised of small, value firms; for example, the monthly R^2_{OS} is 5.73% (0.36%) for the S1/BM5 (S5/BM1) portfolio comprised of firms with the lowest market capitalization and highest book-to-market value (highest market capitalization and lowest book-to-market value). In asset allocation exercises, Kong et al. (2011) show that a variety of component-rotation portfolios based on combination forecasts of component returns outperform portfolios based on historical average forecasts of component returns. Component-rotation portfolios based on the combination forecasts also exhibit significant alpha (after controlling for the three Fama-French factors), as well as significant timing ability according to Lo's (2008) test.[47]

Rapach et al. (2011) compute diffusion index forecasts of 33 industry portfolio returns by extracting the first two principal components from the same 14 economic variables and lagged industry portfolio returns. R^2_{OS} statistics are positive for nearly all industries, and 26 are significant at conventional levels based on the Clark and West (2007) test. Textiles, apparel, furniture, printing and publishing, and transportation equipment are among the most predictable industries, with monthly R^2_{OS} statistics above 3%. In addition, industry-rotation portfolios that utilize diffusion index forecasts of industry returns generate sizable utility gains relative to historical average forecasts for an investor with power utility.

Another interesting forecasting application in the cross-sectional domain is Han et al. (2012), who examine the use of technical indicators to inform asset allocation decisions across portfolios sorted by volatility. They find that moving-average rules generate investment timing portfolios that substantially outperform a buy-and-hold strategy. Furthermore, the timing portfolios have negative or little risk exposures to the three Fama-French factors, signaling abnormal returns from the perspective of the Fama-French three-factor model, with annual alphas exceeding 20% for high-volatility portfolios.

5 Conclusion

The key themes of this chapter can be summarized as follows:

- Theory tells us that the predictable component in stock returns will be small and that monthly R^2_{OS} statistics below 1% can be economically relevant. Forecasting models that purport to explain a large portion of stock return fluctuations imply substantial risk-adjusted abnormal returns and are simply too good to be true.

- As forcefully demonstrated by Goyal and Welch (2008), conventional predictive regression

[47] Although not employing a predictive regression framework *per se*, Tu (2010) finds that responding to regime switching between "bull" and "bear" markets substantially improves portfolio performance for an investor allocating across the 25 Fama-French portfolios.

forecasts of stock returns fail to consistently outperform the simple historical average forecast in terms of MSFE. Model uncertainty and parameter instability render conventional predictive regression forecasts unreliable.

- Recently proposed strategies significantly improve upon conventional predictive regression forecasts. These procedures, which improve forecasting performance by accommodating model uncertainty and parameter instability, include economically motivated model restrictions, forecast combination, diffusion indices, and regime shifts.

- Inferences concerning out-of-sample stock return predictability typically involve comparisons of nested forecasts. Unless statistical tests designed for nested forecast comparisons are used, significant evidence of stock return forecastability can easily be missed.

- It is important to supplement conventional statistical criteria of forecast accuracy with direct profit- or utility-based criteria, since the two types of measures are not necessarily strongly related. In particular, utility-based measures can indicate clear economic significance, even if conventional statistical measures fail to detect out-of-sample gains.

- Stock return forecastability is strongly linked to business-cycle fluctuations, with a substantially greater degree of forecastability evident during recessions vis-á-vis expansions.

- In addition to the U.S. aggregate market, there is significant out-of-sample evidence of stock return predictability for countries outside of the United States, as well as component portfolios of the U.S. market. Exploiting the out-of-sample predictability in international and cross-sectional returns can produce sizable utility gains from an asset allocation perspective.

We conclude by suggesting avenues for future research. The literature on stock return forecasting primarily relies on popular economic variables as predictors. However, other variables that potentially contain relevant information for forecasting stock returns have received less attention. Such variables include options, futures, and other derivative prices; microstructure measures of liquidity; and institutional trading variables such as trading volumes and money flows for mutual and hedge funds. Furthermore, recent studies find significant in-sample evidence of a positive relationship between expected returns and risk (e.g., Guo and Whitelaw, 2006; Lanne and Saikkonen, 2006; Lundblad, 2007; Bali, 2008). It would be interesting to examine whether these approaches could be used to generate reliable out-of-sample stock return forecasts based on the expected return-risk relationship; Ludvigson and Ng (2007) report promising results in this direction. Finally, learning appears to play an important role in stock return predictability (e.g., Timmermann, 1993, 1996; Pástor and Veronesi, 2009). Theoretical models that explain how investors form return forecasts in light of available information and respond to their forecasting errors serve as promising building blocks for forecasting models based on learning.

References

Aït-Sahalia, Y., Brandt, M.W. (2001). "Variable selection for portfolio choice." Journal of Finance 56, 1297–1351.

Alexander, S. (1961). "Price movement in speculative markets: trends or random walks." Industrial Management Review 2, 7–26.

Alexander, S. (1964). "Price movement in speculative markets: trends or random walks, number 2." Industrial Management Review 5, 25–46.

Amihud, Y., Hurvich, C.M. (2004). "Predictive regressions: a reduced-bias estimation method." Journal of Financial and Quantitative Analysis 39, 813–841.

Amihud, Y., Hurvich, C.M., Wang, Y. (2009). "Multi-predictor regressions: hypothesis testing." Review of Financial Studies 22, 413–434.

Andersen, T.G., Bollerslev, T., Christoffersen, P.C., Diebold, F.X. (2006). "Volatility and Correlation Forecasting." In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam.

Ang, A., Bekaert, G. (2002). "International asset allocation with regime shifts." Review of Financial Studies 15, 1137–1187.

Ang, A., Bekaert, G. (2007). "Return predictability: is it there?" Review of Financial Studies 20, 651–707.

Ang, A., Timmermann, A. (2012). "Regimes changes and financial markets." Annual Review of Financial Economics, forthcoming.

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

Avramov, D. (2002). "Stock return predictability and model uncertainty." Journal of Financial Economics 64, 423–458.

Avramov, D. (2004). "Stock return predictability and asset pricing models." Review of Financial Studies 17, 699–738.

Bai, J. (2003). "Inferential theory for factor models of large dimensions." Econometrica 71, 135–

171.

Bai, J., Ng, S. (2002). "Determining the number of factors in approximate factor models." Econometrica 70, 191–221.

Bai, J., Ng, S. (2008). "Forecasting economic time series using targeted predictors." Journal of Econometrics 146, 304–317.

Bai, J., Perron, P. (1998). "Estimating and testing linear models with multiple structural changes." Econometrica 66, 47–78.

Bai, J., Perron, P. (2003). "Computation and analysis of multiple structural change models." Journal of Applied Econometrics 18, 1–22.

Baker, M., Wurgler, J. (2000). "The equity share in new issues and aggregate stock returns." Journal of Finance 55, 2219–2257.

Bali, T.G. (2008). "The intertemporal relation between expected returns and risk." Journal of Financial Economics 87, 101–131.

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

Barberis, N. (2000). "Investing for the long run when returns are predictable." Journal of Finance 55, 225–264.

Batchelor, R. (2011). "Accuracy versus profitability." Foresight: The International Journal of Applied Forecasting 21, 10–15.

Bates, J.M., Granger, C.W.J. (1969). "The combination of forecasts." Operational Research Quarterly 20, 451–468.

Bekaert, G., Hodrick, R.J. (1992). "Characterizing predictable components in excess returns on equity and foreign exchange markets." Journal of Finance 47, 467–509.

van Binsbergen, J.H., Koijen, R.S.J. (2010). "Predictive regressions: a present-value approach." Journal of Finance 65, 1439–1471.

Bossaerts, P., Hillion, P. (1999). "Implementing statistical criteria to select return forecasting models: what do we learn?" Review of Financial Studies 12, 405–428.

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

Boudoukh, J., Richardson, M.P., Whitelaw, R.F. (2008). "The myth of long-horizon predictability." Review of Financial Studies 21, 1577–1605.

Breen, W., Glosten, L.R., Jagannathan, R. (1989). "Economic significance of predictable variation in stock index returns." Journal of Finance 44, 1177–1189.

Brennan, M.J., Jegadeesh, N., Swaminathan, B. (1993). "Investment analysis and the adjustment of stock prices to common information." Review of Financial Studies 6, 799–824.

