Are the Dynamic Linkages Between the Macroeconomy and Asset Prices Time-Varying?*

Massimo Guidolin†
Federal Reserve Bank of St. Louis - Research Division

Sadayuki Ono
University of York, UK

Abstract

We estimate a number of multivariate regime switching VAR models on a long monthly U.S. data set for eight variables that include excess stock and bond returns, the real T-bill yield, predictors used in the finance literature (default spread and the dividend yield), and three macroeconomic variables (inflation, industrial production growth, and a measure of real money growth). Heteroskedasticity may be accounted for by making the covariance matrix a function of the regime. We find evidence of four regimes and of time-varying covariances. We show that the best in-sample fit is provided by a four state model in which the VAR(1) component fails to be regime-dependent. We interpret this as evidence that the dynamic linkages between financial markets and the macroeconomy have been stable over time. The four-state model can be helpful in forecasting applications and provides one-step ahead predicted Sharpe ratios.

JEL codes: E44, G12, C32, C52.
Keywords: Predictability, Multivariate Regime Switching, Predictive Density Tests, Sharpe ratios.

1. **Introduction**

The possibility that macroeconomic aggregates may predict the evolution of asset prices has been attracting the attention of researchers in economics and finance since the late 1970s. Against the background of the efficient market hypothesis (EMH, by which asset prices should follow a random walk or at least be unpredictable, given current information), the existence of statistically detectable predictability patterns has been considered interesting not only for its intrinsic usefulness in asset pricing and portfolio management, but also because a reconciliation between the EMH and the predictive power of macroeconomic variables was perceived as a high-priority research question. Therefore a remarkable bulk of empirical evidence on such predictability relationships linking asset returns and macroeconomic factors has been cumulating.[1]

Another, recent literature has investigated whether asset returns may forecast future realized macroeconomic variables, particularly output and inflation (see e.g. Stock and Watson (2003)). Ultimately, these

*Elizabeth La Jeunesse provided excellent research assistance. We thank two anonymous referees for stimulating comments and helpful suggestions. We are also indebted to session partecipants at the European Financial Management Association in Madrid (July 2005).

†Correspondence to: Research Division, P.O. Box 442, St. Louis, MO 63166, United States. E-mail: Massimo.Guidolin@stls.frb.org; phone: 314-444-8550.

[1] It is now clear that the EMH may be consistent with predictability. In synthesis, the random walk obtains only under special assumptions or after appropriately scaling the asset prices. The EMH implies the existence of a relationship between asset returns and all variables that contain information on the fundamental pricing operator (the stochastic discount factor).

analyses also rely on a rationality restriction: since in equilibrium financial markets should efficiently aggregate into asset prices expectations that relate to future economic conditions (output and relative prices), if the forecasts of market participants are rational one would expect that asset returns could on average predict future macroeconomic conditions (see e.g. Fischer and Merton, 1984). Additionally, financial markets routinely provide data in such quantities that it seems also very convenient for professional macroeconomic forecasters and policy makers to investigate whether any useful information may be extracted from asset prices.

Our paper deals with both aspects of the linkages between financial returns and macroeconomic variables and asks whether it is sensible – as it has been routinely done so far – to assume that such dynamic predictability relationships (if any) have been stable over time in the US. In fact, the US economy (as well as the world economy) has been changing at such a fast pace that attaching a lot of weight to a prior that such dynamic linkages would have remained stable and unchanging requires some careful scrutiny. In particular, the acceleration of the speed of change over the last 30 years, after the two oil shocks, a few experiments concerning the conduct of the monetary policy, and the removal of most of the remaining barriers to international flows of goods, services, and capital, may instead justify putting a lot of trust in the opposite prior that the recent experience might be somehow different. Additionally, stock (with the tech bubble of the 1990s) and bond (with the protracted worldwide decline in long-term interest rates since the early 1980s) markets have been subject to dramatic changes that may lend support to the hypothesis of unstable dynamic linkages.

We investigate the hypothesis of time-varying dynamic linkages across financial markets and the macroeconomy in a highly flexible multivariate regime switching framework in which the existence of breaks and of structural change is captured in the form of a vector autoregressive (VAR) system alternating among a number of recurrent regimes with different statistical properties. We stress that the model is flexible but also required by the question we investigate. It is flexible as it does not impose the presence of regimes but uses a number of econometric tools to test whether multiple regimes are needed in order to fit and/or forecast the data. Moreover, flexibility exists in the way regimes enter the model: in particular – when predictability patterns take the form of a vector autoregressive component by which past values of some variables may influence the conditional predictive mean of other variables – it is possible that regimes may affect parameters (components) of the model that do not directly affect predictability. On the other hand, using models in such class appears to be a minimal requirement: it is well known that in the presence of non-stationarities, standard linear regression models deliver biased estimates.[2] Only by endogenizing the presence of regimes, we can learn in statistically meaningful ways about the issue of predictability.

Our paper offers three contributions. First, it estimates a relatively wide range of multivariate $k-$regime VAR models in which heteroskedasticity may be accounted for by making the covariance matrix a function of the regime. The model is applied to an eight-variable vector that includes stock and bond returns in excess of a T-bill rate, the T-bill yield, typical predictors used in the finance literature (such as the default spread between low- and high-grade bond yields and the dividend yield), and three genuine macroeconomic variables, inflation, industrial production growth, and a measure of real money growth. We use the longest available monthly data base for US data, 1926:12 - 2004:12. We find evidence of four regimes and of time-varying covariances. The four regimes carry a sensible interpretation as a moderately persistent bull-rebound state, a highly persistent stable state, an expansion, high growth state, and a recession-bear state. The last two

[2]To consider a simple example, think of what happens to the estimate of the slope coefficient in a regression model in which the intercept stochastically switches between two alternative values: unless the switching is taken into account, the estimate of the slope coefficient will be biased and inconsistent as its estimate slope will be lower than the true but unknown one.

regimes have low persistence and hence durations limited to 4-5 months.

Second, we provide evidence that the best fit to the joint density of the data is provided by a four state model in which the VAR(1) component fails to be regime-dependent. We interpret this as evidence that the dynamic linkages between financial markets and the macroeconomy have been stable over time, which counters a prior of evolving predictability patterns. To our knowledge such evidence of a stable dynamic relationship between financial markets and the US macroeconomy is new.

Third, we document two of the possible uses one could make of our estimation results. We show that the four-state model can be helpful in forecasting applications, in the sense that for many relevant variables (especially equity-related ones, stock returns and dividend yields) its recursive, out-of-sample predictive performance turns out to be superior to a simpler (and nested) VAR(1), as well as competing regime switching models (e.g. in terms of number of states) and simple benchmarks that earlier papers have proven hard to beat in related applications. Interestingly, such appealing forecasting performance is statistically significant when formal tests of superior predictive accuracy are applied, especially at intermediate and long horizons.[3] Additionally, we provide evidence that the one-step ahead predicted Sharpe ratios for both stocks and bonds are much more sensible when evaluated under the four-state model than under a VAR(1). We argue that this difference may be crucial in financial applications, such as portfolio management.

