Journal of Financial Economics 99 (2011) 560–580

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Time-varying short-horizon predictability ☆

Sam James Henkel [a], J. Spencer Martin [a], Federico Nardari [b,*]

[a] University of Melbourne, Australia
[b] C.T. Bauer College of Business, University of Houston, United States

ARTICLE INFO

Article history:
Received 11 June 2007
Received in revised form
14 September 2009
Accepted 21 September 2009
Available online 29 September 2010

JEL classification:
E44
G11
G12

Keywords:
Stock return predictability
Asset pricing
Business fluctuations
Financial markets and the macroeconomy

ABSTRACT

In the G7 countries, the short-horizon performance of aggregate return predictors such as the dividend yield and the short rate appears non-existent during business cycle expansions but sizable during contractions. This phenomenon appears related to countercyclical risk premiums as well as the time-variation in the dynamics of predictors. Our empirical model outperforms the historical average out-of-sample in the US, but the results throughout the G7 are mixed.

1. Introduction

A number of studies document a disappearance of stock return predictability from US markets. Some researchers point to parameter instability or structural breaks and identify the date of disappearance circa 1991 (Pesaran and Timmermann, 2002; Lettau and Van Nieuwerburgh, 2008). A related hypothesis is that predictability was arbitraged away once discovered, in a scenario similar to attenuation of the January effect. Welch and Goyal (2008) argue that predictability has not been significant in-sample or out-of-sample in the past 30 years. Still others take a Galbraithian view, contending it was never actually there (Bossaerts and Hillion, 1999; Goyal and Welch, 2003).[1]

In this study we reveal predictability as a phenomenon whose strength is distinctively time-varying. The dividend

☆ The original version of this paper was circulated with the title, "Stock Return Predictability: When is it there?" and was inspired by the similarly titled paper by Ang and Bekaert (2008). We thank an anonymous referee, Giuseppe Alesii, George Cashman, Long Chen, Peter Christoffersen, Mike Gallmeyer, Rick Green, John Griffin, Susan Ji, Jerry Kallberg, Patrick Kelly, Praveen Kumar, Zoran Ivkovic, Chris Lamoureux, Crocker Liu, Christian Lundblad, Bradley Paye, Jesper Rangvid (EFA Discussant), Phlippe Rhoner (EFMA Discussant), Ken Roskelley, Sergei Sarkissian, G. William Schwert (the editor), Duane Seppi, Rossen Valkanov (AFA Discussant), Sriram Villupuram, Ziwei Xu (FMA Discussant) and seminar participants at Arizona State University, Barclays Global Investors, Carnegie Mellon University, Indiana University, McGill University, Michigan State University, Queen's University, Thunderbird School of Management, University of Houston, University of Arizona, the 2008 American Finance Association meetings, the 2008 European Finance Association meetings, the 2008 Financial Management Association meetings, the 2009 European Financial Management Association meetings, the 2006 Northern Finance Association meetings and the Duke Financial Econometrics Lunch Group for their useful comments. The usual disclaimer applies.

* Corresponding author.
E-mail addresses: samuel.henkel@yale.edu (S.J. Henkel), martis@unimelb.edu.au (J.S. Martin), fnardari@uh.edu (F. Nardari).

0304-405X/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jfineco.2010.09.008

[1] John Kenneth Galbraith once said, "There are two types of forecasters: those who don't know and those who don't know they don't know."

yield and commonly used term structure variables are effective predictors almost exclusively during recessions. Fig. 1 provides a direct, simple view of this phenomenon. Plotted are time series of the cumulative proportion of recession months in the US data and the adjusted R^2 ($\overline{R}^2$) from a one-month-ahead predictive ordinary least-squares (OLS) regression using the contemporaneously available sample. The $\overline{R}^2$ rises and falls with the proportion of recession months in the available sample. In the US over the 1953–2007 period, the average $\overline{R}^2$ is about 15% during recessions yet less than 1% in expansions. We investigate this basic pattern in each of the G7 countries with more rigorous econometric methods. No country has $\overline{R}^2$ significantly different from zero during expansions, and no individual predictor is more important in expansions than in recessions.

The robust prominence of business cycles in these results suggests a potentially substantial tie to the literature on the dynamics of expected returns. Campbell and Cochrane (1999), Menzly, Santos, and Veronesi (2004), and Bekaert, Engstrom, and Xing (2009) show that risk premiums are countercyclical, and that the time-series behavior of risk premiums governs at least some return predictability. Consistent with this literature, we estimate that the market risk premium is higher during recessions in all seven of the countries studied.

Since a time-varying predictive relation is the byproduct of the interacting dynamics of expected returns and of the predictors, the complex behavior of the predictors themselves must be considered. The underlying primitives are the potential micro-level objectives of firms and central banks whose activities jointly determine aggregate predictor variables. The business cycle is an important driver of these micromotives, and this leads us to re-examine predictability using a regime-switching vector autoregression (RSVAR) framework capable of matching the time-varying dynamics of predictors to the dynamics of expected returns. In support of this view, we find the predictors themselves to be less persistent and more volatile during recessions. The increases in predictor volatility can approach the magnitude of the well-known increase in realized market volatility during recessions.

The countercyclical behavior of short-horizon predictability also provides a historical context in which to understand important elements of prior research. Combining our results with the benefit of hindsight illustrates a link between these predictability findings and their contemporaneous economic history. Fig. 2 shows the cumulative ratio of recession months to all months in the Center for Research in Stock Prices (CRSP) data since its inception in 1963. Overlaid are indicative, not comprehensive, citations of early research on predictability for each variable we consider. Several features stand out: the random walk model of stock prices prevailed in the 1970s, based upon CRSP data from the long 1960s era expansion; predictability emerged in research of the late 1970s and mid-1980s, following several recessions; and predictability was subsequently doubted following the long booms of the 1980s and 1990s. Although great technical strides have been achieved, this figure reinforces the notion that the conclusions drawn are ultimately also dependent on the available data sample.

The remainder of the paper proceeds as follows. We lay the foundations of our work in Section 2. Section 3

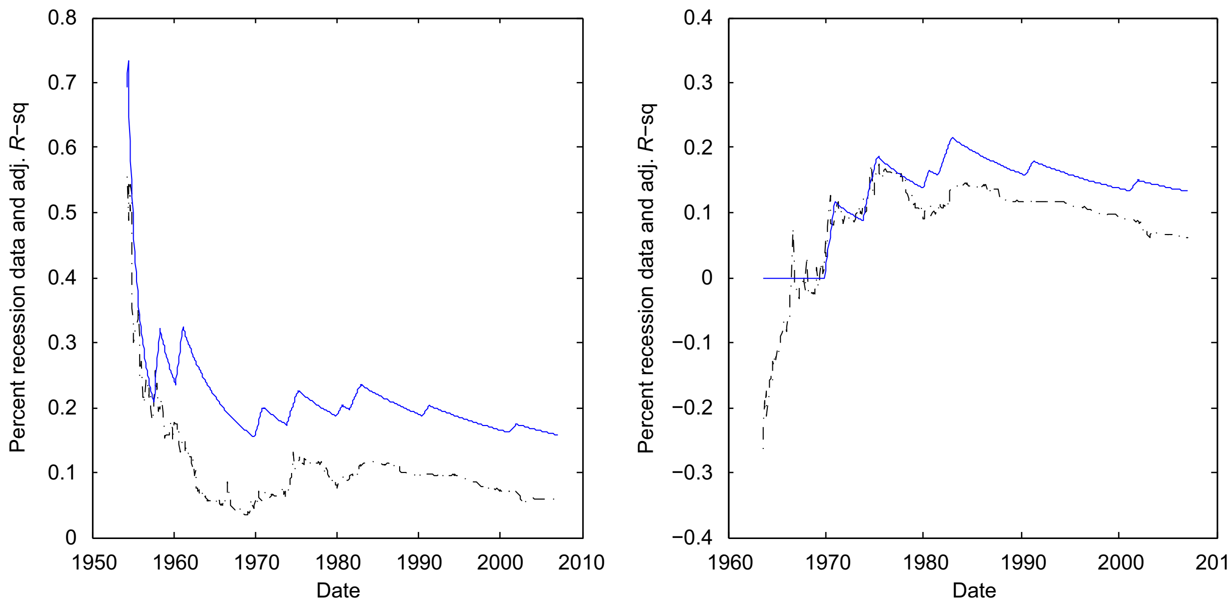

Fig. 1. Recessions and predictability. (a) and (b) plot the adjusted R^2 ($\overline{R}^2$) from a predictive OLS cumulating regression and the cumulative percent NBER recession data starting from 1953:04 and from 1962:07, respectively. The adjusted $\overline{R}^2$ is computed from a predictive OLS regression of the excess market return at time t on the dividend yield, short rate, term spread, and default spread at time $t-1$. For (a), the regressions are run for each t=1954:04 to 2006:12. For (b), the sample is restricted to the originally historically available CRSP sample starting with t=1963:07 to 2006:12. The correlation between proportion of NBER recessions and the corresponding adjusted $\overline{R}^2$ is 70% to 90% depending on time periods used. The charts are for expository purposes only.

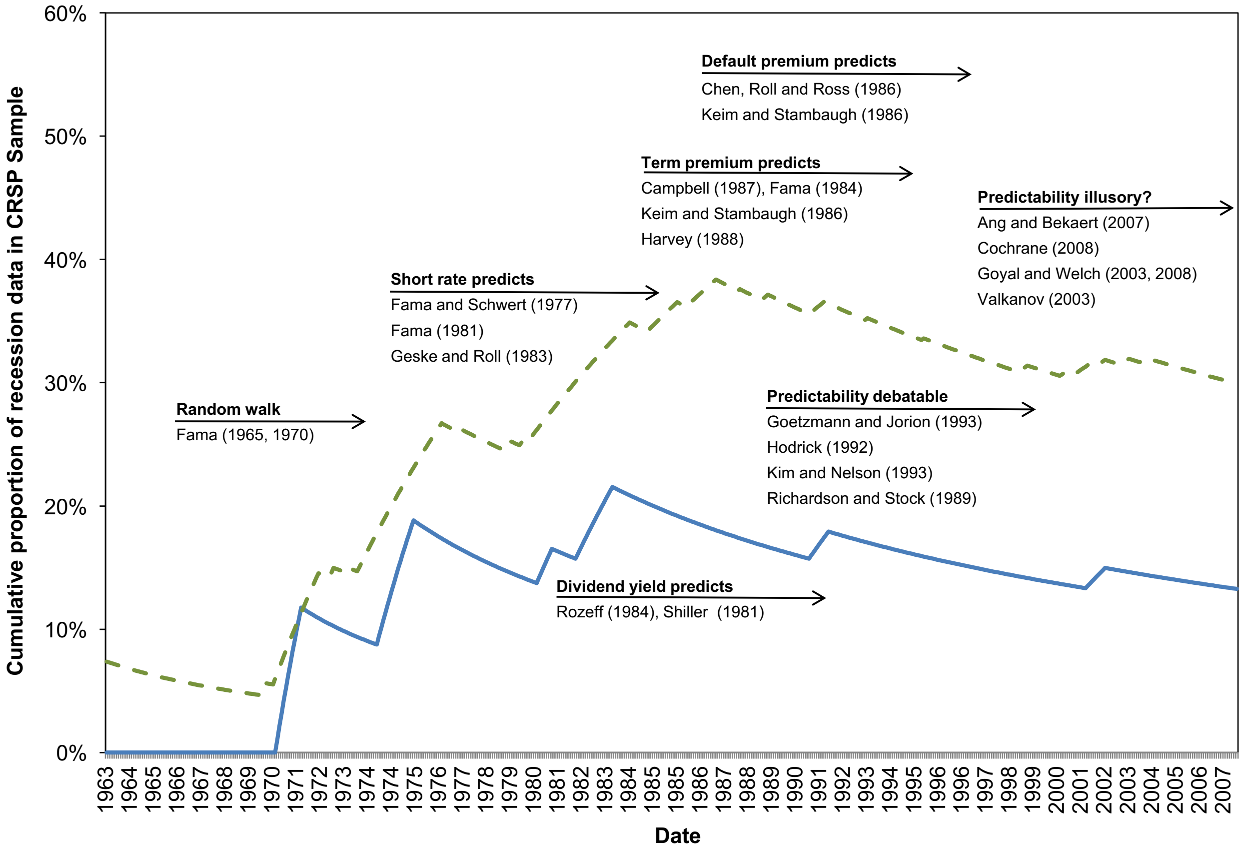

Fig. 2. The time-series of predictability research. The literature on stock return predictability follows closely the availability of recession data as a cumulative proportion of the total data in CRSP which originally started in 1962. Shown are the percentages of recession data as a percentage of the available data at a given date, as measured by NBER (solid line) and RSVAR (dashed line) dates. Both the NBER and RSVAR samples show similar profiles, although RSVAR recession probabilities represent a much larger proportion of the data. Many seminal, and first, papers on return predictability were published just after the peaking of the proportion of recession data to total available data in 1985 and are followed by a decline in the proportion of recession data thereafter. The citations are representative for expository purposes and are not intended to be indicative of initial research, nor a comprehensive literature survey (Ang and Bekaert, 2007; Campbell, 1987; Chen et al., 1986; Cochrane, 2008; Fama, 1965, 1970, 1981, 1984; Fama and Schwert, 1977; Geske and Roll, 1983; Goetzmann and Jorion, 1993; Goyal and Welch, 2003; Harvey, 1988; Hodrick, 1992; Keim and Stambaugh, 1986; Kim and Nelson, 1993; Richardson and Stock, 1989; Rozeff, 1984; Shiller, 1981; Valkanov, 2003; Welch and Goyal, 2008).

expands on our empirical approach and is followed by a description of the data in Section 4. Section 5 reports our empirical findings. Section 6 concludes.