Brennan, M.J., Xia, Y. (2005). "*tay*'s as good as *cay*." Finance Research Letters 2, 1–14.

Britten-Jones, M., Neuberger, A., Nolte, I. (2011). "Improved inference in regression with overlapping observations." Journal of Business, Finance, and Accounting 38, 657–683.

Brock, W., Lakonishok, J., LeBaron, B. (1992). "Simple technical trading rules and the stochastic properties of stock returns." Journal of Finance 47, 1731–1764.

Brown, S.J., Goetzmann, W.N., Kumar, A. (1998). "The Dow Theory: William Peter Hamilton's track record reconsidered." Journal of Finance 53, 1311–1333.

Butler, A.W., Grullon, G., Weston, J.P. (2005). "Can managers forecast aggregate market returns?" Journal of Finance 60, 963–986.

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

Campbell, J.Y. (2000). "Asset pricing at the millennium." Journal of Finance 55, 1515–1567.

Campbell, J.Y. (2008). "Viewpoint: estimating the equity premium." Canadian Journal of Economics 41, 1–21.

Campbell, J.Y., Cochrane, J.H. (1999). "By force of habit: a consumption-based explanation of aggregate stock market behavior." Journal of Political Economy 107, 205–251.

Campbell, J.Y., Hamao, Y. (1992). "Predictable stock returns in the United States and Japan: a study of long-term capital market integration." Journal of Finance 47, 43–69.

Campbell, J.Y., Shiller, R.J. (1988a). "The dividend-price ratio and expectations of future dividends and discount factors." Review of Financial Studies 1, 195–228.

Campbell, J.Y., Shiller, R.J. (1988b). "Stock prices, earnings, and expected dividends." Journal of Finance 43, 661–676.

Campbell, J.Y., Shiller, R.J. (1998). "Valuation ratios and the long-run stock market outlook." Journal of Portfolio Management 24, 11–26.

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

Campbell, J.Y., Vuolteenaho, T. (2004). "Inflation illusion and stock prices." American Economic Review 94, 19–23.

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: a half century of direct evidence." Journal of Business and Economic Statistics 27, 266–278.

Cavanaugh, C.L., Elliott, G., Stock, J.H. (1995). "Inference in models with nearly nonstationary regressors." Econometric Theory 11, 1131–1147.

Cenesizoglu, T., Timmermann, A. (2011). "Do return prediction models add economic value?" Manuscript, HEC Montreal and University of California at San Diego.

Chen, Q., Hong, Y. (2010). "Predictability of equity returns over different time horizons: a nonparametric approach." Manuscript, Cornell University.

Chordia, T., Swaminathan, B. (2000). "Trading volume and cross-autocorrelations in stock returns." Journal of Finance 55, 913–935.

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

Clark, T.E., McCracken, M.W. (2005). "The power of tests of predictive ability in the presence of structural breaks." Journal of Econometrics 124, 1–31.

Clark, T.E., McCracken, M.W. (2009). "Improving forecast accuracy by combining recursive and rolling forecasts." International Economic Review 50, 363–395.

Clark, T.E., McCracken, M.W. (2012). “Reality checks and comparisons of nested predictive models.” Journal of Business and Economic Statistics 30, 53–66.

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

Clements, M.P., Hendry, D.F. (2006). “Forecasting with breaks.” In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam.

Cochrane, J.H. (2005). Asset Pricing (Revised Edition). Princeton University Press, Princeton, N.J.

Cochrane, J.H. (2007). “Financial markets and the real economy.” In: Mehra, R. (Ed.), Handbook of the Equity Premium. Elsevier, Amsterdam.

Cochrane, J.H. (2008). “The dog that did not bark: a defense of return predictability.” Review of Financial Studies 21, 1533–1575.

Cochrane, J.H. (2011). “Presidential address: discount rates.” Journal of Finance 66, 1047–1108.

Cooper, I., Priestly, R. (2009). “Time-varying risk premiums and the output gap.” Review of Financial Studies 22, 2801–2833.

Cootner, P.H. (1962). “Stock prices: random vs. systematic changes.” Industrial Management Review 3, 24–45.

Corradi, V., Swanson, N.R.. “Predictive density evaluation.” In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam.

Cowles, A. (1933). “Can stock market forecasters forecast?” Econometrica 1, 309–324.

Cremers, K.J.M. (2002). “Stock return predictability: a Bayesian model selection perspective.” Review of Financial Studies 15, 1223–1249.

Cumby, R.E., Modest, D.M. (1987). “Testing for market timing ability: a framework for forecast evaluation.” Journal of Financial Economics 19, 169–189.

Cutler, D.M., Poterba, J.M., Summers, L.H. (1991). “Speculative dynamics.” Review of Economic Studies 58, 529–546.

Dangl, T., Halling, M. (2012). “Predictive regressions with time-varying coefficients.” Journal of

Financial Economics, forthcoming.

Della Corte, P., Sarno, L., Valente, G. (2010). "A century of equity premium predictability and the consumption-wealth ratio: an international perspective." Journal of Empirical Finance 17, 313–331.

Diebold, F.X., Mariano, R.S. (1995). "Comparing predictive accuracy." Journal of Business and Economic Statistics 13, 253–263.

Driesprong, G., Jacobsen, B., Maat, B. (2008). "Striking oil: another puzzle?" Journal of Financial Economics 89, 307–327.

Elliott, G., Müller, U. (2006). "Efficient tests for general persistent time variation in regression coefficients." Review of Economic Studies 73, 907–940.

Fama, E.F. (1970). "Efficient capital markets: a review of theory and empirical work." Journal of Finance 25, 383–417.

Fama, E.F. (1991). "Efficient capital markets: II." Journal of Finance 46, 1575–1617.

Fama, E.F., Blume, M.E. (1966). "Filter rules and stock-market trading." Journal of Business 39, 226–241.

Fama, E.F., French, K.R. (1988). "Dividend yields and expected stock returns." Journal of Financial Economics 22, 3–25.

Fama, E.F., French, K.R. (1989). "Business conditions and expected returns on stocks and bonds." Journal of Financial Economics 25, 23–49.

Fama, E.F, Schwert, G.W. (1977). "Asset returns and inflation." Journal of Financial Economics 5, 115–146.

Ferreira, M.I., Santa-Clara, P. (2011). "Forecasting stock market returns: the sum of the parts is more than the whole." Journal of Financial Economics 100, 514–537.

Ferson, W.E., Harvey, C.R. (1991). "The variation in economic risk premiums." Journal of Political Economy 99, 385–415.

Ferson, W.E., Harvey, C.R. (1993). "The risk and predictability of international equity returns." Review of Financial Studies 6, 527–566.

Ferson, W.E., Harvey, C.R. (1999). “Conditioning variables and the cross section of stock returns.” Journal of Finance 54, 1325–1360.

Ferson, W.E., Korajczyk, R.A. (1995). “Do arbitrage pricing models explain the predictability of stock returns?” Journal of Business 68, 309–349.

Ferson, W.E., Sarkissian, S., Simin, T.T. (2003). “Spurious regressions in financial economics?” Journal of Finance 58, 1393–1413.

Foster, F.D., Smith, T., Whaley, R.E. (1997). “Assessing goodness-of-fit of asset pricing models: the distribution of the maximum R^2.” Journal of Finance 52, 591–607.

Franses, P.H., van Dijk, D. (2000). Nonlinear Time Series Models in Empirical Finance. Cambridge University Press, Cambridge, United Kingdom.

Giacomini, R., Rossi, B. (2009). “Detecting and predicting forecast breakdowns.” Review of Economic Studies 76, 669–705.

Giacomini, R., Rossi, B. (2010). “Forecast comparisons in unstable environments.” Journal of Applied Econometrics 25, 595–620.

Goetzmann, W.N., Jorion, P. (1993). “Testing the predictive power of dividend yields.” Journal of Finance 48, 663–679.

González-Rivera, G., Lee, T.-H. (2009). “Nonlinear time series in financial forecasting.” In: Meyer, R.A. (Ed.), Encyclopedia of Complexity and Systems Science. Springer, New York.

Goyal, A., Welch, I. (2003). “Predicting the equity premium with dividend ratios.” Management Science 49, 639–654.