The applications of switching models in macroeconomics and finance are constantly expanding. Following Hamilton (1989), several papers have proposed to improve the empirical fit of standard, single-equation models for short-term interest rates (e.g. Gray (1996) and Ang and Bekaert (2002b)) and stock returns (e.g. Turner, Starz, and Nelson (1989) and Ang and Bekaert (2002a)) by allowing for mixtures of distributions. For instance, Turner, Starz, and Nelson (1989) develop a univariate model with regime shifts in means and variances, showing that mean excess equity returns tend to be low in the high-risk (volatility) period, and vice versa. Allowing for switching in the parameters of an autoregressive conditional heteroskedasticity (ARCH) process, Hamilton and Susmel (1994) report that in-sample performance and out-of-sample forecasts of the regime-switching ARCH are superior to a benchmark single-state GARCH(1,1) specification and that the high-volatility state is likely to occur in recession periods. Guidolin and Timmermann (2006) extend this class of models to multivariate systems including excess returns on a few equity portfolios as well as bond returns. However, to our knowledge ours is the first paper to undertake a thorough investigation of predictability patterns involving stock and bond returns, along with a rich set of macroeconomic variables.

A literature exists that stresses that the forecasting power of financial variables for key macroeconomic aggregates is strongly time-varying. For instance, Stock and Watson (2003) report and discuss a bulk of evidence that shows that the US term structure fails to steadily predict output growth. Davis and Fagan (1997) document similar instability involving the out-of-sample forecasting performance of yield spreads for nine European countries. Emery (1996) makes a similar point with reference to the instability of predictive relations involving the spread between commercial paper (high yield corporate notes) and T-bill yields. However Estrella et al. (2003) have concluded that when there is strong international evidence of forecasting power of the yield spreads for real activity, then the predictive relations also tend to be stable over time. Only a few papers – e.g. Jaditz et al. (1998) – have explored the possibility that carefully specified nonlinear prediction models may reproduce the possible time-variation that many papers have uncovered in the forecasting relations connecting

[3] This is important because many papers have warned that while in-sample tests generally show that a large number of macroeconomic variables appear to predict future stock returns, out-of-sample tests of return predictability (that protect against data mining) typically return disappointing results, see e.g. Rapach and Wohar (2005).

financial variables to output and inflation. Our paper takes a few steps in this direction.

The paper is structured as follows. Section 2 gives a quick literature review that highlights the goals and limitations of our exercise. Section 3 describes our econometric model. Section 4 gives information on the data employed in the paper. Section 5 estimates a range of regime switching models and proceeds to select the one providing the best fit according to a number of statistical criteria. Parameter estimates and interpretation for the resulting regimes are provided. The basic finding of no time-variation in the dynamic relationships between financial markets and the macroeconomy is presented. Section 6 shows that a four-state model produces somewhat useful out-of-sample forecasts. This validates the possibility that such a model may provide an approximation to the data generating process. Section 7 comments on possible financial applications of our findings, and provides simple examples. Section 8 concludes. Two appendices detail selected technical aspects of the econometric methodology.

2. Literature Review

An impressive amount of literature has cumulated that investigates whether US stock and government bond returns are predictable using past values of a number of macroeconomic variables. In fact, using linear regression models, numerous studies have found that in each data set a few macroeconomic variables can be found that systematically predict US stock returns. Fama and French (1988) document that the dividend yield forecasts future returns on common stocks. Fama and Schwert (1977) report that real stock returns are negatively related with expected and unexpected components of inflation, which implies that stocks are not a good inflation hedge. They also show that industrial production and real GNP growth have forecasting power for financial returns. Cutler, Poterba, and Summers (1989) examine the forecasting power of unexpected changes in a number of macroeconomic variables. They found that a positive shock to the rate of growth of industrial production significantly raises returns on a value-weighted NYSE portfolio. Balvers, Cosimano, and Mcdonald (1990) show that industrial production and real GNP predict stock returns with significantly negative coefficients. Several papers have proposed that money supply is a key variable that determines fluctuations in stock prices. For instance, Homa and Jaffee (1971) found that the money supply growth rate contains predictive power for quarterly stock returns in the period 1954-1961. An increase in the growth rate of money increases one quarter-ahead stock returns. Kaul (1987) shows that a negative relationship between real stock returns and inflation in the post-war data may be caused by a counter-cyclical monetary policy. Hardouvelis (1987) examines stock market reactions to announcements on 15 different macroeconomic variables. He finds that monetary news have a significant effect on stock returns in the October 1979 - October 1982 period, when the Federal Reserve followed non-borrowed reserve targets. This type of finding points to the possibility of regime switching in predictability. Campbell (1987) presents evidence that a variety of term structure variables such as two-month and six-month spreads as well as the 1-month T-bill rate, all forecast excess stock returns. Fama and French (1989) confirm this result using data at alternative frequencies and a longer sample period (1927-1987). Similarly, Fama and French (1989) investigate whether default risk is a significant predictor of stock returns using the yield spread between low- and high-grade corporate bonds. They find that higher spreads predict subsequent increases in stock returns.[4]

The early literature has been generalized in three directions. First, a few papers have tried to extend this

[4] There is also abundant international evidence on linear predictability. For instance, using data for 12 industrialized countries, Rapach, Wohar, and Rangvid (2005) show that interest rates have the most significant forecasting ability in almost every countries, both in-sample and out-of-sample tests. Inflation, money, and term spreads also exhibit some predictive ability.

evidence to bond markets. Campbell (1987) shows that term and default spreads forecast excess returns on long-term bonds. Fama and French (1993) confirm this evidence. Second, a literature has checked the robustness of these results within multivariate, full-information models in which not only macroeconomic factors are allowed to predict future asset returns, but also the opposite may occur. Here results are mixed. Using a vector-autoregression moving average (VARMA) approach, James, Koreisha, and Partch (1985) investigate simultaneous relations among stock returns, real activity, money supply, and inflation. Their findings support the notion that stock returns are important predictors of changes in expected inflation and nominal interest rates.[5] Lee (1992) uses a VAR model, finding that an increase in real stock returns forecasts subsequent increases in real activity as measured by the growth rate of industrial production. However, in contrast to the these findings, Canova and De Nicolo (2000) show that US stock returns do not contain significant forecasting power of real activity and inflation, even in open-economy setups. Using a structural VAR framework characterized by long-run monetary neutrality, Rapach (2001) studies the effects of money supply, aggregate spending, and aggregate supply shocks on real stock returns, finding mixed results.

Third, a number of papers have tried to *informally* generalize the evidence on predictability to models in which regimes play a role. For example, Pesaran and Timmermann (1995) introduce a time-varying *choice* of forecasting variables, in which the selection is based on a number of alternative criteria (adjusted R^2, the Akaike, Bayes-Schwartz, and Hannan-Quinn information criteria). They show that the optimal selection of prediction variables significantly changes over time and that only the 1-month T-bill rate is included over the entire sample. Flannery and Protopapadakis (2002) show that macroeconomic announcements concerning inflation and money supply consistently affect the level of stock returns and that market responses on the announcement are time-varying. Employing VAR methods with endogenous break points, Du (2005) offers evidence that the (contemporaneous) correlation between real stock returns and inflation varies over time. He shows that the time-varying correlation is mainly due to changes in monetary policy regimes.