2. Background and motivation

2.1. Dynamics of expected returns

Early empirical evidence of countercyclical risk premiums is in Fama and French (1989) and Ferson and Harvey (1991). The basic intuition for a link between countercyclical risk premiums and return predictability is simple and appealing. If investors demand higher risk premiums in bad times, and volatility is higher in bad times as well, then overall adjustments to discount rates per unit of change in economic state are larger in bad times. Crucially, price–dividend ratios become more volatile and prices more sensitive to changing expectations as conditions worsen. Predictability might, therefore, be a countercyclical phenomenon.

The cyclical dynamics of risk premiums and of return predictability need not be synchronous, however. Using the framework of Campbell and Cochrane (1999), Li (2007) shows, counterintuitively, that changes in risk aversion alone are insufficient to induce any return predictability at all. In another example, Mele (2007) demonstrates that countercyclical risk premiums do not necessarily imply higher return volatility in bad times.

Nevertheless, we account for the possibility of countercyclical return predictability in two ways. First, we decompose the sources of predictability to control for shifts in market volatility relative to predictor volatility. Second, we design tests based upon professional survey data to better distinguish the effects of current conditions from the effects of expectations regarding future economic conditions.

Changes in predictability over time could also result from infrequent, random structural breaks rather than business cycles. Under different assumptions, Pesaran and Timmermann (2002) and Lettau and Van Nieuwerburgh (2008) both identify 1991 as one such structural break. Since there have been further National Bureau of

Economic Research (NBER) recession periods since 1991, the consistency of our countercyclical empirical findings through time would make the structural break hypothesis a less appealing overall explanation.

2.2. Dynamics of predictors

Understanding which predictors work and why is an old question going back at least as far as Dow (1920). We focus on a simple menu of macroeconomic predictors identified since the 1970s: the dividend yield (Rozeff, 1984), the short rate (Fama and Schwert, 1977; Fama, 1981), the slope of the term structure (Keim and Stambaugh, 1986; Campbell, 1987; Fama and French, 1989), and the default premium (Fama and Bliss, 1987; Campbell, 1987; Fama and French, 1989).

A growing body of empirical evidence documents instabilities and nonlinearities in the time-series properties of these predictors, likely induced by the interplay amongst market participants, managers, and central banks. Noticeable and persistent patterns in the microbehavior of central banks and managers over monetary policy regimes and over the business cycle identify one dimension of this interaction, the smoothing of useful information, as shown by Mankiw and Miron (1986), Gray (1996), Davig and Leeper (2007), Lintner (1956), Kumar (1988), and Fudenberg and Tirole (1995). Smoothing is often associated with greater persistence and reduced volatility as individual shocks are spread out over several periods.

Chan, Karolyi, Longstaff, and Sanders (1992), Gray (1996), and Ang and Bekaert (2002a), among others, focus on the short-term interest rate, *SR*, and find it best described by multiple structural breaks or regime shifts. Further, these shifts may be associated with different implications for the rest of the economy. Ang and Bekaert (2002b, p. 166) observe, "In expansions, the interest rate persistence may arise from the smoothing efforts of the monetary authorities." In related work, Mankiw and Miron (1986) examine the relation between Fed targeting and the (univariate) predictability of interest rates. As Hardouvelis (1988, p. 340) summarizes, "In their view, the Fed's targeting of interest rates makes interest rate changes unpredictable."

Although we are aware of no study focused exclusively on instabilities in the dynamics of the term spread (*TERM*), the evidence previously cited regarding the short rate coupled with the findings by Ang and Bekaert (2002b), Ang, Bekaert, and Wei (2008), Bansal and Zhou (2002), Bansal, Tauchen, and Zhou (2004), and Boudoukh, Richardson, Smith, and Whitelaw (1999) suggest the inclusion of term spread information improves the ability to forecast the short rate itself and partially resolves puzzles relating to the expectations hypothesis, indicating the term spread may plausibly be characterized by a similar switching-type structure as the short rate.

These considerations apply to corporate bond yields and, thus, possibly to the default spread (*DEF*). Alternatively, asymmetric accounting disclosures, known as "conservatism," across the business cycle may explain the higher persistence of default spreads in economic expansions. Conservatism in accounting is summarized by the saying "anticipate no profit, but anticipate all losses" (Bliss, 1924; Watts, 2003).[2] Therefore, balance sheets may be quicker to report bad news than to realize good news leading to slower business condition disclosures and information propagation in good times.

Accounting conservatism maps more directly into the reported earnings and resultant dividends of the income statement than to balance sheet measures of credit worthiness (Basu, 1997). Even more so than interest rates, dividend payouts are known to be the subject of explicit policy decisions at the corporate level that persist even today (Lintner, 1956; Brav, Graham, Harvey, and Michaely, 2005). Marsh and Merton (1987) model aggregate dividend dynamics based on managerial smoothing, but do not consider return predictability. Consistent with our conjecture, Chen, Da, and Priestley (2008) look at cross-sectional differences in dividend-smoothing and find that more smoothing is related to less return predictability.

3. Research design

3.1. Choice of horizon

All of our analysis, with the exception of Section 5.4, uses one-month intervals. Substantial evidence shows return predictability is consistent with a short-horizon phenomenon that is magnified at longer horizons (for examples, see Campbell, Lo, and MacKinlay, 1997, p. 271; Cochrane, 2001, p. 393). In a sample of four developed countries, Ang and Bekaert (2007) show, by an exact Gordon growth formulation and proper annualization of returns, that return predictability is concentrated at short horizons. Boudoukh, Richardson, and Whitelaw (2008) provide simulation evidence that long-horizon predictability may result from highly correlated sampling errors. On the other hand, Campbell (2001) argues in favor of improved power in long-horizon tests.

Besides allowing easy comparisons with the literature, our choice of frequency conveniently abstracts away from the econometric issues associated with long-horizon regressions and overlapping observations (Hodrick, 1992). However, more important than the circumvention of these econometric challenges, predictability regressions at annual horizons or longer would be unlikely to detect the results we report because recessions are typically short-lived events. For example, the expected duration of recessions in our sample ranges from two to five months. The expected duration of expansions ranges from seven to 11 months. Thus, as the state of the economy evolves through time, long-horizon regressions would include random combinations of expansionary and recessionary periods, with any differences between predictability estimates in expansions and recessions hopelessly blurred.

3.2. Choice of predictive variables

Following the discussion in Section 2, our set of predictive variables includes *DY*, *SR*, *TERM*, and *DEF*. These

[2] Conservatism is a fundamental principle in the accounting field (Sterling, 1970, p. 256). Basu (1997, p. 8) shows conservatism has been in practice for at least 500 years.

are precisely measured, high-frequency, market-traded *ex ante* quantities, as opposed to quarterly, lagged, or often-revised government statistics. Moreover, except for the default spread which is available only for the US, these quantities are readily comparable across countries. Following the approach of Chordia and Shivakumar (2002) and Avramov and Chordia (2006), among others, we do not impose additional filters on the raw data, such as separating the short rate into real and inflationary components, or take a stand on expected versus unexpected inflation.[3]

Unfortunately, *DEF* is only available for the US. Following Paye and Timmermann (2006), we include the US default spread in the specification for Canada. Because the Canadian business cycle tracks the US cycle closely and given the strong links between US and Canadian markets, the US default spread seems useful in this specification. However, in specifications for European or Asian countries, the inclusion of *DEF* appears to add noise to the regime identification scheme, possibly due to differences between the US and other countries' business cycles.[4]

In addition to the above ex ante state variables, we need indicators of expansions and recessions. The most obvious candidate for the US is the NBER business cycle measure. However, the NBER determines business cycles by looking at many time series and then selecting dates ex post by committee.[5] The unobservable and ex post nature of the determination process makes the measure objectively unknowable and contemporaneously unavailable to either the agent or the econometrician. Still, we cross-check our analyses using the NBER dates as they (1) provide convenient economic intuition, (2) allow our results a connection to the larger macroeconomic literature, and (3) provide a hedge against the risk of data mining. After all, the degree of return predictability is a highly unlikely metric for the NBER to use in classifying business cycles.

We also consider the Survey of Professional Forecasters and the Livingston Survey as additional checks on regime classification. Ang, Bekaert, and Wei (2007) find these surveys effective in forecasting economic conditions such as inflation. Aside from robustness, the main benefit of these surveys is that, unlike NBER dates, they are forward looking (at different horizons) and available to investors in real time.

For the international data, NBER-quality indicators are either unavailable or not in standard widespread use. We do consider, for rough comparison, the ex post indicator dates available from the Economic Cycle Research Institute (ECRI) Web site. Because we are not aware of any real-time indicators, we draw from the business cycle research of Hamilton (1989, 1994) and let the data determine the state based on a regime-switching vector autoregression (RSVAR).

Unlike the RSVARs typically applied in the macroeconomic literature, we do not use government statistics such as gross domestic product (GDP), consumption, investment, or labor income for the same reasons already given. The RSVAR is capable of determining time t latent state probabilities in three ways. Ex ante probabilities use information up to $t-1$, filtered probabilities include information up through t, and finally, smoothed probabilities use all of the information up to the final observation, T. Even so, the RSVAR still does use some advance information, so in Section 5.5 we consider the out-of-sample performance of the RSVAR.

3.3. Empirical framework

In the RSVAR, y_t (a vector of observable variables) is assumed to be drawn from one of M different distributions at each time t. As a consequence, the autoregressive coefficients and the residuals' covariance matrix are allowed to jump among M sets of values. The RSVAR takes the form

$$y_t = c(s_t) + A(s_t)y_{t-1} + \varepsilon(s_t), \tag{1}$$

where $y_t = (r_t, DY_t, SR_t, TERM_t, DEF_t)'$, r_t is the value-weighted market index excess return at time t, and the other elements of y_t are as defined above. Unlike prior studies, we include the dividend yield as part of the switching model to better pick up potential business cycle variation in return predictability. $c(s_t)$ is a vector of constants, $A(s_t)$ is the coefficient matrix in state $s_t \in \{1, \ldots, M\}$, and $\varepsilon(s_t) \sim N(0, \Sigma(s_t))$ is an error vector of zero mean with a generalized (i.e., non-diagonal) state-dependent error covariance matrix.[6] The states s_t follow a Markov chain where the transition probabilities between one regime at time t and the contiguous regime at time $t+1$ are fixed and contained in an $M \times M$ transition matrix P.[7]

A well-known identification problem must be addressed in regime-switching specifications. Although the indices of the states can be permuted without affecting the likelihood function of the process, it is customary (see, for instance, Krolzig, 1998), to impose identification through some prior beliefs about regime

[3] Considerable evidence (Ang and Bekaert, 2007; Menzly, Santos, and Veronesi, 2004) suggests multiple regression predictability models outperform single regressor models. In prior drafts, we find these simpler models follow the same pattern as the more relevant multiple regressor models.

[4] As a result of the differences in business cycles, it is unclear whether the structural breaks observed in Paye and Timmermann (2006) are due to these conflicting business cycles or due to strictly local effects.

[5] Typical announcements have occurred between six and 18 months after the fact. Another unexplored issue is that the NBER dating methodology may evolve over time and across administrations. We collect business cycle peaks and troughs from the NBER Web page, http://www.nber.org/cycles.

[6] Regarding the number of parameters in the RSVAR, there are two states, one lag, and five variables, so the system has 92 unique parameters to be estimated. France, Germany, Italy, Japan, and the UK do not have the default spread, so their systems only have four variables and, subsequently, 62 parameters to be estimated. Although 92 (62) parameters may not seem parsimonious, we jointly fit five (four) time series with 420 to 657 monthly observations each, so the saturation ratios (the number of observations per parameter) range from 22.8 (Canada) to 33.2 (United Kingdom).

[7] These transition probabilities could be made more realistic by allowing them to vary dependent on the state variables. Given the results with fixed probabilities, however, it appears the refinement would not add enough economic insight to warrant the increased complexity and computational costs.

characteristics. Specifically, we set as priors that $M=2$ and that residual market volatility is higher in one of the two regimes. The two-state RSVAR applied here may be seen as an approximation of a more complex, as yet unknown, model of return predictability. Besides having appealing theoretical properties, Markov-switching processes like an RSVAR are convenient econometric simplifications of state-space systems having potentially a continuum of states. The shocks to the system ε_t are assumed to be normally distributed, but this is not an especially restrictive choice. Because many distributions can be approximated by a mixture of normal distributions, the regime-switching approach has had wide appeal in fitting non-normalities, nonlinearity, volatility clustering, and higher distributional moments of returns.[8]

Our RSVAR most closely resembles that of Guidolin and Timmermann (2007), but we use bond yields rather than bond returns. This subtle difference can lead to large differences in results since bond yields are persistent, but bond returns are not. This difference can be seen in the weak role that bond returns play in Chen, Roll, and Ross (1986) and Fama and French (1993) compared to the much stronger role bond yields play in Keim and Stambaugh (1986) and Campbell (1996). As such, Guidolin and Timmermann (2007) is primarily about asset allocation with state-varying return comovement between stocks and bonds.