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

Granger, C.W.J., Machina, M.J. (2006). “Forecasting and decision theory.” In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam.

Guidolin, M., Hyde, S., McMillan, D., Ono, S. (2009). “Non-linear predictability in stock and bond returns: when and where is it exploitable?” International Journal of Forecasting 25, 373–399.

Guidolin, M., Timmermann, A. (2007). "Asset allocation under multivariate regime switching." Journal of Economic Dynamics and Control 31, 3503–3544.

Guo, H. (2006). "On the out-of-sample predictability of stock market returns." Journal of Business 79, 645–670.

Guo, H., Whitelaw, R.F. (2006). "Uncovering the risk-return relation in the stock market." Journal of Finance 61, 1433–1463.

Hahn, J., Lee, H. (2006). "Interpreting the predictive power of the consumption-wealth ratio." Journal of Empirical Finance 13, 183–202.

Hamilton, J.D. (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle." Econometrica 57, 357–384.

Han, Y., Yang, K., Zhou, G. (2012). "A new anomaly: the cross-sectional profitability of technical analysis." Journal of Financial and Quantitative Analysis, forthcoming.

Hansen, B.E. (2000). "Testing for structural change in conditional models." Journal of Econometrics 97, 93–115.

Hansen, P.R. (2005). "A test for superior predictive ability." Journal of Business and Economic Statistics 23, 365–380.

Hansen, P.R., Timmermann, A. (2012). "Choice of sample split in out-of-sample forecast evaluation." European University Institute Working Paper ECO 2012/10.

Harvey, C.R. (1991). "The world price of covariance risk." Journal of Finance 46, 111–157.

Hawawini, G., Keim, D.B. (1995). "On the predictability of common stock returns: world-wide evidence." In: Jarrow, R., Maksimovic, V., Ziemba, W. (Eds.), Handbooks in Operations Research and Management Science, vol. 9. Elsevier, Amsterdam.

Hendry, D.F., Clements, M.P. (2004). "Pooling of forecasts." Econometrics Journal 7, 1–31.

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

Henriksson, R.D., Merton, R.C. (1981). "On market timing and investment performance, II: statistical procedures for evaluating forecasting skills." Journal of Business 54, 513–533.

Hirshleifer, D., Hou, K., Teoh, S.H. (2009). "Accruals, cash flows, and aggregate stock returns." Journal of Financial Economics 91, 389–406.

Hjalmarsson, E. (2010). "Predicting global stock returns." Journal of Financial and Quantitative Analysis 45, 49–80.

Hjalmarrson, E. (2012). "New methods for inference in long-horizon regressions." Journal of Financial and Quantitative Analysis, forthcoming.

Hodrick, R.J. (1992). "Dividend yields and expected stock returns: alternative procedures for inference and measurement." Review of Financial Studies 5, 357–386.

Hong, H., Torous, W., Valkanov, R. (2007). "Do industries lead stock markets?" Journal of Financial Economics 83, 367–396.

Hou, K. (2007). "Industry information diffusion and the lead-lag effect in stock returns." Review of Financial Studies 20, 1113–1138.

Inoue, A., Kilian, L. (2004). "In-sample or out-of-sample tests of predictability: which one should we use?" Econometric Reviews 23, 371–402.

Jensen, M.C., Benington, G.A. (1970). "Random walks and technical theories: some additional evidence." Journal of Finance 25, 469–482.

Kan, R., Zhou, G. (2007). "A new variance bound on the stochastic discount factor." Journal of Business 79, 941–961.

Kandel, S., Stambaugh, R.F. (1996). "On the predictability of stock returns: an asset-allocation perspective." Journal of Finance 51, 385–424.

Kellard, N.M., Nankervis, J.C., Papadimitriou, F.I. (2010). "Predicting the equity premium with dividend ratios: reconciling the evidence." Journal of Empirical Finance 17, 539–551.

Kelly, B., Pruitt, S. (2011). "The three-pass regression filter: a new approach to forecasting using many predictors." University of Chicago Booth School of Business Working Paper No. 11-19.

Kelly, B., Pruitt, S. (2012). "Market expectations in the cross section of present values." University of Chicago Booth School of Business Working Paper No. 11-08.

Kirby, C. (1998). "The restrictions on predictability implied by rational asset pricing models."

Review of Financial Studies 11, 343–382.

Koijen, R.S.J., Van Nieuwerburgh, S. (2011). "Predictability of returns and cash flows." Annual Review of Financial Economics 3, 467–491.

Kong, A., Rapach, D.E., Strauss, J.K., Zhou, G. (2011). "Predicting market components out of sample: asset allocation implications." Journal of Portfolio Management 37, 29–41.

Kothari, S.P., Shanken, J. (1997). "Book-to-market, dividend yield, and expected market returns: a time-series analysis." Journal of Financial Economics 44, 169–203.

Lamont, O. (1998). "Earnings and expected returns." Journal of Finance 53, 1563–1587.

Lamoureux, C.G., Zhou, G. (1996). "Temporary components of stock returns: what do the data tell us?" Review of Financial Studies 9, 1033–1059.

Lanne, M., Saikkonen, P. (2006). "Why is it so difficult to uncover the risk-return tradeoff in stock returns?" Economics Letters 92, 118–125.

Leitch, G., Tanner, J.E. (1991). "Economic forecast evaluation: profit versus the conventional error measures." American Economic Review 81, 580–590.

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

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

Lewellen, J. (2004). "Predicting returns with financial ratios." Journal of Financial Economics 74, 209–235.

Lo, A.W. (2004). "Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis." Journal of Investment Consulting 7, 21–44.

Lo, A.W. (2008). "Where do alphas come from? A new measure of the value of active investment management." Journal of Investment Management 6, 1–29.

Lo, A.W., MacKinlay, A.C. (1990a). "When are contrarian profits due to stock market overreaction?" Review of Financial Studies 3, 175–205.

Lo, A.W., MacKinlay, A.C. (1990b). "Data-snooping biases in tests of financial asset pricing models." Review of Financial Studies 3, 431–467.

Lo, A.W., Mamaysky, H., Wang, J. (2000). "Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation." Journal of Finance 55, 1705–1765.

Ludvigson, S.C., Ng, S. (2007). "The empirical risk-return relation: a factor analysis approach." Journal of Financial Economics 83, 171–222.

Lundblad, C. (2007). "The risk return tradeoff in the long run: 1836–2003." Journal of Financial Economics 85, 123–150.

Malkiel, B.G. (1973). A Random Walk Down Wall Street (First Edition). Norton, New York.

Malkiel, B.G. (2011). A Random Walk Down Wall Street (Revised Edition). Norton, New York.

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

McCracken, M.W. (2007). "Asymptotics for out of sample tests of Granger causality." Journal of Econometrics 140, 719–752.

McCracken, M.W., Valente, G. (2011). "Testing the economic value of asset return predictability." Manuscript, University of Essex.

Meese, R.A., Rogoff, K.S. (1983). "Empirical exchange rate models of the seventies: do they fit out of sample?" Journal of International Economics 14, 3–24.

Neely, C.J., Rapach, D.E., Tu, J., Zhou, G. (2012). "Forecasting the equity risk premium: the role of technical indicators." Federal Reserve Bank of St. Louis Working Paper 2010-008E.

Nelson, C.R. (1976). "Inflation and the rates of return on common stock." Journal of Finance 31, 471–483.

Nelson, C.R., Kim, M.J. (1993). "Predictable stock returns: the role of small sample bias." Journal of Finance 48, 641–661.

Onatski, A. (2010). "Determining the number of factors from empirical distributions of eigenvalues." Review of Economics and Statistics 92, 1004–1016.

Park, C.-H., Irwin, S.H. (2007). "What do we know about the profitability of technical analysis?"

Journal of Economic Surveys 21, 786–826.

Pástor, L., Stambaugh, R.F. (2009). "Predictive systems: living with imperfect predictors." Journal of Finance 64, 1583–1628.

Pástor, L., Stambaugh, R.F. (2012). "Are stocks really less volatile in the long run?" Journal of Finance 67, 431–478.

Pástor, L., Veronesi, P. (2009). "Learning in financial markets." Annual Review of Financial Economics 1, 361–381.