3. **Econometric Methodology**

Suppose that the random vector collecting monthly (excess) returns on l different assets and q macroeconomic variables, possibly predicting (and predicted by) asset returns, follows a $k-$regime Markov switching (MS) VAR(p) process with heteroskedastic components, compactly $MSIAH(k,p)$ (see Krolzig (1997)):

$$\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_{j,S_t}\mathbf{y}_{t-j} + \boldsymbol{\Sigma}_{S_t}\boldsymbol{\epsilon}_t \tag{1}$$

with $\boldsymbol{\epsilon}_t \sim NID(\mathbf{0}, \mathbf{I}_{l+q})$.[6] S_t is a latent state variable driving all the parameters appearing in (1). $\boldsymbol{\mu}_{S_t}$ collects the $l+q$ regime-dependent intercepts, while the $(l+q)\times(l+q)$ matrix $\boldsymbol{\Sigma}_{S_t}$ represents the factor applicable in a state-dependent Choleski factorization of the covariance matrix, $\boldsymbol{\Omega}_{S_t}$. A non-diagonal $\boldsymbol{\Sigma}_{S_t}$ captures *simultaneous* comovements. Moreover, *dynamic* (lagged) linkages across different asset markets and between financial markets and macroeconomic influences are captured by the VAR(p). In fact, conditionally on the unobservable state S_t, (1) defines a standard Gaussian reduced form VAR(p) model. On the other

[5] Early papers in this literature (e.g. Estrella and Hardouvelis (1991), and Bernanke and Blinder (1992)) had shown that when short-term interest rates or interest rate spreads were included in VARs for output and inflation, they tended to eliminate the marginal predictive content of the money growth rate.

[6] Assume the absence of roots outside the unit circle, thus making the process covariance stationary. Ang and Bekaert (2002b) show that for covariance stationarity to obtain, it is sufficient for such a condition to be verified in at least one of the k regimes.

hand, when $k > 1$, alternative hidden states are possible that will influence both the conditional mean and the volatility/correlation structure characterizing (1). The states are generated by a discrete, homogeneous, irreducible, and ergodic first-order Markov chain:[7]

$$\Pr\left(s_t = j | \{s_j\}_{j=1}^{t-1}, \{\mathbf{y}_j\}_{j=1}^{t-1}\right) = \Pr\left(s_t = j | s_{t-1} = i\right) = p_{ij}, \tag{2}$$

where p_{ij} is the generic $[i, j]$ element of the $k \times k$ transition matrix $\mathbf{P}$. Ergodicity implies the existence of a stationary vector of probabilities $\bar{\boldsymbol{\xi}}$ satisfying $\bar{\boldsymbol{\xi}} = \mathbf{P}'\bar{\boldsymbol{\xi}}$. Irreducibility implies that $\bar{\boldsymbol{\xi}} > \mathbf{0}$, meaning that all unobservable states are possible. In practice, $\mathbf{P}$ is unknown and hence $\bar{\boldsymbol{\xi}}$ can be at most estimated given knowledge on $\mathbf{P}$ extracted from the information set $\Im_T = \{\mathbf{y}_j\}_{j=1}^{T}$.

When l and/or q are large, model (1) implies the estimation of a large number of parameters, $k[(l + q) + p(l + q)^2 + (l + q)(l + q + 1)/2 + (k - 1)]$. For instance, for $k = 2$, $l = 3$, $p = 1$, and $q = 5$ (some of the hyper-parameters in our application), this implies estimation of $2 \times [8 + 8^2 + 4 \times 9 + 1] = 218$ parameters. (1) nests a number of simpler models in which either some parameter matrices are not needed or some of the objects become regime-independent. These simpler models may greatly reduce the number of parameters to be estimated. Among them, we will devote special attention to $MSIH(k, p)$ models,

$$\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \boldsymbol{\Sigma}_{S_t}\boldsymbol{\epsilon}_t,$$

in which $p = 0$, $MSIA(k, p)$ homoskedastic models,

$$\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_{j,S_t}\mathbf{y}_{t-j} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t,$$

in which the covariance matrix is constant over time, and $MSIH(k, 0)$-$VAR(p)$ models,

$$\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_j\mathbf{y}_{t-j} + \boldsymbol{\Sigma}_{S_t}\boldsymbol{\epsilon}_t, \tag{3}$$

a special case of (1) in which intercepts the and covariance matrix are regime-dependent, while the VAR(p) coefficients are not. Model (3) implies the estimation of $k[(l + q) + (l + q)(l + q + 1)/2 + (k - 1)] + p(l + q)^2$ parameters. For the same choices above, this means $2 \times [8 + 4 \times 9 + 1] + 8^2 = 154 < 218$. As we will see, this restricted sub-class of models turns out to be important to test the hull hypothesis that predictability patterns involving asset returns and macroeconomic variables are time-varying. A limit case of (1) is obtained when $k = 1$, a standard multivariate Gaussian VAR(p) model:

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} \mathbf{A}_j\mathbf{y}_{t-j} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t. \tag{4}$$

MSIAH models are estimated by maximum likelihood. An appendix in Guidolin and Ono (2005) shows that algorithms are considerably simplified if (1) is first put in its state-space form. In particular, estimation and inferences are based on the EM (Expectation-Maximization) algorithm proposed by Dempster et al. (1977) and Hamilton (1989), a filter that allows the iterative calculation of the one-step ahead forecast of the state vector $\boldsymbol{\xi}_{t+1|t}$ given the information set $\Im_t$ and the consequent construction of the log-likelihood function

[7] The assumption of a first-order Markov process is not restrictive, since a higher order Markov chain can always be reparameterized as a higher dimensional first-order Markov chain.

of the data. Guidolin and Ono (2005) give a few additional details on the EM algorithm. As for the properties of the resulting ML estimators, under standard regularity conditions (such as identifiability, stability and the fact that the true parameter vector does not fall on the boundaries) Hamilton (1989, 1993) and Leroux (1992) have proven consistency and asymptotic normality of the ML estimator $\tilde{\boldsymbol{\gamma}}$:

$$\sqrt{T}\left(\tilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)\xrightarrow{d} N\left(\mathbf{0},\mathcal{I}_a(\boldsymbol{\gamma})^{-1}\right)$$

where $\mathcal{I}_a(\boldsymbol{\gamma})$ is the asymptotic information matrix. In our empirical results we are going to provide standard results based on a 'sandwich' sample estimator of $\mathcal{I}_a(\boldsymbol{\gamma})$ by which:[8]

$$\widetilde{Var}(\tilde{\boldsymbol{\gamma}})=T^{-1}\left[\mathcal{I}_2(\tilde{\boldsymbol{\gamma}})\left(\mathcal{I}_1(\tilde{\boldsymbol{\gamma}})\right)^{-1}\mathcal{I}_2(\tilde{\boldsymbol{\gamma}})\right],$$

where

$$\mathcal{I}_1(\tilde{\boldsymbol{\gamma}})\equiv T^{-1}\sum_{t=1}^{T}\left[\mathbf{h}_t(\tilde{\boldsymbol{\gamma}})\right]\left[\mathbf{h}_t(\tilde{\boldsymbol{\gamma}})\right]' \qquad \mathbf{h}_t(\tilde{\boldsymbol{\gamma}})\equiv\frac{\partial\ln p(\mathbf{y}_t|\Im_{t-1};\tilde{\boldsymbol{\gamma}})}{\partial\boldsymbol{\gamma}} \qquad \mathcal{I}_2(\tilde{\boldsymbol{\gamma}})\equiv -T^{-1}\sum_{t=1}^{T}\left[\frac{\partial^2\ln p(\mathbf{y}_t|\Im_{t-1};\tilde{\boldsymbol{\gamma}})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right],$$

($p(\mathbf{y}_t|\Im_{t-1};\tilde{\boldsymbol{\gamma}})$ is the conditional density of the data). With one important exception, standard inferential procedures are available to test statistical hypothesis.[9] The Appendix details how Likelihood Ratio (LR) tests are implemented in our paper.