Our RSVAR approach is also related to Perez-Quiros and Timmermann (2000), but bears substantive technical and qualitative distinctions. First, we model a complete dynamic system of market returns and predictors in a VAR, rather than a switching regression for market returns only. Portfolio returns are a prime driver of the regimes in Perez-Quiros and Timmermann (2000), since these portfolio returns are the only distinguishing variable across their size portfolios. Obviously, the economy-wide variables are common to all portfolios. Second, we allow the predictive relation with the dividend yield to switch across states. Third, we incorporate inferences about future expected returns through the term spread. Fourth, we rely exclusively on high-precision market-measured real-time quantities. Finally, we rely on a different estimation and inference methodology: while Guidolin and Timmermann (2007) and Perez-Quiros and Timmermann (2000) rely on a maximum likelihood framework implemented through the expectation-maximization (EM) algorithm, we adopt a Bayesian approach. We outline the details and the relative merits of this approach next.

3.3.1. Bayesian estimation: motivation

We employ Bayesian Markov Chain Monte Carlo (MCMC) methods to analyze the RSVAR. Our Bayesian MCMC analysis builds on techniques developed in Kim and Nelson (1999) and in Krolzig (1998). The Bayesian framework possesses several attractive features for our analysis.

First, models with latent variables (regimes and their respective probabilities, in this case) are especially amenable to estimation using MCMC methods. Using the Bayesian approach, the parameter space is augmented with the state probabilities and the Gibbs sampler algorithm effectively "integrates out" these nuisance parameters. The model parameters and latent variables are estimated simultaneously by drawing them from their joint distribution, so posterior densities implicitly incorporate estimation error (i.e., parameter uncertainty). In contrast, in a standard maximum likelihood approach based on, say, the commonly adopted EM algorithm, inference on the unobserved state vector is made conditional on the parameter estimates.

Second, the distribution of test statistics in a frequentist framework relies on asymptotic arguments and on approximations such as the delta method. We aim to investigate the properties of recessionary periods expected to be relatively brief, and that brevity fosters concerns about the small-sample validity of frequentist inference. In addition, we are interested in quantities, such as the RSVAR's R^2 or the half-life of shocks in the RSVAR, which are nonlinear functions of the model parameters; again, their small sample distribution would raise concerns. The Bayesian framework delivers representative finite-sample posterior densities for both parameters and features of interest.

Third, the Bayesian approach naturally allows for the inclusion of prior beliefs about these objects of interest. In our context, plausible prior beliefs about the magnitude of market risk premiums, regime durations and, possibly, the sign of predictive coefficients are easily incorporated.

3.3.2. Bayesian estimation: prior beliefs

Appendix A describes analytically the prior distributions we adopt. As they are quite uninformative and are commonly used in the analysis of predictive regressions, we do not discuss them here. The only exception is the prior distribution on the market risk premium, i.e., on the expected market return from the RSVAR. While eliciting informative prior beliefs directly on the RSVAR parameters may be problematic, imposing some structure on the magnitude of the market risk premium seems more plausible. Following Kandel and Stambaugh (1996) and Pástor and Stambaugh (2009), we use a normal distribution as our prior on the expected market return, with mean set equal to the historical risk premium and a standard deviation of 2% per quarter. Although informative, such prior does allow for a large range of possible risk premiums estimates. For robustness, we experiment with a number of alternative reasonable priors on the market premium.[9]

[8] Although the specification of shocks is rather simple, incorporation of richer multivariate generalized autoregressive conditional heteroskedasticity (GARCH) dynamics should only improve the identification of regimes. On the other hand, the analysis in Hamilton and Susmel (1994) shows that GARCH effects in aggregate stock returns disappear at the monthly horizon once the variance is allowed to follow a Markov-switching process.

[9] If return predictability, which we argue is largely recession specific, arises from predicting negative expected returns, then our finding may either be expected, given the well-known negative skewness of realized returns, or spurious if expected returns are, in fact, always positive. To address this, we impose priors that allow for 10%, 5%, or 0% probability of a negative expected return. In all cases, shrinking the prior (by reducing the standard deviation) towards

3.3.3. Bayesian inference: testing and model comparison

In addition to summarizing posterior distributions through their sample moments, we provide highest posterior density (HPD) intervals. A $100(1-\alpha)$ HPD interval for a parameter λ has the property of being the smallest interval delimiting an area of $1-\alpha$ under the posterior distribution of λ. Heuristically, a Bayesian HPD interval may be seen as similar to a frequentist confidence interval. In other words, when $\alpha=0.05$, the researcher is 95% confident that λ lies within the HPD. Thus, HPD intervals are well-suited to assess how much confidence the econometrician can place on, for instance, the difference in predictive coefficients across regimes.

To conduct more formal inference on model specification and parameter restrictions, we employ Bayes factors as they provide a unified way to compare the relative support that the data provide for competing model specifications (Kass and Raftery, 1995). Bayes factors also have the appealing property of implicitly penalizing models for additional parameters (O'Hagan, 1994). For the models we consider, we rely on Bayes factors to assess the evidence for regime-switching specifications versus linear models, models with switching means and variances versus models where only one of the two sets of moments is allowed to switch, and RSVARs with no restrictions on the companion matrix versus RSVARs with parameter constraints.

The Bayes factor comparing Model f to Model g is defined as

$$BF_{f,g}=\frac{m(y|\mathcal{M}_f)}{m(y|\mathcal{M}_g)}, \tag{2}$$

where $m(y|\mathcal{M}_h)$ is the marginal likelihood of the data, y, given model $\mathcal{M}_h$, h={f,g}. The marginal likelihood is obtained by integrating the likelihood function with respect to the prior density of the parameters. The computation of marginal likelihood is detailed in the last section of Appendix A.

Bayes factors are closely related to posterior-odds ratios. When the prior-odds ratio is 1:1 (i.e., equal prior probabilities), the Bayes factor coincides with the posterior-odds ratio. In our empirical work, we interpret Bayes factors under the assumption that all models are equally likely a priori. This is, by far, the most common approach (e.g., Cremers, 2002; Avramov, 2002; Avramov and Chao, 2006; Nardari and Scruggs, 2007). If one has different prior odds for a particular pair of model specifications, it is simple to compute corresponding posterior odds for a given Bayes factor.

3.3.4. Forecasting

The Bayesian methodology outlined above can be extended to include a forecasting step. In addition to generating draws from the posterior distribution of the model parameters, the Gibbs sampler can generate draws from the predictive density $p(y_{t+k}|Y_t)$ where Y_t denotes all observations up to time t, the specific details of which can be seen in Krolzig (1998, Section 8.5). In our application, we predict the market excess return one step ahead, so for each time t, we draw from the one-step-ahead predictive density at each iteration of the Gibbs sampler. The average of those draws is taken as the point forecast for y_{t+1}. Those forecasts can, then, be compared to their respective realized values. Notice that only information available at time t is used to generate the out-of-sample forecast.

(*footnote continued*)
positive values does not materially affect any of the results obtained under the baseline prior. The stark differences in the behavior of the predictors and, more importantly, in the degree of predictability across regimes are fully confirmed for the G7 sample. Since the results are qualitatively quite similar to those in Table 2 and are reported in prior drafts, we conserve space and do not report them here.

4. Data

The sample includes country index returns, short bond yields, term spreads, and market dividend yields from 1973 to 2007 (Canada, France, Germany, Italy, Japan), 1965 to 2007 (UK), or 1953 to 2007 (US). Panel A of Table 1 shows the span of data available for each of the G7 countries, ranging from 36 to 56 years. Overall, the span is long enough to pick up several business cycles.

4.1. International data

The country index returns for Canada, France, Germany, Italy, Japan, and the UK are from Thomson Financial's Datastream (*TOTMK*) series in local currencies. As a check, we compare these total market returns to local value-weighted returns from Kenneth French's Web site. The respective total return series are at least 95% correlated between the two sources, but we employ the Datastream return series because it has better overall coverage. We form excess returns by subtracting the appropriately de-annualized short-term interest rate. The market dividend yield is the log of the Morgan Stanley Country Indices (MSCI) dividend yield, available through Datastream as *TOTMK*(*DY*).

The term structure data come from the Global Financial Database. We take the monthly short rate as the three-month Treasury bill. The term spread is the difference in yields between the 10-year government bonds and the short rate. For Japan, the three-month Treasury bill series from the Global Financial Database series exhibits strange behavior (a series of flat plateaus or truncations) in the 1973 to 1985 period. Therefore, we draw the same series from Global Insight which exhibits more expected patterns but a slightly shorter span. The two series are virtually the same from 1985 onward (98.5% correlated).

4.2. US data

The US data span April 1953 to December 2007. The data series are limited to 658 periods (657 net observations) because yields are not available prior to April 1953. Even if such data were available, the 1951 Treasury Accord imposes a structural break often considered the advent of the modern macroeconomic era.[10]

[10] The Accord officially ended the artificial pegging of long-term nominal interest rates, and effectively concentrated control of monetary policy in the Federal Reserve as opposed to the Treasury Department.

Table 1
Sample and regime characteristics.

This table shows the sample and regime characteristics for the G7 countries across expansions and recessions. We determine the expansion and recession states with a regime-switching vector autoregression (RSVAR) which takes the form, $y_t = c(s_t) + A(s_t)y_{t-1} + \varepsilon(s_t)$, where $y_t = (r_t, DY_t, SR_t, TERM_t)'$, $c(s_{\cdot t})$ is a vector of constants, $A(s_t)$ is the coefficient matrix, and $\varepsilon(s_t) \sim N(0, \Sigma(s_t))$ is an error vector with a generalized error covariance matrix all in state $s_t \in \{1,2\}$. For the United States and Canada, $y_t = (r_t, DY_t, SR_t, TERM_t, DEF_t)'$. Panel A shows the span of monthly country-level data obtained from Datastream and Global Financial Database. *Recession* is the number of total months flagged by the RSVAR as a recession state, a state with higher volatility and lower persistence. Panel B shows the persistence of expansion and recession states, $\hat{p}_{1,1} = Prob(s_t = Expansion | s_{t-1} = Expansion)$ and $\hat{p}_{2,2} = Prob(s_t = Recession | s_{t-1} = Recession)$, respectively. Panel B also shows summary statistics regarding fit of the RSVAR. The *RCM* statistic measures the ability of the algorithm to distinguish between expansion and recession states on a scale from zero to one, where one is perfect ability to distinguish states; $RCM = 1 - 4 \cdot T^{-1} \sum_{t=0}^{T} [p(s_t = 1 | \mathcal{F}_{t-1}) p(s_t = 2 | \mathcal{F}_{t-1})]$. *Agreement* shows the percentage of observations where the RSVAR state is the same as those given either by the NBER (for the US) or by the ECRI (for the other G7 countries). Panel C shows the conditional and unconditional market risk premiums. Panel D shows differences in predictor volatilities across regimes, and Panel E shows differences in predictor half-lives across regimes. In square brackets below each point estimate in Panels B through E are 95% highest posterior density intervals based upon 50,000 Gibbs sampler iterations from a suitably constructed Markov chain. *Difference* is defined as *Recession* minus *Expansion* and is computed for each iteration.

Panel A: Sample Characteristics

Country	Start	End	Periods	Recession
Canada	1973:01	2007:12	420	69
France	1973:01	2007:12	420	91
Germany	1973:01	2007:12	420	60
Italy	1973:01	2007:12	420	71
Japan	1973:01	2007:12	420	215
United Kingdom	1965:01	2007:12	516	93
United States	1953:03	2007:12	658	184

Panel B: Regime characteristics

Country	$\hat{p}_{1,1}$	$\hat{p}_{2,2}$	RCM	Agreement (%)
Canada	0.93	0.72	0.92	80
	[0.89, 0.96]	[0.60, 0.82]	[0.90, 0.94]	
France	0.85	0.59	0.82	72
	[0.80, 0.90]	[0.47, 0.70]	[0.79, 0.85]	
Germany	0.88	0.51	0.86	42
	[0.82, 0.92]	[0.38, 0.64]	[0.82, 0.90]	
Italy	0.87	0.52	0.89	77
	[0.83, 0.91]	[0.39, 0.64]	[0.87, 0.91]	
Japan	0.88	0.85	0.82	60
	[0.82, 0.93]	[0.75, 0.93]	[0.77, 0.88]	
United Kingdom	0.87	0.56	0.87	77
	[0.83, 0.91]	[0.45, 0.67]	[0.85, 0.89]	
United States	0.91	0.80	0.88	77
	[0.88, 0.94]	[0.73, 0.86]	[0.86, 0.90]	

Panel C: Market risk premiums

Country	Unconditional	Expansion	Recession
Canada	0.297	0.199	0.666
	[0.026, 0.762]	[0.021, 0.519]	[0.064, 1.727]
France	0.633	0.516	0.958
	[0.108, 1.237]	[0.089, 1.017]	[0.162, 1.888]
Germany	0.401	0.369	0.530
	[0.044, 0.902]	[0.039, 0.864]	[0.054, 1.318]
Italy	0.452	0.380	0.718
	[0.049, 1.142]	[0.040, 0.980]	[0.077, 1.142]
Japan	0.126	0.123	0.129
	[0.037, 0.853]	[0.019, 0.798]	[0.007, 1.041]
United Kingdom	0.605	0.375	1.405
	[0.147, 1.073]	[0.091, 0.671]	[0.344, 2.498]
United States	0.525	0.344	0.925
	[0.195, 0.840]	[0.118, 0.580]	[0.323, 1.554]