Paye, B.S., Timmermann, A. (2006). "Instability of return prediction models." Journal of Empirical Finance 13, 274–315.

Pesaran, M.H., Timmermann, A. (1992). "A simple nonparametric test of predictive performance." Journal of Business and Economic Statistics 10, 461–465.

Pesaran, M.H., Timmermann, A. (1995). "Predictability of stock returns: robustness and economic significance." Journal of Finance 50, 1201–1228.

Pesaran, M.H., Timmermann, A. (2002). "Market timing and return prediction under model instability." Journal of Empirical Finance 9, 495–510.

Pesaran, M.H., Timmermann, A. (2004). "How costly is it to ignore breaks when forecasting the direction of a time series?" International Journal of Forecasting 20, 411–425.

Pesaran, M.H., Timmermann, A. (2007). "Selection of estimation window in the presence of breaks." Journal of Econometrics 137, 134–161.

Pettenuzzo, D., Timmermann, A. (2011). "Predictability of stock returns and asset allocation under structural breaks." Journal of Econometrics 164, 60–78.

Pontiff, J., Schall, L.D. (1998). "Book-to-market ratios as predictors of market returns." Journal of Financial Economics 49, 141–160.

Rangvid, J. (2006). "Output and expected returns." Journal of Financial Economics 81, 595–624.

Rapach, D.E., Strauss, J.K., Tu, J., Zhou, G. (2011). "Out-of-sample industry return predictability: evidence from a large number of predictors." Manuscript, Saint Louis University, Singapore Management University, and Washington University in St. Louis.

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.

Rapach, D.E., Strauss, J.K., Zhou, G. (2012). "International stock return predictability: what is the role of the United States?" Journal of Finance, forthcoming.

Rapach, D.E., Wohar, M.E. (2006a). "Structural breaks and predictive regression models of aggregate U.S. stock returns." Journal of Financial Econometrics 4, 238–274.

Rapach, D.E., Wohar, M.E. (2006b). "In-sample vs. out-of-sample tests of stock return predictability in the context of data mining." Journal of Empirical Finance 13, 231–247.

Rapach, D.E., Wohar, M.E., Rangvid, J. (2005). "Macro variables and international stock return predictability." International Journal of Forecasting 21, 137–166.

Ross, S.A. (2005). Neoclassical Finance. Princeton University Press, Princeton, N.J.

Rossi, B. (2005). "Optimal tests for nested model selection with underlying parameter instabilities." Econometric Theory 21, 962–990.

Rossi, B., Inoue, A. (2012). "Out-of-sample forecast tests robust to the window size choice." Journal of Business and Economic Statistics, forthcoming.

Rozeff, M.S. (1984). "Dividend yields are equity risk premiums." Journal of Portfolio Management 11, 68–75.

Santos, T., Veronesi, P. (2006). "Labor income and predictable stock returns." Review of Financial Studies 19, 1–44.

Solnik, B. (1993). "The performance of international asset allocation strategies using conditioning information." Journal of Empirical Finance 1, 33–55.

Stambaugh, R.F. (1986). "Biases in regressions with lagged stochastic regressors." University of Chicago Graduate School of Business Working Paper No. 156.

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

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

Stock, J.H., Watson, M.W. (2006). "Forecasting with many predictors." In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam.

Subrahmanyam, A. (2010). "The cross-section of expected stock returns: what have we learnt from the past twenty-five years of research?" European Financial Management 16, 27–42.

Sullivan, R., Timmermann, A., White, H. (1999). "Data-snooping, technical trading rule performance, and the bootstrap." Journal of Finance 54, 1647–1691.

Sullivan, R., Timmermann, A., White, H. (2001). "Dangers of data mining: the case of calendar effects in stock returns." Journal of Econometrics 105, 249–286.

Teräsvirta, T. (2006). "Forecasting economic variables with nonlinear models." In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam.

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

Timmermann, A. (1996). "Excess volatility and predictability of stock prices in autoregressive dividend models with learning." Review of Economic Studies 63, 523–557.

Timmermann, A. (2006). "Forecast combinations." In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam.

Timmermann, A. (2008). "Elusive return predictability." International Journal of Forecasting 24, 1–18.

Timmermann, A., Granger, C.W.J. (2004). "Efficient market hypothesis and forecasting." International Journal of Forecasting 20, 15–27.

Torous, W., Valkanov, R., Yan, S. (2005). "On predicting stock returns with nearly integrated explanatory variables." Journal of Business 77, 937–966.

Tu, J. (2010). "Is regime switching in stock returns important in portfolio decisions?" Management Science 56, 1198–1215.

Tu, J., Zhou, G. (2004). "Data-generating process uncertainty: what difference does it make in portfolio decisions? Journal of Financial Economics 72, 385–421.

Valkanov, R. (2003). “Long-horizon regressions: theoretical results and applications.” Journal of Financial Economics 68, 201–232.

Wachter, J.A., Warusawitharana, M. (2009). “Predictable returns and asset allocation: should a skeptical investor time the market?” Journal of Econometrics 148, 162–178.

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

West, K.D. (2006). “Forecast evaluation.” In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam.

White, H. (2000). “A reality check for data snooping.” Econometrica 68, 1097–1126.

White, H. (2006). “Approximate nonlinear forecasting methods.” In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam.

Xu, Y. (2004). “Small levels of predictability and large economic gains.” Journal of Empirical Finance 11, 247–275.

Zhou, G. (2010). “How much stock return predictability can we expect from an asset pricing model?” Economics Letters 108, 184–186.

Table 1 Monthly U.S. equity premium out-of-sample forecasting results based on individual economic variables, 1957:01–2010:12

Economic variable	Overall R^2_{OS} (%)	Overall Δ (annual %)	Expansion R^2_{OS} (%)	Expansion Δ (annual %)	Recession R^2_{OS} (%)	Recession Δ (annual %)
Panel A: Unrestricted predictive regression forecasts						
$\log(DP)$	−0.05 [0.10]	0.87	−1.24 [0.42]	−1.47	2.41 [0.00]	11.87
$\log(DY)$	−0.37 [0.07]	1.18	−2.28 [0.40]	−1.98	3.56 [0.00]	16.17
$\log(EP)$	−1.88 [0.28]	0.57	−2.21 [0.31]	−0.41	−1.20 [0.38]	4.99
$\log(DE)$	−2.04 [0.97]	−0.44	−1.26 [0.80]	0.06	−3.67 [0.97]	−2.73
$SVAR$	0.32 [0.17]	−0.11	−0.02 [0.50]	−0.37	1.01 [0.16]	1.08
BM	−1.74 [0.31]	−0.72	−2.56 [0.44]	−2.01	−0.04 [0.28]	5.12
$NTIS$	−0.91 [0.41]	−0.21	0.50 [0.03]	0.72	−3.82 [0.94]	−4.71
TBL	−0.01 [0.09]	1.53	−0.84 [0.30]	0.24	1.71 [0.10]	7.58
LTY	−1.17 [0.12]	1.29	−2.37 [0.38]	−0.21	1.32 [0.11]	8.38
LTR	−0.08 [0.20]	0.57	−0.85 [0.63]	−0.35	1.52 [0.05]	4.74
TMS	0.06 [0.16]	1.15	−0.40 [0.34]	0.01	1.00 [0.09]	6.49
DFY	−0.04 [0.59]	0.29	−0.06 [0.64]	0.00	−0.01 [0.48]	1.48
DFR	−0.01 [0.38]	0.39	0.12 [0.25]	0.15	−0.28 [0.48]	1.53
$INFL$	−0.09 [0.50]	0.34	0.10 [0.22]	0.19	−0.48 [0.66]	1.16
Panel B: Predictive regression forecasts with Campbell and Thompson (2008) restrictions						
$\log(DP)$	0.15 [0.07]	0.87	−0.92 [0.38]	−1.47	2.36 [0.00]	11.87
$\log(DY)$	0.17 [0.04]	1.18	−1.33 [0.40]	−1.98	3.26 [0.00]	16.17
$\log(EP)$	−0.82 [0.24]	0.57	−1.19 [0.30]	−0.41	−0.06 [0.32]	4.99
$\log(DE)$	−1.74 [0.98]	0.01	−1.19 [0.78]	0.01	−2.88 [0.98]	0.00
$SVAR$	0.00 [–]	−0.26	0.00 [–]	−0.36	0.00 [–]	0.18
BM	−1.17 [0.29]	−0.72	−1.68 [0.40]	−2.01	−0.13 [0.30]	5.12
$NTIS$	−0.91 [0.41]	−0.21	0.50 [0.03]	0.72	−3.82 [0.94]	−4.71
TBL	0.21 [0.10]	1.53	−0.25 [0.27]	0.24	1.16 [0.10]	7.58
LTY	−0.01 [0.09]	1.29	−0.67 [0.29]	−0.21	1.36 [0.07]	8.38
LTR	0.22 [0.12]	0.63	−0.47 [0.52]	−0.36	1.64 [0.03]	5.15
TMS	0.12 [0.15]	1.15	−0.42 [0.37]	0.01	1.23 [0.06]	6.49
DFY	−0.01 [0.50]	0.29	−0.03 [0.55]	0.00	0.01 [0.45]	1.48
DFR	−0.16 [0.49]	0.56	0.09 [0.27]	0.15	−0.68 [0.66]	2.51
$INFL$	−0.06 [0.46]	0.34	0.10 [0.22]	0.19	−0.38 [0.63]	1.16