Under a mean squared forecast error (MSFE) criterion, the forecasting algorithms are simple in spite of the nonlinearity of these processes. Considering a MSIAH(k,p) process, the function minimizing the MSFE is the standard conditional expectation function. For instance, for a one-step ahead forecast:

$$E[\mathbf{y}_{t+1}|\Im_t]=\mathbf{X}_{t+1}\hat{\Psi}\left(\hat{\boldsymbol{\xi}}_{t+1|t}\otimes\boldsymbol{\iota}_{l+q}\right)$$

where $\mathbf{X}_{t+1}=[1\ \mathbf{y}_t'...\mathbf{y}_{t-p+1}']\otimes\boldsymbol{\iota}_{l+q}$, $\hat{\Psi}$ collects the estimated conditional mean parameters, and $\hat{\boldsymbol{\xi}}_{t+1|t}$ is the one-step ahead, predicted latent state vector to be filtered out of the available information set $\Im_t$ according to transition equation

$$\hat{\boldsymbol{\xi}}_{t+1|t}=\hat{\mathbf{P}}'\hat{\boldsymbol{\xi}}_{t|t},$$

where also the transition matrix $\mathbf{P}$ will have to be estimated. It follows that

$$E[\mathbf{y}_{t+1}|\Im_t]=\mathbf{X}_{t+1}\hat{\Psi}\left(\hat{\mathbf{P}}'\hat{\boldsymbol{\xi}}_{t|t}\otimes\boldsymbol{\iota}_{l+q}\right). \tag{5}$$

4. The Data

We use monthly data for the longest available period, 1926:12 - 2004:12, a total of 937 observations per variable. Asset returns are from the Center for Research on Security Prices (CRSP) at the University of Chicago. In particular, we use data on the three most important segments of the US market: stocks (value-weighted stock

[8] Under the null of no misspecification, $\mathcal{I}_1(\tilde{\gamma})$ and $\mathcal{I}_2(\tilde{\gamma})$ should be identical. Since in our paper we do not perform misspecification tests based on the 'distance' between $\mathcal{I}_1(\tilde{\gamma})$ and $\mathcal{I}_2(\tilde{\gamma})$, we base our inferences on the "sandwich" form that combines information from both $\mathcal{I}_1(\tilde{\gamma})$ and $\mathcal{I}_2(\tilde{\gamma})$. Notice that under the null of a correctly specified model, this implies no loss of generality as assuming that $\mathcal{I}_1(\tilde{\gamma})=\mathcal{I}_2(\tilde{\gamma})$, we obtain $\mathcal{I}_3(\tilde{\gamma})\equiv\mathcal{I}_2(\tilde{\gamma})\left(\mathcal{I}_1(\tilde{\gamma})\right)^{-1}\mathcal{I}_2(\tilde{\gamma})=\mathcal{I}_1(\tilde{\gamma})=\mathcal{I}_2(\tilde{\gamma})$.

[9] The exception concerns the *number of non-zero rows* of the transition matrix $\mathbf{P}$, i.e. the number of regimes k. In this case, even under the assumption of asymptotic normality of the estimator $\tilde{\gamma}$, standard testing procedures suffer from non-standard asymptotic distributions of the likelihood ratio test statistic due to the existence of nuisance parameters under the null hypothesis.

returns for the NYSE, NASDAQ, and the AMEX exchanges), bonds (a CRSP index of 10-year to maturity US government bonds), and money market instruments (30-day Treasury bills, again from CRSP).[10]

Additionally, we employ 5 predictor variables which either correspond to macroeconomic aggregates or that have been associated to business cycle conditions in previous research. In the first group we have the CPI inflation rate (seasonally adjusted), the rate of growth of industrial production (seasonally adjusted), and the rate of growth of a measure of adjusted monetary base. These three series are available from FRED® II at the Federal Reserve Bank of St. Louis. The practice of seasonally adjusting the data in real time experiments (see Section 6) assumes that market participants are effectively able to 'see through' the veil of time series variation purely caused by seasonal factors. In the latter group we have two variables. The first is the dividend yield (calculated from CRSP data), and defined as aggregate dividends on the value-weighted CRSP portfolio of stocks over the previous twelve month period divided by the current stock price. The second is the default spread, defined as the differential yield on Moody's Bbb (low rating) and Aaa (high rating) seasoned corporate bonds with similar maturities. These two variables have played a key role in the recent literature on optimal asset allocation under predictable asset returns (see e.g. Brandt, 1999).

In our empirical analysis we use the following transformed variables. Given their crucial role in financial decisions, we study *excess* stock and bond returns, defined as the difference between nominal, realized monthly returns and the 1-month T-bill rate. For similar reasons, given the important literature on term spreads in the US yield curve, we use the long-short bond term spread (which is a notion of term premium) defined as the difference between the CRSP long-term bond and 1-month T-bill returns. Finally, also the monetary base growth is measured in real terms, by deducting from nominal rates of growth the realized CPI inflation rate.

Tables 1 and 2 report basic summary statistics. Mean values are consistent with commonly known facts: for instance, the mean excess stock return is 0.65% per month, i.e. 7.80% per year, which represents a typical value in the equity premium literature, with an annualized volatility of 19.1%; the mean term premium is 0.14% per month, i.e. 1.68% per year, a moderate but plausible average slope of the US term structure; the average annualized *real* T-bill rate is 0.60% which, summed to a mean annualized inflation rate of 3.12%, delivers a mean annualized nominal short-term interest rate of 3.72%, once more in line with the typical values reported in the asset pricing literature. Both real money and industrial production growth are positive on average, 0.48 and 2.52 percent in annualized terms, respectively. All the series display evident departures from a (marginal) Gaussian distribution, which would imply zero skewness (i.e. a symmetric distribution) and a kurtosis coefficient of 3. On the opposite, both excess stock returns and all macroeconomic variables are characterized by huge kurtosis values (in excess of 10), an indication of distributions with tails considerably fatter than a normal. The dividend yield has only moderate kurtosis, but it is also skewed to the right (which is to be expected, since the dividend yield cannot be negative by construction). Even in the case of excess bond returns, a formal Jarque-Bera test of marginal normal distribution rejects with a 0.000 p-value.

For all series but one (bond returns) there is evidence of statistically significant first-order serial correlation, as evidenced by Portmanteau Ljung-Box statistics (of order 4) in excess of the 1% critical value under a $\chi^2_{(4)}$. Similarly, there is evidence of volatility clustering (heteroskedasticity), as all Ljung-Box statistics (of order 4) applied to squared values of the variables are highly significant. We also examine pairwise correlation coefficients (see Guidolin and Ono, 2005). Although many coefficients are statistically different from zero, the largest correlations are between the dividend yield and the default spread (positive), and between the real

[10]The bond returns data are completed by using the Ibbotson-Sinquifeld data for the period 1926-1946. However, these early series are constructed following criteria that are identical to the ones employed by CRSP.

1-month T-bill rate and the inflation rate (negative, capturing the imperfect reaction of short-term rates to inflation).