Panel D: Differences in volatilities across regimes

Country	*r*	*DY*	*SR*	*TERM*	*DEF*
Canada	3.175	−0.036	0.985	0.199	0.214
	[2.674, 3,546]	[−0.068, 0.005]	[0.692, 1.224]	[0.061, 0.327]	[0.178, 0.250]
France	1.978	0.021	−0.133	0.138	
	[1.312, 2.749]	[0.000, 0.043]	[−0.467, 0.192]	[0.261, 0.991]	
Germany	0.982	0.039	0.553	0.101	
	[−0.828, 2.353]	[−0.000, 0.073]	[0.384, 0.730]	[−0.004, 0.207]	
Italy	2.410	0.042	−0.341	0.150	
	[1.223, 3.687]	[0.021, 0.064]	[−0.677, 0.006]	[0.031, 0.302]	

Table 1 (*continued*)

Panel D: Differences in volatilities across regimes

Country	*r*	*DY*	*SR*	*TERM*	*DEF*
Japan	−0.162	−0.207	−0.449	−1.292	
	[−0.912, 0.356]	[−0.269, −0.162]	[−0.670, −0.224]	[−1.365, −1.175]	
United Kingdom	4.133	−0.056	−0.077	0.310	
	[3.657, 4.689]	[−0.081, −0.026]	[−0.265, 0.125]	[0.202, 0.427]	
United States	2.407	0.032	1.368	0.245	0.237
	[2.193, 2.624]	[0.006, 0.059]	[1.254, 1.488]	[0.195, 0.298]	[0.206, 0.254]

Panel E: Differences in half-lives across regimes

Country	*DY*	*SR*	*TERM*	*DEF*
Canada	−54	−28	−41	−7
	[−58, −38]	[−55, 15]	[−59, −10]	[−15, −3]
France	−17	−35	−21	
	[−43, 2]	[−57, 12]	[−59, −6]	
Germany	−24	−53	−34	
	[−55, 15]	[−59, −32]	[−56, −1]	
Italy	−39	−52	−17	
	[−58, −11]	[−59, −30]	[−47, −3]	
Japan	−38	27	7	
	[−51, 0]	[−7, 49]	[−1, 25]	
United Kingdom	−33	−45	−43	
	[−59, −11]	[−57, −21]	[−59, −11]	
United States	−6	−44	−15	−8
	[−29, 0]	[−53, −19]	[−36, 5]	[−18, −2]

Excess monthly returns are from Ken French's Web site. As in Campbell and Shiller (1988), we construct the dividend yield, *DY*, using the sum of dividends on the CRSP value-weighted market portfolio smoothed over the past 12 months and scaled by the price level of the index. The dividends are calculated as the log of the difference between the total returns with dividends (*VWRETD*) and the capital gains (*VWRETX*). The US dividend yield from CRSP and from Datastream are 96% correlated. To account for share repurchases, we follow the procedure of Bansal, Dittmar, and Lundblad (2005). Net share repurchases can be calculated using changes in the number of shares outstanding along with the CRSP split factor.

The term structure variables are from the Federal Reserve Economic Database (*FRED*), the Web-based economic data source made available by the Federal Reserve Bank in St. Louis. The Treasury bill yield, *SR*, is the 30-day constant maturity yield. The term spread, *TERM*, is the difference between the 10-year Treasury bond and the one-year constant maturity Treasury bill. We use the 10-year Treasury bond because the 20-year and 30-year series have several years of missing (discontinued) data and the 30-year bond has been discontinued. The default spread, *DEF*, is the difference between Moody's BAA and AAA bond yields.

5. Results

In this section, we first report the return predictability results for our international sample. We examine risk premiums, the statistical properties of predictors, explanatory power, and changes in coefficients across good and bad times. Next, using US data, we evaluate alternative definitions of regimes at various horizons. Finally, we briefly investigate the out-of-sample performance of the RSVAR compared to the historical average and to a VAR.

5.1. Regime characteristics across the G7 countries

For each of the G7 countries, we estimate the RSVAR specification from Eq. (1) of Section 3.3. We run the Gibbs sampler outlined in Appendix A and summarize each posterior distribution of interest by its posterior mean. Immediately beneath each mean in the tables, we report the accompanying 95% HPD interval in square brackets.

Panel A of Table 1 displays the country-by-country data span and the frequency of recessions flagged by the RSVAR. The recession regime, identified by higher market and predictor volatility, ranges from about 14% of the sample for Germany to over 50% in Japan. The recession regime represents 28% of the US sample.

Panel B shows that the regime transition probabilities vary considerably across countries. The relatively narrow HPD intervals indicate that these estimates are quite precise. Across all countries, the probability of going from expansion state to expansion state, denoted by $\hat{p}_{1,1}$ in Panel B, is highest for Canada and the US, at 93% and 91%, respectively. This probability is somewhat lower for the other countries, at 85% to 88%. As expected, the probability of persisting in the recession regime, denoted by $\hat{p}_{2,2}$, is much lower, ranging from 51% to 80%. The expected duration of an expansion averages 9.3 months (evaluated at the posterior means) compared to less than half that, or four months, for recessions.

Despite the variation in the evolution of the state variables across countries, the *RCM* statistic shows that regimes can be easily distinguished. The RCM measure used by Ang and Bekaert (2002b) is

$$RCM(M=2) = 1 - 4T^{-1} \sum_{t=1}^{T} \left(\prod_{m=1}^{2} p(s_t = m | \mathcal{F}_{t-1}) \right), \qquad (3)$$

where M is the number of regimes and $p(s_t=m|\mathcal{F}_{t-1})$ is the probability that the process is in state m conditional on the filtration $\mathcal{F}$ at time $t-1$. If the algorithm cannot distinguish between two regimes from the data, then $p(s_t=1|\mathcal{F}_{t-1}) = p(s_t=2|\mathcal{F}_{t-1}) = 0.5 \forall t \in \{1,2,\ldots,T\}$ and $RCM(2)=0$. If the regimes are perfectly distinguishable (i.e., $p(s_t=1|\mathcal{F}_{t-1})=1$ or $p(s_t=2|\mathcal{F}_{t-1})=1\ \forall t \in \{1,2,\ldots,T\}$), then $RCM(2)=1$. Therefore, the *RCM* statistic is a measure of the fit of the regimes, similar in spirit to the R^2. The *RCM* statistic is highest for Canada (92%), followed by Italy (89%), and the US (88%); it is lowest for Japan (82%) and France (82%). Even an *RCM* of 82% suggests relative certainty about the state of the process: it corresponds to an average state probability that is within 0.0475 of absolute certainty of being in either state 1 or 2.

For external validation, we next compare our RSVAR regimes to those provided by the NBER for the US and by the ECRI for the other G7 countries. In the last column of Table 1 Panel B, we define the *Agreement* between the RSVAR-identified states and those provided by the NBER and the ECRI as the percentage of observations where both measures indicate the same state. Since the RSVAR provides a continuous measure of state probabilities between 0 and 1, we assign the state as recession for values of 0.5 to 1 inclusive. *Agreement* levels are between 72% and 80% for the US, Canada, France, Italy, and UK. The numbers are lower for Japan (60%) and Germany (42%).[11]

Fig. 3 depicts RSVAR recession probabilities as stars (under 0.5 is expansion, above 0.5 is contraction), and real GDP growth (from Global Financial Database) as a line using a right y-axis scaled from -2% to 3% for all charts. Overall, the data-driven RSVAR contraction regimes appear correlated with below-average real GDP growth. Compared to the NBER sample, the RSVAR identifies more frequent recessions, but of shorter average duration. Short duration events are unlikely to warrant labeling by the NBER. The more frequent, shorter recession indications from the RSVAR are also consistent with Paul A. Samuelson's quip that "Wall Street indexes predicted nine of the last five recessions!"[12] To compare our classification of regimes to other studies, we compute the simple correlation of RSVAR and NBER states for the US as 40%. This compares to 31% correlation for the large portfolio of Perez-Quiros and Timmermann (2000, p. 1245) and 11% correlation for the regime-switching term structure model of Bansal, Tauchen, and Zhou (2004, p. 13).

In Panel C we report posterior summaries for the market risk premium implied by the RSVAR estimates. In order to control for the effects of shifts in investment opportunity sets across regimes, the computations follow the methodology of Mayfield (2004). The realized return moments are used to solve for the preference parameters of the Mayfield model, and these in turn give the conditional risk premiums reported. We repeat the computations for the full set of simulations to obtain HPD intervals for the estimates. Overall, the HPD intervals are large but do not include zero.[13] Although statistical precision is certainly an issue, the posterior means are higher in recessions than in expansions for all markets. Albeit limited, this evidence is consistent with higher expected returns during high volatility periods. The findings are also consistent with the view that market participants price in higher risk premiums as economic conditions worsen.

A common feature across countries is the increased persistence and reduced volatility of the predictors in the expansion regime. Panel D shows the posterior summaries for differences in volatility between bad and good times for each of the RSVAR variables. Overwhelmingly, the posterior means are positive, indicating higher volatility in bad times. The exceptions all have zero inside their HPD intervals, save Japan and the dividend yield in the UK.

Panel E reports the differences in the persistence of the predictors across economic times. As customary in multiple time-series analysis, we measure persistence by computing the half-life of shocks to each variable in the system. The half-life is computed as the number of periods needed to halve the impulse response function following an initial unit shock. The reported numbers are the differences between half-lives in bad and good times for each variable in each country. With the exception of *SR* and *TERM* in Japan, the posterior means are negative and, with a few exceptions, the HPD intervals exclude zero, indicating that most predictors tend to be substantially more persistent during good times.

5.2. Predictability across the G7 countries

Table 2 reports the overall $\overline{R}^2$ of a full sample VAR of the market return, the dividend yield, the short rate, and the term spread (with a constant included), followed by the $\overline{R}^2$ s for the regressions conditional on the state. The seven-country average full sample $\overline{R}^2$ is 2.88%. The individual values range from 0.9% for Italy up to 5.0% for the United Kingdom.

More importantly, for the expansion regime the 95% HPD intervals for the estimated $\overline{R}^2$s include zero without exception. As a result, very little confidence can be placed on the existence of any predictability at all during expansions. The seven-country average $\overline{R}^2$ is 1.63%, with a range from 0.4% for the United States up to 3.5% for France.

[11] For Germany, we actually find little evidence that the data can be characterized by two regimes. As we will show, the model tests fail to reject the single-regime null, the regimes show no difference in risk premiums, and the difference in predictability across regimes is negligible. Given the rather modest agreement between the RSVAR estimates and the ECRI dates, we re-estimated the RSVAR with the state indicator predetermined by the ECRI dates. We find similar results to those reported, confirming that the German market does not appear to share the countercyclical predictability patterns that dominate other developed markets. The contrast with Germany highlights the stylized facts shown for the other economies.

[12] From Paul A. Samuelson's "Science and Stocks" column in Newsweek, September 19th, 1966.

[13] In solving the Mayfield preference model for all Gibbs sampler draws of the return moment estimates, there are instances of non-convergence in the numerical procedure. Discarding those draws, less than 10% of the total, leaves a still sizable sample for determining the HPD intervals reported in Panel C.