Notes: R^2_{OS} measures the percent reduction in mean squared forecast error (MSFE) for the predictive regression forecast based on the economic variable given in the first column relative to the historical average benchmark forecast. Brackets report p-values for the Clark and West (2007) *MSFE-adjusted* statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the predictive regression MSFE against the alternative that the historical average MSFE is greater than the predictive regression MSFE (corresponding to H_0: $R^2_{OS} \leq 0$ against H_A: $R^2_{OS} > 0$). Average utility gain (Δ) is the portfolio management fee (in annualized percent return) that an investor with mean-variance preferences and risk aversion coefficient of five would be willing to pay to have access to the predictive regression forecast based on the economic variable given in the first column relative to the historical average benchmark forecast. 0.00 indicates less than 0.005. R^2_{OS} statistics and average utility gains are computed for the entire 1957:01–2010:12 forecast evaluation period and separately for NBER-dated expansions and recessions.

Table 2 Monthly U.S. equity premium out-of-sample forecasting results based on multiple economic variables, 1957:01–2010:12

	Overall		Expansion		Recession	
Method	R^2_{OS} (%)	Δ (annual %)	R^2_{OS} (%)	Δ (annual %)	R^2_{OS} (%)	Δ (annual %)
Panel A: Unrestricted forecasts						
Kitchen sink	−8.43 [0.42]	0.24	−9.41 [0.68]	−1.60	−6.38 [0.28]	8.94
SIC	−5.61 [0.99]	−1.77	−5.80 [1.00]	−3.24	−5.21 [0.79]	5.00
POOL-AVG	0.44 [0.03]	1.25	0.12 [0.21]	0.41	1.10 [0.01]	5.14
POOL-DMSFE	0.51 [0.02]	1.52	0.08 [0.25]	0.40	1.39 [0.01]	6.76
Diffusion index	0.68 [0.01]	1.65	−1.00 [0.27]	−1.46	4.15 [0.00]	16.42
Sum-of-the-parts	0.93 [0.01]	2.47	0.29 [0.13]	0.27	2.24 [0.01]	12.87
Panel B: Forecasts with nonnegativity restrictions						
Kitchen sink	−2.57 [0.59]	0.24	−3.38 [0.64]	−1.60	−0.89 [0.45]	8.94
SIC	−4.47 [0.99]	−1.77	−4.73 [0.99]	−3.24	−3.03 [0.85]	5.00
POOL-AVG	0.44 [0.03]	1.25	0.12 [0.21]	0.41	1.10 [0.01]	5.14
POOL-DMSFE	0.51 [0.02]	1.52	0.08 [0.25]	0.40	1.39 [0.01]	6.76
Diffusion index	0.70 [0.01]	1.65	−0.68 [0.24]	−1.46	3.47 [0.00]	16.42
Sum-of-the-parts	0.99 [0.00]	2.47	0.31 [0.12]	0.27	2.40 [0.00]	12.87

Notes: R^2_{OS} measures the percent reduction in mean squared forecast error (MSFE) for the forecasting method given in the first column relative to the historical average benchmark forecast. Brackets report p-values for the Clark and West (2007) *MSFE-adjusted* statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the forecasting method MSFE against the alternative that the historical average MSFE is greater than the forecasting method MSFE (corresponding to H_0: $R^2_{OS} \leq 0$ against H_A: $R^2_{OS} > 0$). Average utility gain (Δ) is the portfolio management fee (in annualized percent return) that an investor with mean-variance preferences and risk aversion coefficient of five would be willing to pay to have access to the forecasting method relative to the historical average benchmark forecast. 0.00 indicates less than 0.005. R^2_{OS} statistics and average utility gains are computed for the entire 1957:01–2010:12 forecast evaluation period and separately for NBER-dated expansions and recessions.

Table 3 Quarterly U.S. equity premium out-of-sample forecasting results based on individual economic variables, 1957:1–2010:4

	Overall		Expansion		Recession	
Economic variable	R^2_{OS} (%)	Δ (annual %)	R^2_{OS} (%)	Δ (annual %)	R^2_{OS} (%)	Δ (annual %)
Panel A: Unrestricted predictive regression forecasts						
$\log(DP)$	−0.80 [0.07]	2.40	−6.92 [0.43]	−1.88	7.23 [0.00]	19.19
$\log(DY)$	0.36 [0.07]	3.00	−4.71 [0.47]	−1.19	7.02 [0.00]	19.39
$\log(EP)$	−6.51 [0.26]	1.61	−10.19 [0.36]	−0.05	−1.68 [0.29]	7.43
$\log(DE)$	−4.93 [0.97]	−0.02	−2.98 [0.75]	−0.17	−7.48 [0.97]	0.57
$SVAR$	−0.47 [0.92]	0.27	−0.40 [0.93]	0.18	−0.57 [0.80]	0.50
BM	−6.87 [0.20]	0.19	−13.53 [0.45]	−1.59	1.87 [0.13]	6.43
$NTIS$	−5.38 [0.48]	−1.12	1.01 [0.02]	0.44	−13.76 [0.95]	−7.32
TBL	−0.56 [0.16]	2.38	−1.96 [0.31]	0.36	1.28 [0.19]	10.22
LTY	−3.43 [0.20]	2.25	−5.62 [0.36]	0.17	−0.55 [0.22]	10.39
LTR	−0.72 [0.38]	0.14	0.70 [0.14]	0.20	−2.58 [0.73]	−0.33
TMS	−0.12 [0.25]	1.44	−1.71 [0.49]	−0.24	1.95 [0.13]	7.77
DFY	0.15 [0.31]	1.33	0.00 [0.40]	0.51	0.34 [0.33]	4.04
DFR	−4.79 [0.95]	−1.71	−1.93 [0.51]	−0.24	−8.54 [0.98]	−7.33
$INFL$	0.16 [0.33]	1.89	−0.25 [0.72]	−0.35	0.70 [0.23]	10.64
Panel B: Predictive regression forecasts with Campbell and Thompson (2008) restrictions						
$\log(DP)$	0.66 [0.04]	2.40	−3.90 [0.42]	−1.88	6.66 [0.00]	19.19
$\log(DY)$	1.48 [0.03]	3.00	−2.68 [0.38]	−1.19	6.94 [0.00]	19.39
$\log(EP)$	−2.22 [0.21]	1.61	−4.68 [0.36]	−0.05	1.00 [0.22]	7.43
$\log(DE)$	−4.22 [0.97]	0.04	−2.74 [0.71]	−0.10	−6.17 [0.98]	0.66
$SVAR$	−0.07 [0.87]	0.27	−0.01 [0.52]	0.18	−0.16 [0.91]	0.50
BM	−3.75 [0.22]	0.19	−6.68 [0.42]	−1.59	0.09 [0.21]	6.43
$NTIS$	−5.34 [0.48]	−1.12	1.03 [0.02]	0.44	−13.70 [0.95]	−7.32
TBL	0.50 [0.14]	2.38	−0.74 [0.30]	0.36	2.12 [0.16]	10.22
LTY	0.09 [0.11]	2.25	−1.43 [0.25]	0.17	2.08 [0.15]	10.39
LTR	−0.31 [0.30]	0.14	0.86 [0.13]	0.20	−1.85 [0.66]	−0.33
TMS	0.10 [0.23]	1.44	−1.63 [0.49]	−0.24	2.37 [0.09]	7.77
DFY	0.15 [0.31]	1.33	0.00 [0.40]	0.51	0.34 [0.33]	4.04
DFR	−0.13 [0.60]	−0.01	0.27 [0.25]	0.00	−0.64 [0.91]	−0.03
$INFL$	0.17 [0.32]	1.89	−0.23 [0.70]	−0.35	0.70 [0.23]	10.64