5. **Empirical Results**

This section contains the bulk of our estimation, in-sample results, and an anticipation of some tests based on the one-step ahead predictive performance of regime switching models. Section 5.1 presents a number of model selection criteria and tests (formally, these are misspecification tests). Section 5.2 resolves some uncertainty on the model specification using relatively recent tests that base their power on certain properties of the density of the one-step ahead forecast errors under the null of no misspecifications. We end up estimating a particularly interesting four-state model in which the VAR component of the model is time-homogeneous, which implies stability of the predictability patterns involving the financial and macroeconomic variables under investigation. Section 5.3 presents parameter estimates for such a model.

5.1. *Model Selection*

The first task of our empirical analysis is the selection of an appropriate econometric model to represent the dynamic linkages between asset returns and macroeconomic forces. Since testing the hypothesis that the predictability patterns involving macroeconomic and financial variables has been stable over time revolves around achieving an accurate specification of a sufficiently rich model such as (1), we make an extensive effort. We estimate a large number of variants of (1) and use six alternative criteria to gauge the correct specification of the candidate models:

1. Davies (1977)-corrected likelihood ratio tests of the presence of multiple regimes $k \geq 2$, i.e. formal tests of the null hypothesis of $k = 1$ against the alternative of $k \geq 2$. As discussed in Garcia (1998), testing for the number of regimes may be tricky as under the null a few parameters of the unrestricted model (under the alternative) − i.e. the elements of the transition probability matrix associated to the rows that correspond to "disappearing states" — can take any values without influencing the likelihood function; these parameters are said to become a nuisance to the estimation.[11] Davies (1977) derives an upper bound for the significance level of the LR test under nuisance parameters:

$$\Pr\left(LR > x\right) \leq \Pr\left(\chi_1^2 > x\right) + \sqrt{2x}\exp\left(-\frac{x}{2}\right)\left[\Gamma\left(\frac{1}{2}\right)\right]^{-1},$$

   where $\Gamma(\cdot)$ is the standard gamma function.

2. Wolfe (1971)-modified LR tests for multiple regimes $k \geq 2$ (see e.g. Turner, Starz, and Nelson, 1989):

$$LR^{Wolf} = -\frac{2}{T}(T-3)\left[\ln L(\tilde{\gamma}) - \ln L(\tilde{\gamma}_r)\right] \xrightarrow{d} \chi_r^2$$

   where $\tilde{\gamma}_r$ is obtained under the null of a single-state model, T is the sample size, and $r = k(k-1)$ since in the absence of regime switching there are $k(k-1)$ parameters which cannot be estimated.

3. - 5. Three standard information criteria, i.e. the Akaike (AIC), Bayes-Schwartz (BIC), and Hannan-Quinn (H-Q) criteria. These statistics are supposed to trade-off in-sample fit with prediction accuracy and rely

[11] In the presence of nuisance parameters, even asymptotically the LR statistic fails to have a standard chi-square distribution.

on the principle that a correctly specified model should not only provide an accurate in-sample fit of the data at hand, but also prove useful to precisely forecast out-of-sample. In practice, information criteria identify the ex-ante potential out-of-sample by penalizing models with a large number of parameters. A well-performing model ought to minimize each of the information criteria. Information criteria do not explicitly suffer from nuisance parameter issues and are therefore employed to compare models with different number of regimes, as well model in given k-class but with different structure.[12]

6. Standard LR tests may be used within classes of models characterized by the same number of regimes, i.e. for which nuisance parameter problems do not exist so that standard asymptotic results apply.

Table 2 reports the outcomes of these model selection/specification tests.[13] Tests appear in the same order in which they have been listed above, and each of them corresponds to columns [4] through [8] of the table; columns [2] and [3] simply report the number of parameters implied by each MSIAH(k, p) model and the corresponding maximized log-likelihood function. The fourth column of Table 2 systematically tests the null of $k = 1$ against $k > 1$ (the exact number of regimes varies with the different models) and reports p-values calculated under Davies' upper bound. Obviously, even adjusting for the presence of nuisance parameters, the evidence against specifying traditional single-state models is overwhelming: the smallest LR statistic takes a value of 139, which is clearly above any conceivable critical value regardless of the number of restrictions imposed. We also perform (but do not report, to save space) Wolfe's modified LR-test and find anyway at least triple digit values, which again points to overwhelming rejections of the null of one regime only. This gives a first, crucial implication: the data propose strong evidence of time-variation in the coefficients of models capturing the dynamic linkages between financial and macroeconomic variables in the US. However, notice that this does not yet imply that patterns of predictability can effectively be treated as time-varying.

Once we establish that $k \geq 2$ is appropriate, this only rules out models of type $MSI(1, p)$, i.e. the first few rows of Table 2 only. We therefore proceed to select an appropriate model within the more general regime switching class $MSIAH(k, p)$ with $k \geq 2$. The range of models estimated in Table 3 is wide and spans models with $k = 2, 3, 4$, $p = 1, 2$, and with and without a regime-dependent covariance matrix. When possible, also models like (3) are estimated, since they are relatively parsimonious as well as economically interesting, implying the contemporaneous presence of regimes (in intercepts and covariances) along with dynamic linkages which are constant over time.[14] Columns [5]-[7] of Table 2 show that some tension exists among different criteria. The AIC is minimized by a richly parameterized MSIAH(4,1) model in which 444 parameters have to be estimated. We notice that although the MLE estimation could be carried out, issues may exist with a model that implies a saturation ratio (i.e. the number of available observations per estimated parameter) of only 16.9.[15] However, this is less than surprising as the AIC is generally known to select large

[12] These criteria are now relatively well-established in the regime switching literature, see e.g. Sola and Driffill (1994) and Guidolin and Timmermann (2006). Roeder and Wasserman (1997) formally argue in favor of using information criteria in mixtures of normals.

[13] In Table 2, the switching models are classified as MSIAH(k, p), where I, A and H refer to state dependence in the intercept, vector autoregressive terms and heteroskedasticity. p is the autoregressive order. Models in the class MSIH$(k, 0)$-$VAR(p)$ have regime switching in the intercept but not in the VAR coefficients. A MSI$(1, 0)$ is a simple multivariate Gaussian IID model; a MSIA$(1, p)$ is a Gaussian VAR(p) model.

[14] It is challenging to estimate models with $k \geq 5$ since the number of parameters grows to levels that cause the MLE-EM routines to fail. For instance, a MSIAH(5,1) model implies 560 paramters.

[15] A commo rule of thumb proposes that nonlinear estimation results based on saturation ratios less than to 20 ought to be taken with great caution.

models in nonlinear frameworks (see e.g. Fenton and Gallant (1996)). Next, the H-Q seems to be undecided between a relatively parsimonious MSIH(4,0)-VAR(1) model (with a saturation ratio of almost 30) and a richer MSIAH(3,1) (saturation ratio of 23). Notice that these two models imply a different number of regimes, 3 vs. 4. So, if on the one hand it seems obvious that regime switching matters, on the other hand the precise number of states required seems to be debatable. Finally, the BIC selects a relatively tight MSIH(4,0)-VAR(1) model.