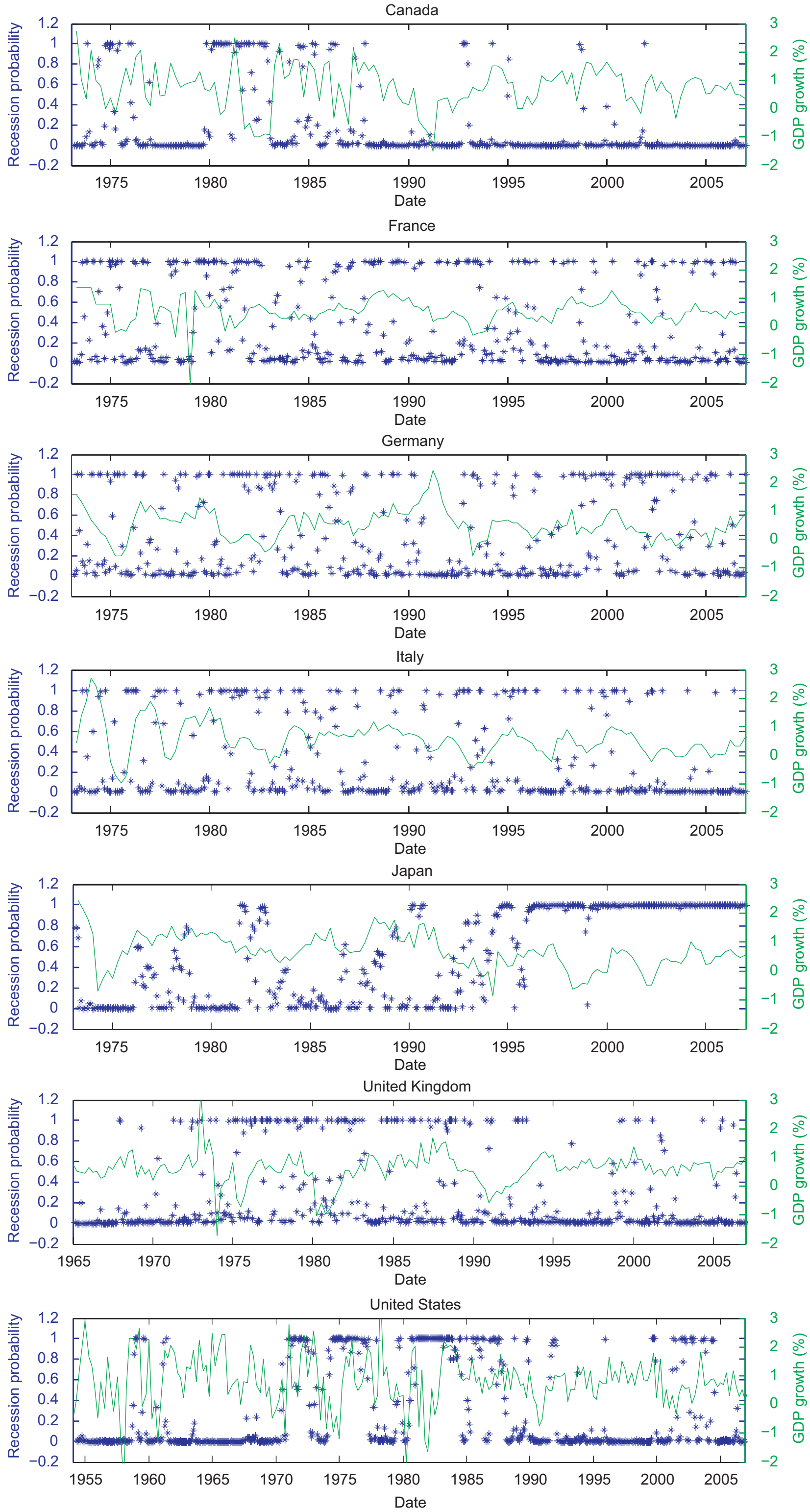

Fig. 3. RSVAR estimates of regimes for the G7. Plotted are the monthly probability of recessions (asterisks) as flagged by the RSVAR and quarterly real GDP growth (solid line). The *y*-axis on the left is the probability of recession. A value of zero indicates an expansion (state 1) and a value of one indicates a recession (state 2). A value of 0.5 represents uncertainty about the regime, or equal probability that the state is an expansion or a recession. The scale for GDP growth on the second *y*-axis is from −2% to 3% for all charts. The RSVAR was estimated using Markov Chain Monte Carlo (MCMC) techniques. The estimates of the posterior distributions are based on 50,000 Gibbs sampler iterations from a suitably constructed Markov chain.

Table 2
Predictive explanatory power in good and bad times.

This table shows the explanatory power of return predictability for the G7 sample of countries across expansions and recessions. We determine the expansion and recession states with a regime-switching vector autoregression (RSVAR) which takes the form, $y_t = c(s_t)+A(s_t)y_{t-1}+\varepsilon(s_t)$, where $y_t = (r_t, DY_t, SR_t, TERM_t)'$, $c(s_t)$ is a vector of constants, $A(s_t)$ is the coefficient matrix, and $\varepsilon(s_t) \sim N(0, \Sigma(s_t))$ is an error vector with a generalized error covariance matrix all in state $s_t \in \{1,2\}$. For the United States and Canada, $y_t = (r_t, DY_t, SR_t, TERM_t, DEF_t)'$. The table reports the (average) adjusted R^2's from predictive regressions of next month's country-level market excess return on this month's short rate, term spread, and dividend yield. The adjusted R^2s are first for the full sample and then shown for the expansion and recession periods. The 95% highest posterior density confidence intervals (in square brackets) are based on 50,000 Gibbs sampler iterations from a suitably constructed Markov chain. *Difference* is defined as *Recession* minus *Expansion* and is computed for each iteration.

Country	Full sample	Expansion	Recession	Difference
Canada	0.030	0.011	0.152	0.141
	[0.002, 0.067]	[−0.010, 0.046]	[0.009, 0.315]	[0.006, 0.307]
France	0.030	0.035	0.073	0.038
	[0.002, 0.068]	[0.000, 0.084]	[−0.016, 0.194]	[−0.064, 0.165]
Germany	0.015	0.022	0.090	0.068
	[−0.005, 0.046]	[−0.005, 0.065]	[−0.047, 0.313]	[−0.074, 0.290]
Italy	0.009	0.006	0.129	0.123
	[−0.007, 0.037]	[−0.010, 0.036]	[−0.016, 0.343]	[−0.024, 0.334]
Japan	0.013	0.017	0.152	0.135
	[−0.005, 0.042]	[−0.016, 0.080]	[0.051, 0.254]	[0.011, 0.246]
United Kingdom	0.050	0.015	0.253	0.238
	[0.018, 0.090]	[−0.005, 0.048]	[0.109, 0.398]	[0.089, 0.386]
United States	0.049	0.004	0.175	0.171
	[0.019, 0.085]	[−0.008, 0.027]	[0.082, 0.276]	[0.076, 0.273]

In contrast, for recessions the seven-country average is 15.1%, with a range from 7.3% for France up to 25.3% for the United Kingdom. Comparing regimes, the *Difference* column of Table 2 indicates that $\overline{R}^2$s in the recession regime are larger in all seven countries. For all countries except Germany, 90% of the posterior distribution of the difference does not include zero. For five countries, the difference is statistically significant at the 5% level.

Having assessed the overall predictive power of the RSVAR, we now analyze the sources of the observed differences between regimes. We begin with a joint test of no shift across regimes in any of the predictive coefficients in the market equation. From a Bayesian perspective, this amounts to a model comparison between an unrestricted specification where all autoregressive coefficients are allowed to change and a restricted model where the coefficients in the first equation of the RSVAR are forced to be constant across regimes. Assuming equal prior probabilities for the two specifications, the evidence against the constant coefficients is provided by the posterior-odds ratio. Results are in Table 3 Panel A. With the exception of Germany, the posterior odds indicate that the predictive coefficients are, indeed, different across regimes. We also compute Bayes factors to compare the one-regime (i.e., linear) homoskedastic VAR with the two-regime VAR. For all seven countries the (unreported) posterior odds are all larger than 10^6, a magnitude denoting overwhelming evidence in favor of the two-state specification, possibly due to heteroskedasticity.

We also investigate differences between regimes at the individual predictor level. Two features of the posterior estimates of the predictive coefficients reported in Table 3 Panel B are especially relevant for our study. First, it is never the case that a predictor's 95% HPD interval excludes zero in expansions but not in recessions. Moreover, no predictor in any country appears to be more important in good times than in bad times. Second, the differences in predictive coefficients between regimes are comfortably distributed away from zero for most countries. Where there are differences in predictive power, the larger predictive power occurs in bad times.

5.3. Decomposition of R^2

The higher R^2 of predictive regressions during recessions is not a variance effect due to conditioning bias. Here we decompose the R^2 to consider the source of increased explanatory power. For a regression of r_t (the excess market return, in our analyses) on predictors $y_1, \ldots, y_k$ at $t-1$, we can write the R^2 as

$$R^2 = \sum_{i=1}^{k} \beta_i^2 \left(\frac{\sigma_i^2}{\sigma_r^2}\right) + \sum \sum_{i \neq j} \frac{\beta_i \beta_j \sigma_{ij}}{\sigma_r^2}, \tag{4}$$

where $\beta_1, \ldots, \beta_k$ correspond to the first row of the matrix $A(s_t)$ in Eq. (1), $\sigma_{i,j}$ corresponds to the $i,j \in \{1, \ldots, k\}$ element of the variance–covariance matrix of predictors, $y_{t-1}y'_{t-1}$, and σ_i corresponds to the ith diagonal element of $y_{t-1}y'_{t-1}$. Thus, each of the first k terms is the product of a squared beta and a relative volatility. Is there a potential conditioning bias in our application? Each of the first k terms could increase either due to an increase in β_i, or to a disproportionate increase in the predictor variance σ_i^2 relative to σ_r^2.

Table 4 Panel A shows the overall R^2 for expansion and for recession, and this figure is decomposed into the first k terms and the sum of the cross terms, as in Eq. (4). For every country, the increased R^2 in recessions is associated with increases in the first k terms, and a decrease in the cross-term sums. The first k terms are associated with predictors individually, and the cross terms with predictors in covariance pairs.

Table 3

Comparison of RSVAR coefficients country by country.

This table shows the mean RSVAR coefficients for the G7 sample of countries across expansions and recessions and the difference between the two. We determine the expansion and recession states with a regime-switching vector autoregression (RSVAR) which takes the form, $y_t = c(s_t) + A(s_t)y_{t-1} + \varepsilon(s_t)$, where $y_t = (r_t, DY_t, SR_t, TERM_t)'$, $c(s_t)$ is a vector of constants, $A(s_t)$ is the coefficient matrix, and $\varepsilon(s_t) \sim N(0, \Sigma(s_t))$ is an error vector with a generalized error covariance matrix all in state $s_t \in \{1,2\}$. For the US and Canada samples, $y_t = (r_t, DY_t, SR_t, TERM_t, DEF_t)'$. Panel A shows the posterior odds against the hypothesis of jointly constant coefficients across regimes. The RSVAR is estimated using Markov Chain Monte Carlo (MCMC) techniques. Panel B shows the posterior mean and the 95% highest posterior density intervals (in square brackets) based on 50,000 Gibbs sampler iterations from a suitably constructed Markov chain. *Difference* is defined as *Recession* minus *Expansion* and is computed for each iteration.

Panel A: Posterior-odds ratios against jointly constant coefficients

Country	Odds ratio	Evidence against constant coefficients
Canada	20.15	Strong
France	15.66	Strong
Germany	0.006	Minimal/not worth mentioning
Italy	5.13	Substantial
Japan	151.36	Decisive
United Kingdom	11.12	Strong
United States	199.04	Decisive

Panel B: Regression coefficients

Canada	VAR	Expansion	Recession	Difference
DY_{t-1}	1.690	0.719	4.663	3.944
	[−0.093, 3.487]	[−0.855, 2.362]	[−2.887, 12.358]	[−0.373, 11.981]
SR_{t-1}	−0.372	−0.105	−1.108	−0.976
	[−0.621, −0.122]	[−0.342, 0.136]	[−1.851, −0.345]	[−1.793, −0.186]
$TERM_{t-1}$	−0.170	0.046	−0.892	−0.938
	[−0.569, 0.228]	[−0.316, 0.426]	[−2.266, 0.456]	[−2.373, 0.482]
DEF_{t-1}	1.154	−0.385	5.198	5.584
	[−0.172, 2.453]	[−1.869, 1.088]	[0.601, 9.734]	[0.644, 10.432]
France	**VAR**	**Expansion**	**Recession**	**Difference**
DY_{t-1}	4.259	2.172	10.115	7.943
	[1.275, 7.322]	[0.154, 5.118]	[2.884, 17.519]	[0.010, 15.894]
SR_{t-1}	−0.387	−0.210	−0.948	−0.739
	[−0.670, −0.102]	[−0.483, 0.028]	[−1.845, −0.117]	[−1.690, 0.166]
$TERM_{t-1}$	−0.260	0.357	−1.681	−2.037
	[−0.924, 0.410]	[−0.368, 1.050]	[−3.237, −0.098]	[−3.787, −0.223]
Germany	**VAR**	**Expansion**	**Recession**	**Difference**
DY_{t-1}	0.881	1.063	3.523	2.460
	[−0.700, 2.555]	[0.057, 2.682]	[0.189, 9.202]	[−1.535, 8.461]
SR_{t-1}	−0.188	−0.273	−0.312	−0.039
	[−0.543, 0.172]	[−0.657, 0.110]	[−1.473, 0.842]	[−1.319, 1.271]
$TERM_{t-1}$	0.159	0.171	−0.656	−0.827
	[−0.420, 0.743]	[−0.443, 0.799]	[−2.556, 1.225]	[−2.860, 1.232]
Italy	**VAR**	**Expansion**	**Recession**	**Difference**
DY_{t-1}	0.904	1.131	3.455	2.324
	[−1.110, 2.995]	[0.056, 2.980]	[0.111, 8.429]	[−1.843, 7.609]
SR_{t-1}	0.012	−0.000	0.218	0.219
	[−0.126, 0.150]	[−0.170, 0.163]	[−0.261, 0.849]	[−0.318, 0.959]
$TERM_{t-1}$	0.229	0.247	−0.016	−0.263
	[−0.279, 0.740]	[−0.347, 0.836]	[−1.443, 1.508]	[−1.889, 1.510]
Japan	**VAR**	**Expansion**	**Recession**	**Difference**
DY_{t-1}	1.100	−0.151	3.486	3.637
	[0.115, 2.214]	[−2.514, 2.173]	[1.446, 5.732]	[0.328, 6.551]
SR_{t-1}	−0.114	0.106	−0.544	−0.689
	[−0.343, 0.108]	[−0.852, 0.971]	[−0.866, −0.229]	[−1.545, −0.401]
$TERM_{t-1}$	0.018	0.283	0.268	−0.018
	[−0.457, 0.488]	[−0.859, 1.357]	[−0.740, 1.287]	[−1.518, 1.487]
United Kingdom	**VAR**	**Expansion**	**Recession**	**Difference**
DY_{t-1}	5.728	1.896	26.384	24.488
	[3.113, 8.381]	[0.203, 4.054]	[16.294, 35.388]	[14.043, 33.814]
SR_{t-1}	−0.318	−0.167	−1.611	−1.444
	[−0.544, −0.091]	[−0.378, 0.030]	[−2.514, −0.658]	[−2.377, −0.464]

Table 3 (*continued*)

Panel B: Regression coefficients Canada	VAR	Expansion	Recession	Difference
$TERM_{t-1}$	−0.371	−0.207	−1.626	−1.419
	[−0.818, 0.073]	[−0.656, 0.213]	[−3.015, −0.178]	[−2.888, 0.112]
United States	VAR	Expansion	Recession	Difference
DY_{t-1}	1.076	0.466	3.114	2.648
	[0.209, 1.955]	[0.027, 1.193]	[0.707, 5.750]	[0.127, 5.383]
SR_{t-1}	−0.381	−0.080	−0.634	−0.554
	[−0.573, −0.193]	[−0.286, 0.121]	[−1.070, −0.210]	[−1.042, −0.066]
$TERM_{t-1}$	−0.015	−0.159	0.305	0.464
	[−0.431, 0.402]	[−0.586, 0.264]	[−0.744, 1.356]	[−0.699, 1.633]
DEF_{t-1}	1.906	0.534	3.879	3.346
	[0.726, 3.046]	[−1.181, 2.241]	[1.504, 6.254]	[0.355, 6.322]

Table 4 Panel B shows the squared beta, β_i^2, for each predictor $i \in \{1,\ldots,k\}$ in expansion and recession regimes. Panel C shows the change in relative volatility, σ_i^2/σ_r^2, for each predictor $i \in \{1,\ldots,k\}$ in expansion and recession regimes. Whereas the squared betas in Panel B are notably larger in recessions, the relative volatilities in Panel C are roughly comparable across regimes if not actually smaller in recessions. Thus, the increased R^2 in recessions is due to the beta terms, not the relative volatility terms.