Notes: R^2_{OS} measures the percent reduction in mean squared forecast error (MSFE) for the predictive regression forecast based on the economic variable given in the first column relative to the historical average benchmark forecast. Brackets report p-values for the Clark and West (2007) *MSFE-adjusted* statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the predictive regression MSFE against the alternative that the historical average MSFE is greater than the predictive regression MSFE (corresponding to H_0: $R^2_{OS} \leq 0$ against H_A: $R^2_{OS} > 0$). Average utility gain (Δ) is the portfolio management fee (in annualized percent return) that an investor with mean-variance preferences and risk aversion coefficient of five would be willing to pay to have access to the predictive regression forecast based on the economic variable given in the first column relative to the historical average benchmark forecast. 0.00 indicates less than 0.005. R^2_{OS} statistics and average utility gains are computed for the entire 1957:1–2010:4 forecast evaluation period and separately for NBER-dated expansions and recessions.

Table 4 Quarterly U.S. equity premium out-of-sample forecasting results based on multiple economic variables, 1957:1–2010:4

	Overall		Expansion		Recession	
Method	R^2_{OS} (%)	Δ (annual %)	R^2_{OS} (%)	Δ (annual %)	R^2_{OS} (%)	Δ (annual %)
Panel A: Unrestricted forecasts						
Kitchen sink	−29.82 [0.76]	0.01	−25.89 [0.43]	−2.07	−34.97 [0.85]	8.22
SIC	−19.16 [0.95]	−3.64	−15.27 [0.59]	−4.40	−24.28 [0.97]	−1.31
POOL-AVG	0.73 [0.09]	1.74	0.40 [0.20]	0.98	1.15 [0.13]	4.42
POOL-DMSFE	1.09 [0.05]	2.27	0.23 [0.26]	0.68	2.22 [0.05]	8.41
Diffusion index	2.10 [0.01]	3.40	−3.58 [0.23]	−1.26	9.56 [0.00]	21.81
Sum-of-the-parts	2.02 [0.03]	3.61	0.24 [0.22]	0.63	4.35 [0.03]	15.34
Panel B: Forecasts with nonnegativity restrictions						
Kitchen sink	−10.35 [0.77]	0.01	−8.06 [0.45]	−2.07	−13.34 [0.87]	8.22
SIC	−15.37 [0.91]	−3.64	−12.36 [0.55]	−4.40	−19.31 [0.95]	−1.31
POOL-AVG	0.73 [0.09]	1.74	0.40 [0.20]	0.98	1.15 [0.13]	4.42
POOL-DMSFE	1.09 [0.05]	2.27	0.23 [0.26]	0.68	2.22 [0.05]	8.41
Diffusion index	2.06 [0.01]	3.40	−2.09 [0.23]	−1.26	7.51 [0.00]	21.81
Sum-of-the-parts	2.26 [0.02]	3.61	0.51 [0.18]	0.63	4.55 [0.02]	15.34

Notes: R^2_{OS} measures the percent reduction in mean squared forecast error (MSFE) for the forecasting method given in the first column relative to the historical average benchmark forecast. Brackets report p-values for the Clark and West (2007) *MSFE-adjusted* statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the forecasting method MSFE against the alternative that the historical average MSFE is greater than the forecasting method MSFE (corresponding to H_0: $R^2_{OS} \leq 0$ against H_A: $R^2_{OS} > 0$). Average utility gain (Δ) is the portfolio management fee (in annualized percent return) that an investor with mean-variance preferences and risk aversion coefficient of five would be willing to pay to have access to the forecasting method relative to the historical average benchmark forecast. 0.00 indicates less than 0.005. R^2_{OS} statistics and average utility gains are computed for the entire 1957:1–2010:4 forecast evaluation period and separately for NBER-dated expansions and recessions.

Table 5 Annual U.S. equity premium out-of-sample forecasting results, 1957–2010

Economic variable or method	Unrestricted		Campbell and Thompson (2008) restrictions		Nonnegativity restrictions	
	R^2_{OS} (%)	Δ (annual %)	R^2_{OS} (%)	Δ (annual %)	R^2_{OS} (%)	Δ (annual %)
Panel A: Forecasts based on individual economic variables						
$\log(DP)$	1.76 [0.09]	4.13	3.51 [0.07]	4.13	–	–
$\log(DY)$	−17.34 [0.21]	4.28	−1.26 [0.13]	4.28	–	–
$\log(EP)$	−4.18 [0.16]	1.30	0.76 [0.09]	1.30	–	–
$\log(DE)$	−7.89 [0.95]	−0.66	−0.05 [0.84]	0.02	–	–
$SVAR$	−2.34 [0.94]	−0.02	0.00 [–]	0.02	–	–
BM	−8.16 [0.17]	1.63	−4.99 [0.14]	1.63	–	–
$NTIS$	−16.89 [0.71]	−1.32	−16.89 [0.71]	−1.32	–	–
TBL	−4.86 [0.20]	2.88	0.95 [0.14]	2.88	–	–
LTY	−9.13 [0.18]	2.46	0.74 [0.11]	2.46	–	–
LTR	−6.35 [0.04]	1.92	−5.03 [0.04]	1.92	–	–
TMS	−0.34 [0.22]	1.48	0.00 [0.22]	1.48	–	–
DFY	−2.07 [0.97]	0.00	−0.01 [0.91]	0.02	–	–
DFR	−4.69 [0.58]	−0.04	−4.34 [0.57]	−0.04	–	–
$INFL$	−1.06 [0.70]	−0.22	−1.06 [0.70]	−0.22	–	–
Panel B: Forecasts based on multiple economic variables						
Kitchen sink	−99.18 [0.70]	−2.93	–	–	−57.39 [0.66]	−2.93
SIC	−45.37 [0.56]	−1.02	–	–	−32.74 [0.53]	−1.02
POOL-AVG	3.11 [0.07]	1.36	–	–	3.11 [0.07]	1.36
POOL-DMSFE	1.71 [0.17]	1.29	–	–	1.71 [0.17]	1.29
Diffusion index	7.14 [0.03]	4.89	–	–	6.98 [0.03]	4.89
Sum-of-the-parts	5.60 [0.04]	2.85	–	–	5.85 [0.04]	2.85

Notes: R^2_{OS} measures the percent reduction in mean squared forecast error (MSFE) for the predictive regression forecast based on the economic variable or forecasting method given in the first column relative to the historical average benchmark forecast. Brackets report p-values for the Clark and West (2007) *MSFE-adjusted* statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the predictive regression or forecasting method MSFE against the alternative that the historical average MSFE is greater than the predictive regression or forecasting method MSFE (corresponding to H_0: $R^2_{OS} \leq 0$ against H_A: $R^2_{OS} > 0$). Average utility gain (Δ) is the portfolio management fee (in annualized percent return) that an investor with mean-variance preferences and risk aversion coefficient of five would be willing to pay to have access to the predictive regression forecast or forecasting method relative to the historical average benchmark forecast. 0.00 indicates less than 0.005.