Column [8] in Table 3 shows the outcomes of standard LR tests within classes of models characterized by the same number of regimes, i.e. for which nuisance parameter problems do not exist so that standard asymptotic results apply. The column should be read as testing the null that augmenting a smaller model by a certain feature − either increasing p to the next highest integer (type A) or making the covariance matrix regime-dependent (type H) − does not significantly increase the maximized log-likelihood. For instance, in the row of the MSIA(3,1) model we read: 'A: 0.000', meaning that going from a MSIH(3,0) to a MSIAH(3,1) the log-likelihood increases by more than one would impute to random chance; 'H: 0.000' to imply that the log-likelihood increase caused by a move from a MSIA(3,1) to a MSIAH(3,1) is highly significant. As previously observed in other nonlinear estimation contexts (see e.g. Fenton and Gallant (1996)), LR tests tend to be not very selective, as they fail to trade-off in-sample fit for parsimony. On any account, we find evidence that relatively rich regime switching models are required to fit the data at hand.

All in all, we are left with two plausible and competing candidate models. The first one is a four-regime MSIH(4,0)-VAR(1) model that is directly selected by both the H-Q and the parsimonious BIC criterion. The second is a three-regime MSIAH(3,1) model that obtains a good 'score' in a H-Q metric.[16] Notice at this point that these two models are structurally different both in a statistical and in an economic sense:

- MSIAH(3,1) is obviously a three-regime model while MSIH(4,0)-VAR(1) is a four-regime model. It is clearly important to understand how many regimes can be reliably singled out in the law of motion of US asset returns and macroeconomic variables.
- There is a deeper difference: MSIAH(3,1) relies on regime switching VAR(1) coefficients, i.e. in this model the dynamic linkages between financial markets and the macroeconomy are regime-dependent and therefore time-varying. This means that both the ways in which the current macroeconomic stance predicts subsequent asset prices and in which asset returns may forecast future economic conditions may change with time. On the contrary, MSIH(4,0)-VAR(1) implies constant VAR(1) coefficients and hence time homogeneous predictability involving financial and macroeconomic variables.

Since these differences appear crucial both under an econometric and an economic perspective, Section 5.2 uses additional tools to select between MSIAH(3,1) and MSIH(4,0)-VAR(1).

5.2. *Density Specification Tests*

The seminal work of Diebold et al. (1998) has spurred increasing interest in specification tests based on the h-step ahead accuracy of fit of a model for the underlying density. These tests are based on the probability integral transform or z-score. This is the probability of observing a value smaller than or equal to the

[16] We do not pursue estimation of the richer MSIAH(4,1) (selected by the AIC) because of the high probability of it being over-parameterized (its saturation ratio is almost half the MSIH(4,0)-VAR(1)).

realization $\tilde{\mathbf{y}}_{t+1}$ (assuming $h = 1$) under the null that the model is correctly specified. Under a k-regime mixture of normals, this is given by

$$\begin{aligned}\Pr\left(\mathbf{y}_{t+1} \leq \tilde{\mathbf{y}}_{t+1}|\Im_t\right) &= \sum_{i=1}^{k} \Pr\left(\mathbf{y}_{t+1} \leq \tilde{\mathbf{y}}_{t+1}|\Im_t, S_{t+1} = i\right) \Pr(S_{t+1} = i|\Im_t) \\ &= \sum_{i=1}^{k} \Phi_{q+l}\left(\boldsymbol{\Sigma}_i^{-1}\left[\mathbf{y}_{t+1} - \boldsymbol{\mu}_i - \sum_{j=1}^{p} \mathbf{A}_{j,i}\mathbf{y}_{t+1-j}\right]\right) \Pr(S_{t+1} = i|\Im_t) \equiv z_{t+1} \in \mathcal{R}, \quad (6)\end{aligned}$$

where $\Phi_{q+l}(\cdot)$ is the standard $(q + l)-$variate normal cdf. As stressed by Rosenblatt (1952), if the model is correctly specified, z_{t+1} should be independently and identically distributed (IID) and uniform on the interval $[0, 1]$. The uniform requirement relates to the fact that deviations between realized values and predicted ones should be conditionally normal and as such describe a uniform distribution once it is 'filtered through' an appropriate Gaussian cdf. The IID condition reflects the fact that if the model is correctly specified, forecast errors ought to be unpredictable and fail to show any detectable structure.

Unfortunately, testing whether a distribution is uniform is not a simple task, as test statistics popular in the statistics literature often rely on the IID-ness of the series, which is here at stake as well. Berkowitz (2001) has recently proposed a likelihood-ratio test that inverts Φ to get a transformed z-score,

$$z_{t+1}^{*} \equiv \Phi^{-1}(z_{t+1}),$$

which essentially turns the z-score back into a bell-shaped random variable. Provided that the model is correctly specified, z^{*} should be IID and normally distributed (IIN(0, 1)). We follow Berkowitz (2001) and use a likelihood ratio test that focuses on a few salient moments of the return distribution. Suppose the log-likelihood function is evaluated under the null that $z_{t+1}^{*} \sim IIN(0, 1)$:

$$L_{IIN(0,1)} \equiv -\frac{T}{2}\ln(2\pi) - \sum_{t=1}^{T} \frac{(z_t^{*})^2}{2}.$$

Under the alternative of a misspecified model, the log-likelihood function incorporates deviations from the null, $z_{t+1}^{*} \tilde{}\ IIN(0, 1)$:

$$z_{t+1}^{*} = \alpha + \sum_{j=1}^{w}\sum_{i=1}^{r} \beta_{ji}(z_{t+1-i}^{*})^{j} + \sigma u_{t+1}, \tag{7}$$

where $u_{t+1} \sim IIN(0, 1)$. The null of a correct return model implies $w \times r + 2$ restrictions – i.e., $\alpha = \beta_{ji} = 0$ ($j = 1, ..., w$ and $i = 1, ..., r$) and $\sigma = 1$ – in equation (7). Let $L(\hat{\alpha},\ \{\hat{\beta}_{ji}\}_{j=1}^{w}\,_{i=1}^{r},\ \hat{\sigma})$ be the maximized log-likelihood obtained from (7). To test that the null model (some version of (1)) is correctly specified, we can then use the following test statistic:

$$LR_{wr+2} \equiv -2\left[L_{IIN(0,1)} - L(\hat{\alpha}, \{\hat{\beta}_{ji}\}_{j=1}^{w}\,_{i=1}^{r}, \hat{\sigma})\right] \xrightarrow{d} \chi_{wr+2}^{2}.$$

In addition to the standard Jarque-Bera test that considers skew and kurtosis in the z-scores to detect non-normalities in z_{t+1}^{*}, it is customary to present three likelihood ratio tests, namely a test of zero-mean and unit variance ($w = r = 0$), a test of lack of serial correlation in the z-scores ($w = 1$ and $r = 1$) and a test that further restricts their squared values to be serially uncorrelated in order to test for omitted volatility dynamics ($w = 2$ and $r = 2$). Notice that a rejection of the null of normal transformed z-scores has the same meaning as rejecting the null of a uniform distribution for the raw z-scores, i.e. the model fails to generate a density with the appropriate shape. A rejection of the zero-mean, unit variance restriction points to specific problems in the location and scale of the density underlying the model. A rejection of the restriction that $\{z^*_{t+1}\}$ is IID points to dynamic misspecifications (serial correlation or heteroskedasticity).