In turn, we infer that the changing betas are driven by changing correlations between the predictors $y_{i,t\text{-}1}$ and r_t since

$$\beta_i = \rho_{ir}\frac{\sigma_r}{\sigma_i}, \tag{5}$$

and Table 4 Panel C shows that the relative volatilities change little during recessions. In a simple regression with only one y, the correlation term ρ in (5) would be the simple Pearson correlation between r and y. With multiple regressors, ρ_{ir} is instead the part correlation between r and y_i, a measurement of comovement between r and y_i on a basis independent of the other predictors.[14] In summary, short-horizon return predictability seems due to changing correlations rather than to changes in the volatility of the market relative to predictors.

5.4. Current vs. expected business conditions

Table 5 first shows the posterior summaries from the estimation of predictive VARs where the regimes are predetermined by monthly NBER business cycle dates for the 1953 to 2007 period. These VARs are special cases of the RSVARs considered earlier, so estimation and inference through Bayesian MCMC methods present no additional challenge. The results cleanly reaffirm the previous findings using the RSVAR approach and the international sample. The differences in R^2 and in coefficient values across regimes are comparable to those in Table 3. Full-sample return predictability is driven predominantly by NBER recession periods, in which the $\overline{R}^2$s rise from less than 1.5% during expansions to 27.8% during recessions. If predictability exists, it does so primarily in recessions.

To investigate whether the RSVAR tends to pick up expected but unrealized recessions, we apply alternative definitions of good and bad times, based on expectations provided by professional surveys. We consider the Survey of Professional Forecasters (SPF) and the Livingston Survey. These surveys provide forecasts at the quarterly and semiannual horizons, and so may identify whether the RSVAR picks up current, short-term, or longer-term expectations about the economy.

For the SPF, we use the forecast of next quarter's GDP deflated by the forecast of the next quarter Consumer Price Index (CPI), and define the bottom 30% of the sample to be "bad" times. Although not reported, the results are essentially unchanged when using 20% or 25% cutoffs. The results are also very similar when using an absolute standard of zero real GDP growth to distinguish good from bad times. The SPF is at quarterly frequency from 1968 through 2007. The results are in Table 5, together with results using NBER regimes at quarterly frequency over the same period. The results using SPF expectations about the economy compare quite favorably to those using NBER determinations of the actual current state of the economy. If anything, the SPF-based results are even stronger.

Next, we employ the Livingston Survey to classify good and bad times. Livingston Survey respondents are asked to forecast GDP growth over a six-month period. We use the forecast of the next six months' GDP (deflated by the forecast of the CPI over the same time span), and define the bottom 30% of the sample to be "bad" times. Although not reported, the results are essentially unchanged when using 20% or 25% cutoffs. Using an absolute standard of zero real GDP growth to distinguish good from bad times is not feasible with the Livingston Survey data as only a handful of observations would define the bad state.

Results from predictive VARs on semiannual data are in Table 5. The aggregation of quarterly NBER dates to the semiannual frequency is done as follows. If two consecutive quarters during the first or second half-year period are flagged as a recession (expansion), the corresponding semester is flagged the same way. If the first quarter of the half-year period is a recession (expansion) and the second is an expansion (recession), the semiannual period is flagged as a recession

[14] Also known as semipartial correlation, ρ_{ir} is a reminder that multiple regression coefficients must be interpreted on a marginal basis, holding all other right-hand-side variables constant.

Table 4
A decomposition of return predictability.

For the regression of y. ($R^e_{VW,t+1}$) on predictors $x_1, \ldots, x_k$ (r_t, DY_t, SR_t, $TERM_t$, DEF_t), the R^2 is

$$R^2 = \sum_{i=1}^{k} \beta_i^2 \left(\frac{\sigma_i^2}{\sigma_y^2} \right) + \sum \sum_{i \neq j} \frac{\beta_i \beta_j \sigma_{ij}}{\sigma_y^2}.$$

Panel A shows the overall R^2 for expansions and for recessions, decomposed into the first k terms and the sum of the cross terms. Each of the first k terms is the product of a squared beta and a variance ratio. Panel B shows the squared beta, β_i^2, for each predictor $i \in \{1,\ldots,k\}$ in expansion and recession regimes. Panel C shows the change in relative volatility, σ_i^2/σ_y^2, for each predictor.

	Canada		France		Germany		Italy		Japan		United Kingdom		United States	
	Exp	Rec	Exp	Rec	Exp	Rec	Exp	Rec	Exp	Rec	Exp	Rec	Exp	Rec
Panel A: Decomposition of conditional R^2														
R^2	0.011	0.152	0.035	0.073	0.022	0.090	0.021	0.182	0.017	0.152	0.015	0.253	0.004	0.175
Sum of:														
r	0.002	0.001	0.006	0.006	0.000	0.054	0.000	0.028	0.010	0.054	0.010	0.018	0.000	0.006
DY	0.004	0.044	0.017	0.233	0.005	0.067	0.014	0.306	0.000	0.059	0.013	0.492	0.003	0.046
SR	0.008	0.431	0.020	0.189	0.012	0.024	0.001	0.410	0.004	0.064	0.013	0.291	0.002	0.130
TERM	0.000	0.054	0.006	0.092	0.002	0.025	0.017	0.054	0.012	0.001	0.003	0.084	0.002	0.004
DEF	0.001	0.153											0.001	0.109
Cross terms	−0.005	−0.532	−0.014	−0.447	0.004	−0.080	−0.012	−0.617	−0.008	−0.027	−0.024	−0.633	−0.004	−0.119
Panel B: Squared Betas														
r	0.002	0.002	0.005	0.008	0.000	0.071	0.000	0.040	0.001	0.059	0.008	0.025	0.000	0.007
DY	0.516	21.741	4.717	102.317	1.130	12.409	1.596	38.479	0.023	12.151	3.594	696.130	0.217	9.698
SR	0.011	1.169	0.044	0.899	0.074	0.097	0.002	1.909	0.011	0.295	0.028	2.596	0.006	0.402
TERM	0.002	0.796	0.127	2.824	0.029	0.430	0.624	2.083	0.080	0.070	0.043	2.642	0.025	0.093
DEF	0.149	27.022											0.285	15.049
Panel C: Relative volatility														
r	1.194	0.716	1.151	0.695	1.055	0.762	1.162	0.704	1.096	0.929	1.223	0.723	1.188	0.834
DY	0.009	0.002	0.004	0.002	0.004	0.005	0.009	0.008	0.012	0.005	0.004	0.001	0.012	0.005
SR	0.714	0.368	0.461	0.210	0.162	0.244	0.460	0.215	0.315	0.216	0.469	0.112	0.364	0.324
TERM	0.174	0.068	0.044	0.033	0.052	0.059	0.028	0.026	0.145	0.015	0.075	0.032	0.083	0.040
DEF	0.008	0.006											0.005	0.007

(expansion). While the NBER data once again show much stronger predictability in bad times than in good times, the Livingston data do not produce any significant predictability differences between good and bad times. HPD confidence intervals for regression coefficient differences and for the R^2 difference all include zero.

Overall, the results suggest that our core RSVAR methodology may be identifying short-term (a few months to a quarter) expected conditions in addition to current conditions rather than expected conditions six months to a year in the future.

5.5. Out-of-sample performance

Although our primary interest is in the controversy surrounding the academic literature rather than portfolio applications, the disparity between in-sample and out-of-sample (OOS) performance of models is itself another puzzling feature of the data. Since the RSVAR is not actually a real-time model (full-sample information is used to determine the states), it is a worthwhile question whether such a highly parametrized model can be as effective out-of-sample as the simpler historical average (HA) null of Welch and Goyal (2008).

After some experimentation, we conclude that 15 years of monthly observations appear sufficient to get sensible state and parameter estimates for the US and UK. However, at least 20, if not 25, years are needed to produce reliable inferences for the remaining markets. The forecasting exercise occurs recursively on an expanding window. For instance, the RSVAR for the US is first estimated over the 1953:04–1968:03 interval and draws from the predictive density are generated for 1968:04. Next, the model is re-estimated over 1953:04–1968:04 and a forecast is generated for 1968:05, and so forth. We use the full CRSP sample from 1926:01 to $t-1$ to estimate the historical average market return for the HA, but we also verify that the results are similar using the RSVAR training period starting in 1953:04.

Following Campbell and Thompson (2008), we compute the OOS R^2 as

$$R^2_{OOS} = 1 - \frac{\sum_{t=1}^{T} (r_t - \hat{r}_t)^2}{\sum_{t=1}^{T} (r_t - \tilde{r}_t)^2}, \tag{6}$$

where $\hat{r}_t$ is the expected return from the candidate model using data up to and including time $t-1$ and $\tilde{r}_t$ is the expected return from the null model using data up to and including time $t-1$.

For diagnostic purposes, we split the sample using three possible regime indicators. The first is the OOS state prediction at $t-1$ (*RSVAR, OOS splits*). Second, we use the RSVAR smoothed probabilities (*RSVAR, IS splits*). Last, for the US we split the sample on the basis of NBER regimes (*NBER splits*). The OOS RSVAR prediction of the state of the economy matches that from NBER dates in 72% of the observations,

Table 5
Alternative classifications of good and bad times.

This table shows the performance of predictive RSVARs for the US in expansions and recessions, and the difference between the two. The RSVAR is constrained to prespecified regime states based upon three classifications: first, NBER business cycle dates; second, the Survey of Professional Forecasters (SPF); and third, the Livingston Survey. Panel A shows the conditional R^2s. Panel B shows the regression coefficients. The 95% highest posterior density intervals (in square brackets) are based on 50,000 Gibbs sampler iterations from a suitably constructed Markov chain. *Difference* is defined as *Recession* minus *Expansion* and is computed for each iteration, then averaged.

Panel A: Adjusted R^2 Classification	Expansion	Recession	Difference
RSVAR (Monthly)	0.005	0.167	0.161
	[−0.008, 0.029]	[0.073, 0.266]	[0.065, 0.263]
NBER (Monthly)	0.015	0.278	0.263
	[−0.003, 0.042]	[0.136, 0.411]	[0.119, 0.398]
SPF (Quarterly)	0.026	0.453	0.426
	[−0.032, 0.119]	[0.242, 0.618]	[0.201, 0.607]
NBER (Quarterly)	0.037	0.372	0.335
	[−0.013, 0.108]	[0.143, 0.563]	[0.098, 0.535]
Livingston (Semiannual)	0.212	0.192	−0.020
	[0.060, 0.366]	[−0.042, 0.420]	[−0.300, 0.256]
NBER (Semiannual)	0.053	0.486	0.433
	[−0.034, 0.172]	[0167, 0.701]	[0.101, 0.674]

Panel B: Regression coefficients RSVAR (Monthly)	Expansion	Recession	Difference
DY_{t-1}	0.466	3.114	2.648
	[0.027, 1.193]	[0.707, 5.750]	[0.127, 5.383]
SR_{t-1}	−0.080	−0.634	−0.554
	[−0.286, 0.121]	[−1.070, −0.210]	[−1.042, −0.066]
$TERM_{t-1}$	−0.159	0.305	0.464
	[−0.586, 0.264]	[−0.744, 1.356]	[−0.699, 1.633]
DEF_{t-1}	0.534	3.879	3.346
	[−1.181, 2.241]	[1.504, 6.254]	[0.355, 6.322]
NBER (Monthly)	Expansion	Recession	Difference
DY_{t-1}	1.022	1.847	0.825
	[0.185, 1.937]	[0.138, 4.227]	[−1.148, 3.314]
SR_{t-1}	−0.226	−0.589	−0.363
	[−0.427, 0.035]	[−1.093, −0.069]	[−0.911, 0.191]
$TERM_{t-1}$	−0.083	1.436	1.519
	[−0.504, 0.327]	[−0.141, 3.066]	[0.115, 3.199]
DEF_{t-1}	1.041	3.862	2.822
	[−0.320, 2.421]	[1.001, 6.656]	[0.352, 5.933]
SPF (Quarterly)	Expansion	Recession	Difference
DY_{t-1}	3.232	16.350	13.118
	[−1.556, 8.238]	[7.125, 25.424]	[2.785, 23.381]
SR_{t-1}	−0.226	−3.219	−2.393
	[−1.973, 0.309]	[−5.131, −1.293]	[−4.617, −0.177]
$TERM_{t-1}$	−0.857	−0.334	0.522
	[−2.823, 1.120]	[−3.700, 3.092]	[−3.379, 4.445]
DEF_{t-1}	2.404	16.050	13.646
	[−2.295, 7.083]	[7.336, 24.644]	[3.745, 23.483]
Livingston (Semiannual)	Expansion	Recession	Difference
DY_{t-1}	12.347	3.545	−8.802
	[4.582, 20.244]	[−5.705, 12.535]	[−20.956, 3.222]
SR_{t-1}	−2.380	−1.927	0.453
	[−3.968, −0.820]	[−4.351, 0.409]	[−2.419, 3.267]
$TERM_{t-1}$	−0.049	1.490	1.539
	[−2.960, 2.895]	[−6.269, 9.401]	[−6.748, 9.973]
DEF_{t-1}	11.740	7.751	−3.989
	[3.477, 20.110]	[−12.089, 27.334]	[−25.508, 17.241]

Table 6
Out-of-sample return predictability.