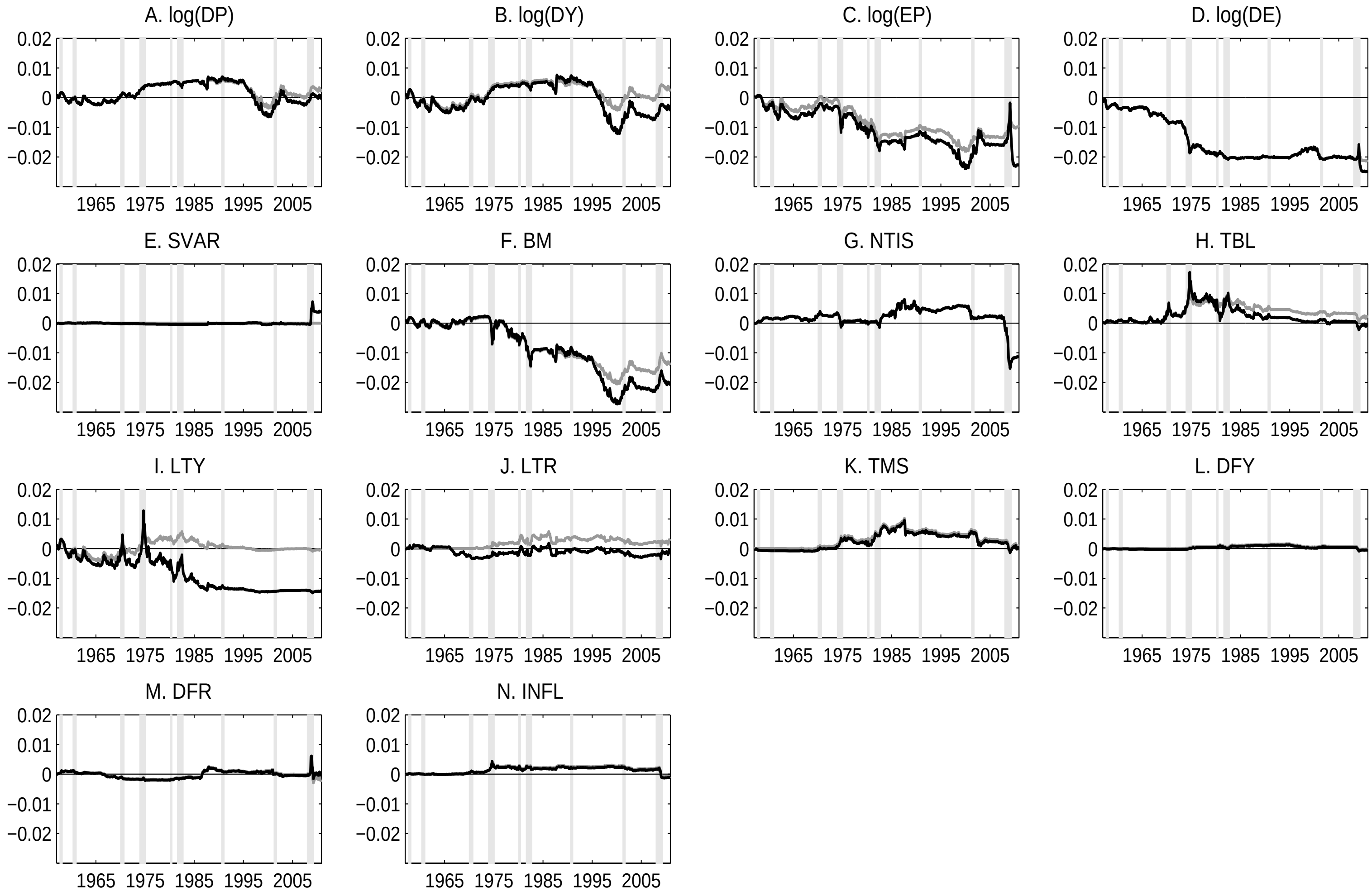

Figure 1 Cumulative differences in squared forecast errors, monthly U.S. equity premium out-of-sample forecasts based on individual economic variables, 1957:01–2010:12. Black (gray) lines in each panel delineate the cumulative difference in squared forecast errors for the historical average forecast relative to the unrestricted predictive regression forecast (predictive regression forecast with Campbell and Thompson (2008) restrictions imposed) based on the economic variable given in the panel heading. Vertical bars depict NBER-dated recessions.

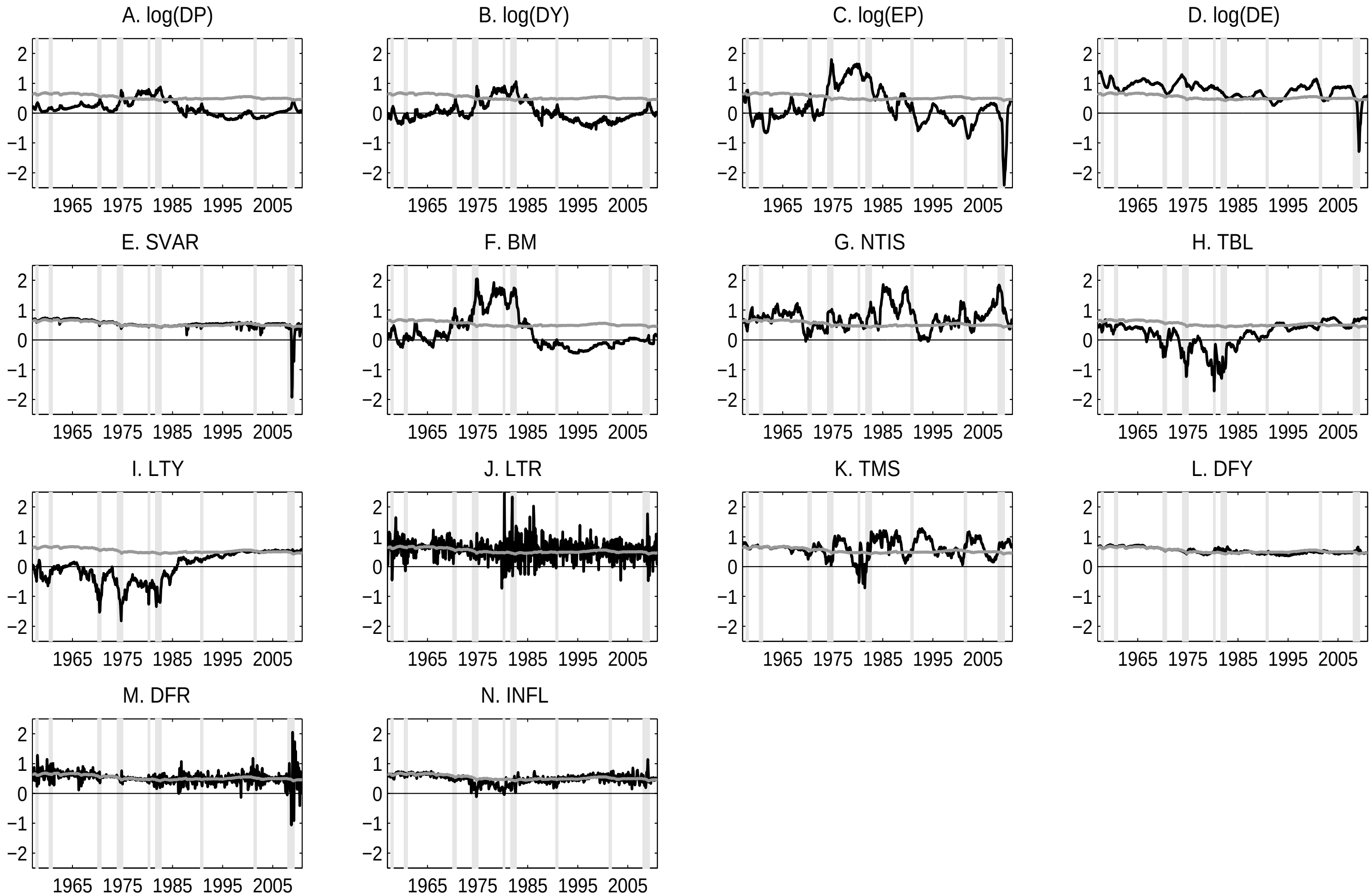

Figure 2 Monthly U.S. equity premium out-of-sample forecasts (in percent) based on individual economic variables, 1957:01–2010:12. Black (gray) lines delineate unrestricted predictive regression forecasts based on the economic variable given in the panel heading (historical average forecast). Vertical bars depict NBER-dated recessions.