Sometimes density specification tests are applied not the vector $\mathbf{y}_{t+1}$, but to each of its components:

$$\begin{aligned} \Pr\left(y^j_{t+1} \leq \tilde{y}^j_{t+1} | \Im_t\right) &= \sum_{i=1}^{k} \Phi\left(\sigma^{-1}_{j,i}\left[y^j_{t+1} - \mu_{j,i} - \mathbf{e}'_j \sum_{u=1}^{p} \mathbf{A}_{j,i}\mathbf{y}_{t+1-u}\right]\right) \Pr(S_{t+1} = i|\Im_t) \\ &\equiv z^j_{t+1} \qquad j = 1, ..., q + l, \end{aligned}$$

where $\sigma_{j,i}$ is the volatility of variable j in state i, and $\mathbf{e}_j$ is a vector with a one in position j and zeros elsewhere. The reason for this choice is that – especially when the dimension of $\mathbf{y}_{t+1}$ is high (like $q + l = 8$) – rejections from the z-scores based on the generalized multivariate residuals in (6) may often provided limited information on which variables are responsible for the rejection, i.e. on the dimensions over which the regime switching model is failing. This is also our choice.

Table 3 reports Berkowitz-style, transformed z-score tests for three models: a benchmark Gaussian VAR(1) inspired by the literature on the linear predictability of financial and macroeconomic variables; of course, the MSIAH(3,1) and MSIH(4,0)-VAR(1) models. Strikingly, a simple yet popular VAR(1) is resoundingly rejected by *all* tests and for *all* variables. Rejections tend to be harsh: the highest VAR(1) p-value appearing in the table is 0.001, i.e. there is actually a very thin chance that the data might have been generated by a simple linear Gaussian homoskedastic model. In fact, the rejections are so strong that it becomes difficult to understand in which direction one should be moving to amend the VAR(1) model.

The picture improves, albeit not drastically, when a MSIAH(3,1) model is estimated. For most tests and variables, the LR test statistics decline by a factor between 30 and 200% when we move from a single- to a multi-state model. However all (but one, the real T-bill rate when testing the scale/location properties) of the related p-values remain highly significant, indicating strong rejection of the null of correct specification of the three-state model in which the predictability patterns are time-varying. This means that specifying $k = 3$ and allowing the VAR(1) coefficients to change with the regime produces a density which is structurally different from the density that has generated the data.

Figures 1 provides further evidence on the sources of misspecifications within a MSIAH(3,1) model, by displaying the empirical distributions of $\{z^*_{t+1}\}$ for each of the eight variables (continuous lines) and comparing them with a normal variate with identical mean and variance (dotted lines). The model's faults are obvious for most of the variables, with the exceptions of T-bill short-term real yields and (possibly) the default spread. In many cases, the score distributions are either leptokurtic (too much mass at the center and in the tails, e.g. excess stock returns, CPI inflation, IP, and monetary base real growth rates) or even multi-modal (excess bond returns and the dividend yield). We also investigate (see Guidolin and Ono, 2005) quantile-quantile (q-q) plots for each of the eight variables that compare empirical quantiles with the standard normal ones that should apply under the null of no misspecifications. While under the null the q-q should approximately look like a 45-degree straight line in the q-q plane, in practice this happens only for real T-bill yields.[17]

On the contrary, the improvement is significant when we fit a four-state model. The p-values associated to the various tests generally increase and out of 32 combinations test/variable, we have that the null of no

[17] In fact, many of these plots assume an S-shape, i.e. the slope is too high at the center of the distribution (i.e. more mass is put under the distribution of $\{z^*_{t+1}\}$ than under a $N(0, 1)$) and too flat for intermediate values in the support.

misspecification fails to be rejected in 16 cases, with p-values exceeding 0.05. Of the remaining 17 tests, in 8 the p-values are between 0.01 and 0.05, i.e. the rejection is mild. However, the (marginal) conditional density of real T-bill yields and the default spread remains hard to model: for these two variables (which are responsible for 7 of the 10 highly significant rejections) there are signs of consistent departures from normality, of serial correlation in the scores, and of volatility clustering. The bright side is that for 3 variables − remarkably all of the financial variables, including the dividend yield − the tests give evidence of correct specification, with only weak signs of additional volatility clustering not simply accommodated by regime switching covariance matrices in excess stock returns.[18] Moreover, the improvement vs. the three-state model is clear: in only one case the LR statistic increases when the number of regimes is increased. Figure 2 visualizes the marked improvement (with the exception of some residual deviations for the default spread). Unreported qq-plots turn out to be almost perfect, i.e. roughly aligned around a 45 degrees line.

Finally, we apply density specification tests also at the multivariate level. We obtain a Jarque-Bera statistic of 6.45 (p-value of 0.040) and

$$
\begin{aligned}
LR_2 &= 5.56 \quad \text{p-value: } 0.062 \\
LR_3 &= 7.44 \quad \text{p-value: } 0.059 \\
LR_6 &= 13.04 \quad \text{p-value: } 0.042.
\end{aligned}
$$

Even if some issues remain regarding the overall shape of the distribution of the scores and possibly omitted heteroskedasticity, this is considerable evidence in favor of four states over three. Therefore in the following we present estimates from the MSIH(4,0)-VAR(1) model.

5.3. *A Four-State Model*

Before commenting on the nature of the estimated MSIH(4,0)-VAR(1) model, we want to stress what the selection of this model over a three-state MSIAH(3,1) means for the main thesis of this paper. The four-state model in Table 4 implies that although multiple regimes are required to approximate the joint conditional density of financial returns and macroeconomic variables, there is no evidence of time-variation in the structure of the predictability patterns linking financial markets and the economy at large. This means that even though expected returns and economic conditions (as captured by inflation, industrial production growth, and possibly the default and term spread) have been subject to recurring states, the dynamic linkages between financial prices and monetary and economic variables have been stable over time. This is quite a remarkable result: in spite of the flexibility of (1) in making the vector-autoregressive coefficients a function of the underlying latent regime, such an hypothesis is rejected by the BIC, cast in doubt by the H-Q, and again strongly rejected by the density specification tests of Section 5.2. In plain terms, this means that there does not seem to be a statistically sound way to differentiate between the measured response of stock prices to inflation news or of IP growth to movements in the bond term spread (just to cite two among the many interesting links) during the 1929 crash, the rapid growth of the post-WWII period, and the booming economy of the 1990s.[19]

Table 4 presents parameter estimates. Panel A reports standard ML estimates of a benchmark, single-state VAR(1) model. Panel B shows MLE-EM estimates of the four-state model.[20] Panel A shows that in a standard

[18] The good results for financial variables confirm Ang and Bekaert's (2002a) and Guidolin and Timmermann's (2004) findings that regime switching models provide an excellent description of the dynamic behavior of asset returns.

[19] The only related finding we are aware of is Estrella et al.'s (2003) conclusion that in a cross-section of countries where the term structure forecasts real activity, the relationship seems to be stable over time. We specialize our investigation to the US, but extend the analysis to a larger set of financial and macroeconomic variables.