This table compares the out-of-sample (OOS) performance of three predictability models. *HA* is the expected return based on the historical average return up to time $t-1$. *VAR* is a vector autoregression of y using data up to and including $t-1$. *RSVAR* is the corresponding regime-switching VAR. As have Campbell and Thompson (2008), we compute the OOS R^2 as

$$R^2_{OOS} = 1 - \frac{\sum_{t=1}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=1}^{T}(r_t - \tilde{r}_t)^2},$$

where $\hat{r}_t$ is the expected return from the candidate model using data up to and including time $t-1$ and $\tilde{r}_t$ is the expected return from the null model using data up to and including time $t-1$. *Model/Null* identifies the candidate and null models. *RSVAR, OOS splits* uses the regime specifications from the OOS RSVAR. *RSVAR, IS splits* is the regime classification using smoothed probabilities. *NBER splits* is the regime classification by the NBER for the US.

Panel A: US 1968–2007							
		RSVAR, *OOS* splits		RSVAR, *IS* splits		NBER splits	
Model/Null	1968:04–2007:12	Exp	Rec	Exp	Rec	Exp	Rec
RSVAR/HA	3.13	0.47	6.62	−12.35	11.19	−4.73	21.08
VAR/HA	2.68	−3.01	10.18	−16.78	12.83	−5.40	21.15
RSVAR/VAR	0.46	3.38	−3.96	3.80	−1.87	0.64	−0.08
Panel B: US 1980–2007							
US		RSVAR, *OOS* splits		RSVAR, *IS* splits		NBER splits	
Model/Null	1980:01–2007:12	Exp	Rec	Exp	Rec	Exp	Rec
RSVAR/HA	1.93	0.71	3.67	−7.38	8.00	−3.51	21.52
VAR/HA	1.31	−2.75	7.11	−10.36	−8.91	−4.49	22.25
RSVAR/VAR	0.63	3.37	−3.70	2.70	−1.00	0.93	−0.93
Panel C: UK 1980–2007							
UK		RSVAR, *OOS* splits		RSVAR, *IS* splits			
Model/Null	1980:1–2007:12	Exp	Rec	Exp	Rec		
RSVAR/HA	−0.74	−0.91	0.76	−1.96	2.50		
VAR/HA	−1.80	−1.92	−0.73	−4.33	2.56		
RSVAR/VAR	1.04	0.99	1.48	−2.27	4.93		

compared to in-sample RSVAR which shows 77% agreement in state designations with the NBER dates. In other words, with the NBER dates as a benchmark, the real-time model makes errors 5% more frequently than the in-sample RSVAR. We further find that the state uncertainty accounts for a considerable loss in the performance of the RSVAR.

Table 6 Panel A reports R^2_{OOS} for the US from 1968:04–2007:12 for the whole sample and under these splits. A positive *RSVAR/HA* number means that the mean squared prediction errors were smaller for the RSVAR than for the HA. Also shown are the results for *RSVAR/VAR* and for *VAR/HA*. For the total available sample, the RSVAR outperforms both the VAR and HA. In the splits, we see that it outperforms each model for different reasons. In expansions, the RSVAR is slightly better than HA in the OOS-based split, but underperforms in the other splits. However, the RSVAR is clearly better than the VAR in expansions. That the RSVAR underperforms the HA in expansions is consistent with the overall message of this paper, that the information contained in predictors appears useless during expansions—restricting the RSVAR coefficients in this state to zero would appear to improve performance of the RSVAR. On the contrary, the RSVAR also appears to underperform the VAR in recessions, possibly due to small-sample estimation errors for the RSVAR. In recessions, either predictability model, the VAR or the RSVAR, would be preferable to the historical average.

Welch and Goyal (2008, p. 1456) argue (emphasis original), "For many models, any earlier apparent statistical significance was often based exclusively on years up to *and especially on* the years of the Oil Shock of 1973–1975." Accordingly, we report the R^2_{OOS} performance from 1980–2007 for the US (Panel B) and UK (Panel C). We include the UK since the UK data begin in 1965 and 15 years of data first become available in 1980. Both panels demonstrate the same general patterns as Panel A. That the RSVAR underperforms the HA in expansions is consistent with the notion that the information contained in predictors appears useless during expansions; restricting the RSVAR coefficients in the expansion state to zero would improve performance of the RSVAR. In recessions, either predictability model, the VAR or the RSVAR, would be preferable to the historical average.

In international tests the results are mixed, with R^2_{OOS} ranging from −1.86 (France) to −8.14 (Canada). Several factors seem responsible. First, the estimation requirements leave only a short OOS sample of 1993–2007, a period when recessions are rare. Even in the US, this period largely favors the HA model ($R^2_{OOS} = -4.74$). Second, construction of the dividend yield variable has great influence on the conclusions drawn, consistent with Boudoukh, Michaely, Richardson, and Roberts (2007). Using the standard dividend yield excluding repurchases for the US, we find an R^2_{OOS} very close to zero for the entire

1968–2007 period, albeit still positive in recessions. Finally, the default spread appears to be a key indicator of the state of the economy. Its exclusion from the international sample might add considerable uncertainty regarding the regime out-of-sample.

6. Conclusion

We find a strong and robust link between the extent of aggregate return predictability and the business cycle in all of the G7 countries except Germany. Predictors such as the dividend yield and term structure variables are typically important only during poor economic times. We use an RSVAR to separate regimes and find that these regimes are representative of business cycles determined by NBER and ECRI (Economic Cycle Research Institute), as well as by professional survey forecasts. The link is further confirmed in a recent working paper by Rapach, Strauss, and Zhou (2010) finding independent evidence of countercyclical predictability in several equity markets.

Overall, the results are broadly consistent with the prevailing literature on countercyclical risk premiums. We find higher risk premiums in recession regimes as well as higher, potentially priced volatility in returns and economic fundamentals. Although current theoretical models that produce counter-cyclical risk premiums can imply long-horizon return predictability, it may be a greater challenge to incorporate recession-centric short-horizon predictability of the type that we show. Investigations along these lines appear in Menzly, Santos, and Veronesi (2004) and Bekaert and Engstrom (2009).

A key piece of the puzzle may lie in the complex dynamics of the predictors over the business cycle. These dynamics could be responsible for our finding of a complete absence of return predictability during expansions. Notably, most predictors exhibit smoothed properties in expansions, with higher persistence and lower volatility than in recessions. Corroborative evidence consistent with this link is considered in more detail in Chen, Da, and Priestley (2008).

The out-of-sample experiments further underscore our main findings. While both the RSVAR and a simple VAR clearly outperform the historical average in recessions, the historical average is the best performer in expansions. A full exploration of out-of-sample results would range beyond the scope of this paper, but we expect that improved OOS results might be achieved by imposing economically sensible restrictions on the RSVAR coefficients, by tightening the priors on the conditional expected returns, and by incorporating real-time external information such as the Survey of Professional Forecasters, which in the US appears to segment predictability regimes even better than the NBER. Another important avenue of future inquiry is determination of the implications of recession-centric return predictability for asset allocation problems.

Appendix A. MCMC analysis

Bayesian estimation requires three elements: the data, a likelihood function dictated by the model, and prior densities for the model's parameters. Let D denote the data, $p(D|\theta)$ denote the likelihood function, and $\pi(\theta)$ denote the prior density for the parameter set θ. The object of interest is the joint posterior density of the parameters given the data, $\pi(\theta|D)$. Following Bayes rule, this joint posterior density is proportional to the product of the likelihood function and the prior density on the parameters:

$$\pi(\theta|D) \propto p(D|\theta)\pi(\theta).$$

Bayesian inference is accomplished by analyzing the joint posterior density of the model's parameters, or other functions of interest.

A.1. Likelihood function

For notational convenience, let $\mathbf{y} = \{y_1, \ldots, y_T\}$ denote the dependent variables of the k-dimensional RSVAR system. Let θ denote the vector of model parameters. The likelihood function (i.e., sampling density) for the RSVAR model in Eq. (1) is

$$p(\mathbf{y};\theta) = \prod_{t=1}^{T}\sum_{m=1}^{M} f(y_t|\mathcal{F}_{t-1}, s_t = m;\theta)p(s_t = m|\mathcal{F}_{t-1};\theta),$$

where $\mathcal{F}_{t-1}$ denotes information available at time $t-1$ and $f(\cdot)$ is the k-variate Normal density. In our applications, we set M=2 so that m={1,2}.

A.2. Priors

Prior densities summarize the researcher's prior beliefs about the model's parameters. It is possible to model prior beliefs on the parameters by adopting virtually any reasonable distributional form. The choice is by nature subjective. In our empirical applications we aim to minimize the impact of the priors on the estimates. At the same time, we want to impose some minimal economic plausibility. Hence, we impose stationarity on the system within regimes and we put a prior on the market risk premium, as described in the main text. Denote with A_m the companion matrix $A(s_t = m)$ where m={1,2}, with B_m the matrix containing the vector of constants in the first column and the matrix A_m, in the remaining columns, and with b_m the vector stacking the columns of B_m. We further denote Σ_m as the state-dependent error variance–covariance matrix. We assume that the parameters are mutually independent within and across regimes and, thus, can be factored as

$$\pi(\theta) = \pi(b_1)\pi(b_2)\pi(\Sigma_1)\pi(\Sigma_2)\pi(vec(P)).$$

We choose

$$\pi(b_m) = \mathcal{N}_{k(k+1)}(\gamma_0, \Gamma_0^{-1}),$$
$$\pi(\Sigma_m^{-1}) = \mathcal{W}(S_0, s_0),$$
$$\pi(p_{m,m}) = \mathcal{B}(p_0, \nu_0)$$

for m={1,2}, where $\mathcal{N}_k$ denotes a k-variate Normal density, $\mathcal{W}$ denotes the Wishart density, and $\mathcal{B}$ denotes the Beta density. These are fairly standard choices in the analysis of vector autoregression models.

As for the prior parameters, we set

$$\gamma_0 = 0, \quad \Gamma_0 = I_{k(k+1)} \times 10^{-3},$$
$$s_0 = 8, \quad S_0 = I_k \times 10^{-9},$$
$$p_0 = 2.7, \quad v_0 = 0.3.$$

The hyperparameters on the diagonal of the transition probability matrix $P_{M\times M}$ imply a prior mean of 0.9 (equivalent to an expected duration of the regime of 10 months) and a prior standard deviation of 0.15. Although fairly diffuse, they reflect the expected persistence in economic regimes. The priors about the b's and the Σ's reflect diffused prior beliefs. It follows that the posterior densities will be largely determined by the sample data. As an aside, when the priors are completely uninformative (improper), the posterior means from the Gibbs sampler will equal the maximum likelihood point estimates. On the other hand, besides having little economic plausibility in several instances, a completely diffused prior would also raise a technical challenge as it would make the marginal likelihood and, thus, the posterior-odds ratios undefined.

A.3. Prior–posterior analysis

Combining the likelihood function and the priors via Bayes rule, one obtains the joint posterior density of the model's parameters given the data. Given the form of the likelihood functions in Appendix A.1 and given the priors described above, the joint posterior cannot be estimated (i.e., sampled) directly. Fortunately, the Gibbs sampling method bypasses the computation of the likelihood function and computation of the joint posterior density. Rather, the Gibbs sampler algorithm generates draws from the conditional distribution of each block of parameters (i.e., the distribution of each block given the data, the prior, and the other blocks of parameters). The draws from these conditional densities eventually converge to draws from the joint posterior density. Inference is based on the moments describing the distribution of the sample draws of the model's parameters, and of functions thereof.

In our applications, the Gibbs sampler consists of four blocks. Namely, the sampling proceeds with the following structure:

1. Initialize A_m, Σ_m, and $p_{m,m}$, m={1,2}.
2. Draw the latent states s_t, {t=1, …, T}, using the multi-move filtering/smoothing algorithm first proposed by Carter and Kohn (1994).
3. Draw the free elements of the transition matrix, P, from a beta variate.
4. Conditioning on the latent states
   (a) Draw the elements of the matrix B_m, m={1, 2} from a multivariate normal variate. Discard the draws that violate stationarity of the VAR.
   (b) Draw the elements of the matrix Σ_m, m={1,2} from an Inverse Wishart variate. Discard the draws that violate the identification condition $\Sigma_1[1,1] < \Sigma_2[1,1]$, where $\Sigma_m[1,1]$ denotes the conditional market variance in regime m.
5. Go to step 2 and repeat.