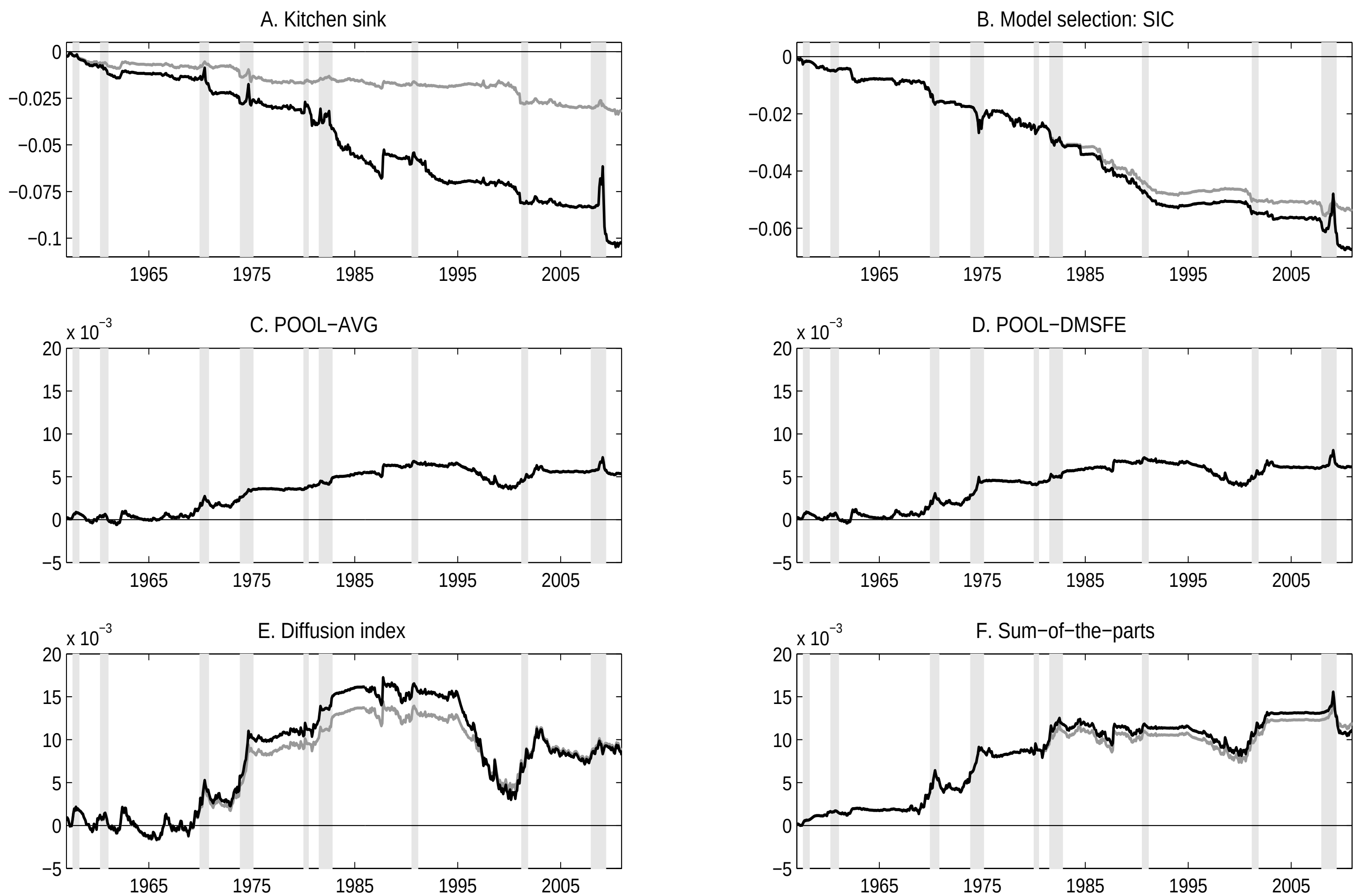

Figure 3 Cumulative differences in squared forecast errors, monthly U.S. equity premium out-of-sample forecasts based on multiple economic variables, 1957:01–2010:12. Black (gray) lines in each panel delineate the cumulative difference in squared forecast errors for the historical average forecast relative to the forecasting method given in the panel heading (forecasting method with nonnegativity restrictions imposed). Vertical bars depict NBER-dated recessions.

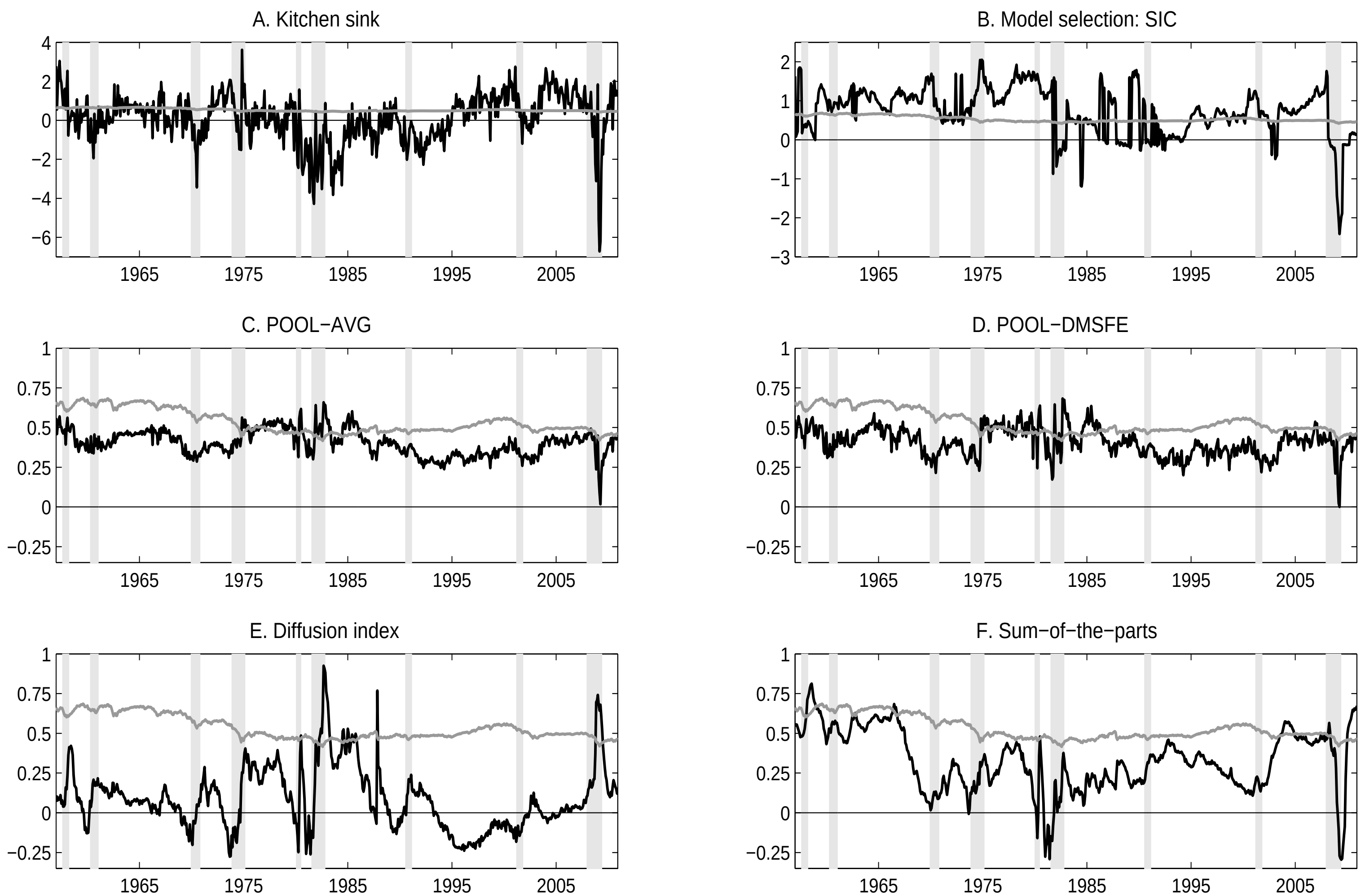

Figure 4 Monthly U.S. equity premium out-of-sample forecasts (in percent) based on multiple economic variables, 1957:01–2010:12. Black (gray) lines delineate forecasts based on the forecasting method given in the panel heading (historical average forecast). Vertical bars depict NBER-dated recessions.