After discarding the first 5,000 iterations, we collect the subsequent 50,000 draws. Our implementation of the Bayesian MCMC approach follows closely the scheme proposed by Krolzig (1998, Chapters 8 and 9). As the steps are now standard in Bayesian analysis, we refer the reader to that book's treatment for details.

A.4. Marginal likelihood calculation

We follow the approach suggested by Chib (1995) for the computation of marginal likelihood estimates. The essential steps of the method are as follows. Using Bayes rule, the marginal likelihood of the data given a model specification, $\mathcal{M}$, is

$$m(y|\mathcal{M}) = \frac{p(y|\mathcal{M},\theta^*)p(\theta^*|\mathcal{M})}{p(\theta^*|y,\mathcal{M})}. \tag{7}$$

Starting from this basic marginal likelihood identity (see Chib, 1995), the log of the Bayes factor for comparing (potentially, non-nested) models $\mathcal{M}_f$ to $\mathcal{M}_g$ is written as

$$\begin{aligned}\log p(y|\mathcal{M}_f) - \log p(y|\mathcal{M}_g) \\ = \{\log p(y|\mathcal{M}_f,\theta_f^*) + \log p(\theta_f^*|\mathcal{M}_f) - \log \pi(\theta_f^*|y,\mathcal{M}_f)\} \\ -\{\log p(y|\mathcal{M}_g,\theta_g^*) + \log p(\theta_g^*|\mathcal{M}_g) - \log \pi(\theta_g^*|y,\mathcal{M}_g)\},\end{aligned} \tag{8}$$

where $p(y|\mathcal{M}_h,\theta_h^*)$ is the likelihood function under $\mathcal{M}_h$, h={f,g}. $p(\theta_h^*|\mathcal{M}_h)$ and $\pi(\theta_h^*|y,\mathcal{M}_h)$ are the corresponding prior and posterior densities, each evaluated at the point θ_h^*. The choice of the points θ_f^* and θ_g^* is arbitrary but for computational accuracy we take these to be the posterior means estimated from the MCMC procedure. Suppressing the model subscript for convenience, there are two key quantities that must be computed for each model under study: the posterior ordinate $\pi(\theta^*|y,\mathcal{M})$ and the likelihood ordinate $p(y|\mathcal{M},\theta^*)$. The latter is computed using the filtering recursions of Carter and Kohn (1994). Because of the high dimensionality of the vector θ and the presence of the latent states, calculation of $\pi(\theta^*|y,\mathcal{M})$ is not analytically tractable. Fortunately, Chib (1995) and Chib and Jeliazkov (2001) provide simulation-based strategies to efficiently estimate the posterior ordinate. Complete details of the posterior ordinate estimation via reduced blocking are available from the authors upon request.

Following, among others, Kass and Raftery (1995), we evaluate the significance of a Bayes factor based on the following scale:

$BF_{f,g}$	Evidence against $\mathcal{M}_g$
1 to 3.16	Not worth more than a bare mention
3.16 to 10	Substantial
10 to 100	Strong
>100	Decisive

References

Ang, A., Bekaert, G., 2002a. International asset allocation with regime shifts. Review of Financial Studies 15 (4), 1137–1187.

Ang, A., Bekaert, G., 2002b. Regime switches in interest rates. Journal of Business and Economic Statistics 20 (2), 163–182.

Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there? Review of Financial Studies 3, 651–707.
Ang, A., Bekaert, G., Wei, M., 2007. Do macro variables, asset markets, or surveys forecast inflation better? Journal of Monetary Economics 54 (4), 1163–1212.
Ang, A., Bekaert, G., Wei, M., 2008. The term structure of real rates and expected inflation. Journal of Finance 63 (2), 797–849.
Avramov, D., 2002. Stock return predictability and model uncertainty. Journal of Financial Economics 64 (3), 423–458.
Avramov, D., Chao, J.C., 2006. An exact Bayes test of asset pricing models with application to international markets. Journal of Business 79 (1), 293–324.
Avramov, D., Chordia, T., 2006. Predicting stock returns. Journal of Financial Economics 82 (2), 387–415.
Bansal, R., Dittmar, R.F., Lundblad, C., 2005. Consumption, dividends, and the cross-section of equity returns. Journal of Finance 60, 1639–1672.
Bansal, R., Tauchen, G., Zhou, H., 2004. Regime shifts, risk premiums in the term structure, and the business cycle. Journal of Business and Economic Statistics 22, 396–409.
Bansal, R., Zhou, H., 2002. The term structure of interest rates with regime shifts. Journal of Finance 57, 1997–2043.
Basu, S., 1997. The conservatism principle and the asymmetric timeliness of earnings. Journal of Accounting and Economics 24 (1), 3–37.
Bekaert, G., Engstrom, E., 2009. Asset return dynamics under bad environment good environment fundamentals. NBER Working Paper No. 15222.
Bekaert, G., Engstrom, E., Xing, Y., 2009. Risk, uncertainty, and asset prices. Journal of Financial Economics 91 (1), 59–82.
Bliss, J.H., 1924. Management Through Accounts. The Ronald Press Co, New York.
Bossaerts, P., Hillion, P., 1999. Implementing statistical criteria to select return forecasting models: what do we learn? Review of Financial Studies 12 (2), 405–428.
Boudoukh, J., Michaely, R., Richardson, M., Roberts, M.R., 2007. On the importance of measuring payout yield: implications for empirical asset pricing. Journal of Finance 62 (2), 877–915.
Boudoukh, J., Richardson, M., Smith, T., Whitelaw, R.F., 1999. Regime shifts and bond returns. Working Paper No. S-MF-99-12, NYU.
Boudoukh, J., Richardson, M., Whitelaw, R.F., 2008. The myth of long-horizon predictability. Review of Financial Studies 21 (4), 1577–1605.
Brav, A., Graham, J.R., Harvey, C., Michaely, R., 2005. Payout policy in the 21st century. Journal of Financial Economics 77, 483–527.
Campbell, J.Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.
Campbell, J.Y., 1996. Understanding risk and return. Journal of Political Economy 104 (2), 298–345.
Campbell, J.Y., 2001. Why long horizons? A study of power against persistent alternatives. Journal of Empirical Finance 8 (5), 459–491.
Campbell, J.Y., Cochrane, J., 1999. Force of habit: a consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107 (2), 205–251.
Campbell, J.Y., Lo, A., MacKinlay, C., 1997. The Econometrics of Financial Markets. Princeton University Press, NJ.
Campbell, J.Y., Shiller, R.J., 1988. The dividend–price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–228.
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 (4), 1509–1531.
Carter, C.K., Kohn, R.J., 1994. On Gibbs sampling for state space models. Biometrika 81 (3), 541–553.
Chan, K.C., Karolyi, G.A., Longstaff, F.A., Sanders, A.B., 1992. An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 47 (3), 1209–1228.
Chen, L., Da, Z., Priestley, R., 2008. Dividend smoothing and predictability. Unpublished Working Paper, University of Washington-St. Louis, University of Notre Dame, and Norwegian School of Management.
Chen, N.-F., Roll, R., Ross, S.A., 1986. Economic forces and the stock market. Journal of Business 59 (3), 383–403.
Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321.
Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis–Hastings output. Journal of the American Statistical Association 96, 270–281.
Chordia, T., Shivakumar, L., 2002. Momentum, business cycle, and time-varying expected returns. Journal of Finance 57 (2), 985–1019.
Cochrane, J.H., 2001. Asset Pricing. Princeton University Press, Princeton, NJ.
Cochrane, J.H., 2008. The dog that did not bark: a defense of return predictability. Review of Financial Studies 21 (4), 1533–1575.
Cremers, K.J.M., 2002. Stock return predictability: a Bayesian model selection perspective. Review of Financial Studies 15 (4), 1223–1249.
Davig, T., Leeper, E.M., 2007. Generalizing the Taylor principle. American Economic Review 97 (3), 607–635.
Dow, C.H., 1920. Scientific Stock Speculation. The Magazine of Wall Street, New York.
Fama, E.F., 1965. The behavior of stock-market prices. Journal of Business 38 (1), 34–105.
Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. Journal of Finance 25 (2), 383–417.
Fama, E.F., 1981. Stock returns, real activity, inflation, and money. American Economic Review 71, 545–565.
Fama, E.F., 1984. The information in the term structure. Journal of Financial Economics 13, 509–528.
Fama, E.F., Bliss, R.R., 1987. The information in long-maturity forward rates. American Economic Review 77, 680–692.
Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25 (1), 23–49.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3–56.
Fama, E.F., Schwert, G.W., 1977. Asset returns and inflation. Journal of Financial Economics 5, 115–146.
Ferson, W.E., Harvey, C.R., 1991. The variation of economic risk premiums. Journal of Political Economy 99 (2), 385–415.
Fudenberg, D., Tirole, J., 1995. A theory of income and dividend smoothing based on incumbency rents. Journal of Political Economy 103 (1), 75–93.
Geske, R., Roll, R., 1983. The fiscal and monetary linkage between stock returns and inflation. Journal of Finance 38 (1), 1–33.
Goetzmann, W.N., Jorion, P., 1993. Testing the predictive power of dividend yields. Journal of Finance 48 (2), 663–679.
Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. Management Science 49, 639–654.
Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27–62.
Guidolin, M., Timmermann, A.G., 2007. Asset allocation under multivariate regime switching. Journal of Economic Dynamics and Control 31 (11), 3503–3544.
Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.
Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.
Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64 (1–2), 307–333.
Hardouvelis, G.A., 1988. The predictive power of the term structure during recent monetary regimes. Journal of Finance 43 (2), 339–356.
Harvey, C.R., 1988. The real term structure and consumption growth. Journal of Financial Economics 22 (2), 305–333.
Hodrick, R.J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. Review of Financial Studies 5 (3), 357–386.
Kandel, S., Stambaugh, R.F., 1996. On the predictability of stock returns: an asset-allocation perspective. Journal of Finance 53 (2), 385–424.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.
Keim, D.B., Stambaugh, R.F., 1986. Predicting returns in the stock and bond markets. Journal of Financial Economics 17 (2), 357–390.
Kim, C.-J., Nelson, C.R., 1993. Predictable stock returns: the role of small sample bias. Journal of Finance 48, 641–661.
Kim, C.-J., Nelson, C.R., 1999. State-Space Models with Regime-Switching. The MIT Press, Cambridge, MA.
Krolzig, H.M., 1998. Markov-Switching Vector Autoregressions. Springer, New York.
Kumar, P., 1988. Shareholder-manager conflict and the information content of dividends. Review of Financial Studies 1 (2), 111–136.
Lettau, M., Van Nieuwerburgh, S., 2008. Reconciling the return predictability evidence. Review of Financial Studies 21 (4), 1607–1652.
Li, G., 2007. Time-varying risk aversion and asset prices. Journal of Banking and Finance 31 (1), 243–257.
Lintner, J., 1956. Distribution of incomes of corporations among dividends, retained earnings, and taxes. American Economic Review 46 (2), 97–113.

Mankiw, N.G., Miron, J.A., 1986. The changing behavior of the term structure of interest rates. Quarterly Journal of Economics 101 (2), 211–228.
Marsh, T.A., Merton, R.C., 1987. Dividend behavior for the aggregate stock market. Journal of Business 60 (1), 1–40.
Mayfield, E.S., 2004. Estimating the market risk premium. Journal of Financial Economics 73 (3), 465–496.
Mele, A., 2007. Asymmetric stock market volatility and the cyclical behavior of expected returns. Journal of Financial Economics 86 (2), 446–478.
Menzly, L., Santos, T., Veronesi, P., 2004. Understanding predictability. Journal of Political Economy 112, 1–47.
Nardari, F., Scruggs, J., 2007. Bayesian analysis of linear factor models with latent factors, multivariate stochastic volatility, and APT pricing restrictions. Journal of Financial and Quantitative Analysis 42 (4), 857–892.
O'Hagan, A. 1994. Bayesian Inference. Kendall's Advanced Theory of Statistics, vol. 2B. Edward Arnold, London.
Pástor, Ľ., Stambaugh, R.F., 2009. Predictive systems: living with imperfect predictors. Journal of Finance 64 (4), 1583–1628.
Paye, B.S., Timmermann, A., 2006. Instability of return prediction models. Journal of Empirical Finance 13 (3), 274–315.
Perez-Quiros, G., Timmermann, A.G., 2000. Firm size and cyclical variations in stock returns. Journal of Finance 55 (2), 1229–1262.
Pesaran, M.H., Timmermann, A.G., 2002. Market timing and return prediction under model instability. Journal of Empirical Finance 9 (5), 495–510.
Rapach, D., Strauss, J., Zhou, G., 2010. International stock return predictability: What is the role of the United States? Unpublished Working Paper, St. Louis University and Washington University-St. Louis.
Richardson, M., Stock, J.H., 1989. Drawing inferences from statistics based on multiyear asset returns. Journal of Financial Economics 25 (2), 323–348.
Rozeff, M., 1984. Dividend yields are equity risk premiums. Journal of Portfolio Management 11, 68–75.
Shiller, R.J., 1981. Do stock prices move too much to be justified by subsequent changes in dividends? American Economic Review 71 (3), 421–436.
Sterling, R.R., 1970. The Theory of the Measurement of Enterprise Income. University of Kansas Press, Lawrence, KS.
Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications. Journal of Financial Economics 68, 201–232.
Watts, R.L., 2003. Conservatism in accounting part I: explanations and implications. Accounting Horizons 17 (3), 207–221.
Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21 (4), 1455–1508.