[20] HAC standard errors are reported in parenthesis for the conditional mean function parameters (i.e. intercepts and VAR

VAR(1) many (if not the majority!) of the estimated coefficients are not significant. The implications for the predictability of financial returns are rather interesting: apart from a weak own serial correlation (coefficient is 0.10), excess stock returns are essentially unpredictable using any of the macro variables. A minor exception is the real rate of growth of the monetary base (coefficient 0.13), which forecasts the equity risk premium as in Homa and Jaffee (1971). The same is true for excess bond returns, which can be just (weakly) predicted from past excess stock returns (coefficient -0.02). Much more predictability characterizes real short-term interest rates, which are (as expected) highly persistent (coefficient 0.58) and can also be predicted off past default spreads (coefficient 1.28) and real IP growth (-0.05). Finally, there is limited evidence that past asset returns (e.g. past real T-bill yields) predict inflation and real growth (weakly forecastable using excess stock returns, with a positive coefficient), as in Lee (1992) or Plosser and Rouwenhorst (1994).

However, only limited confidence should be attributed to these results for three reasons. First, we know from Section 5.1 that single-state VAR(1) models are rejected even when account is taken of nuisance parameter issues. If there are multiple regimes in the data, we can expect that all estimates obtained from single-regime models might be biased. Second, notice that even the significant VAR coefficients in Panel A of Table 4 are often small. For instance, a one-standard deviation increase in the rate of growth of the real adjusted monetary base would translate (assuming that such a shock could be identified without other contemporaneous effects) into a 0.29% spike in excess stock return and a 0.09% increase in IP growth, and both are negligible in economic terms. Even a one standard deviation increase in the default spread would cause a decline in subsequent inflation of 0.08% only. Third, the fit provided by a VAR(1) is rather unsatisfactory. In Table 5 we calculate unconditional means from the VAR(1) using the standard result that under stationarity

$$E[\mathbf{y}_{t+1}] = (\mathbf{I}_{l+q} - \mathbf{A})^{-1}\boldsymbol{\mu},$$

and also period-specific means of the type

$$E_{\tau_0\to\tau_1}[\mathbf{y}_{t+1}] = \frac{1}{(\tau_1 - \tau_0 + 1)}\sum_{t=\tau_0}^{\tau_1}\left[\boldsymbol{\mu} + \sum_{j=1}^{p}\mathbf{A}_j\mathbf{y}_{t-j}\right],$$

where $\{\mathbf{y}_t\}_{t=\tau_0}^{\tau_1}$ consists of the values actually realized over the sample $[\tau_0, \tau_1]$. We isolate four sub-periods, 1926-1946, 1947-1966, 1967-1986, and 1987-2004. When τ_0 = 1926:12 and τ_1 = 2004:12, the mean refers to the full sample period. Comparing panels A (data) and B (VAR(1)) in Table 5, it is clear that $E_{26\to04}[\mathbf{y}_{t+1}]$ under the VAR matches the sample means but: (i) this does not apply to unconditional means, i.e. over the long run the VAR(1) forecasts values for the variables that often radically differ from the average of the data;[21] (ii) sub-samples can be found in which the observed means are different from $E_{\tau_0\to\tau_1}[\mathbf{y}_{t+1}]$ implied by the VAR. For instance, the VAR predicts too high an equity premium over 1926-1946 and too low a premium in 1987-2004. Similarly, IP growth is grossly over-estimated over the first part of the sample while the opposite occurs with reference to the last 18 years. In a sense, a VAR(1) model presents a 'rosy' picture of the Great Depression and misses altogether the stable period of phenomenal growth and of bull stock markets of the 1990s. This is not completely surprising as one of the roles of regime switching is to accommodate within the mixture bad and good periods as separate states.

coefficients). The VAR matrices ought to be read as the variables in the left flank being predicted by lagged values of the variables in the table's head. Parameter estimates in the last column of the transition matrix do not have standard errors because these parameters have to sum to one (by row) which constrains the value of the last coefficient.

[21] For instance, the annualized unconditional mean excess bond return is 1.08% $<$ 1.68% observed in sample; the annualized unconditional inflation rate is 5.88% $>$ 2.62% observed in sample.

Things greatly improve under a well-specified regime switching model. Table 4, panel B, starts by showing that the fraction of parameters in the conditional mean function that get precisely estimated substantially grows when multiple states are allowed. For instance, most of the intercept parameters are highly significant. However the most visible changes concern the "amount" (and in some cases, the structure) of the predictability patterns implied by the model. On one hand, modeling regimes erases all traces of own- and cross-serial correlation involving excess asset returns. This is unsurprising as structural breaks (regimes) are well known to artificially inflate the degree of persistence of the series. On the other hand, excess stock and bond returns now become highly predictable using lagged values of three variables: real T-bill rates (which forecast lower excess returns since the real short-term rate enters the discount rate in asset pricing models) as in Campbell (1987), the default spread (which forecasts higher excess returns, as a reward to increased risk, as in Fama and French (1989)), and the inflation rate (which forecasts lower future excess returns, presumably as a consequence of recessions that need to be induced to bring inflation under control) as in Fama and Schwert (1977). Excess stock returns are also predicted by past IP growth (as in Cutler, Poterba, and Summers (1989)), although the economic effect is rather negligible. Similarly, real T-bill yields are partially predictable from past default spreads (even if a one standard deviation increase in the spread increases T-bill real rates by 0.02% only).

We also obtain evidence of predictability of macro variables, especially IP growth which is not only persistent, but also predictable by past T-bill yields and default spreads with significant coefficients, similarly to Bernanke (1983). CPI inflation remains predictable from past interest rates and the default spread, besides being highly persistent. Interestingly, the ability of asset returns to predict future growth seems confined to single-regime models (see e.g. James, Koreisha, and Partch, 1985). In this sense, our results are similar to Canova and De Nicolo's (2000). When $k = 4$ most coefficients lose significance. This means that the bulk of predictability for inflation and real growth comes from the regime switching structure (and past macroeconomic conditions), and not from financial markets.[22]

Table 4 also reports the regime-dependent estimated volatilities and pairwise correlations implied by estimated variances and covariances. With limited exceptions, regimes 1 and 2 are characterized by moderate volatilities (of the shocks) and by correlations which tend to be smaller (in absolute value) than in the single-state VAR(1) of Panel A. On the opposite, regimes 3 and 4 imply higher volatilities and (at least for a majority of pairs) higher correlations. In fact, Tables 4, 5 and Figure 3 help us give an economic interpretation to the four regimes. Regime 1 is a bull/rebound state characterized (see the regime-specific unconditional means in Table 5) by high equity risk premia (14.5% on annualized basis), low or negative real short term interest rates, relatively high inflation (4.6% on annual basis), and high dividend yields. In this regime, all variables display moderate volatility, e.g. 13% for excess equity returns, 2.4% for excess bond returns, and 2.5% for inflation. This is a rebound state because its persistence is moderate (approximately 10 months) and it tends to follow bear regimes: the estimated transition matrix in Table 4 shows that starting from a bear/recession regime, 17% of the time the system transitions to a rebound (76% of the time, it stays in a bear regime). As a result, the mean dividend yield tends to be exceptionally high (5.2% vs. a historical mean of 3.8%), indication of the existence of good bargains in the stock market. The exceptional stock market performance tends to be

[22]Notice that while in a VAR(1), high growth forecasts future high inflation, this is not the case under the four-state model. However in both cases currently high inflation forecasts lower future growth, an sort of inverted Phillips curve. The effect may have some economic relevance: a one-standard deviation increase in inflation predicts a 0.50% decline in monthly growth in the former case, and 0.49% in the four-state model.

disjoint from real growth, which has in fact an unconditional mean of only 0.84% per annum. Consistently, the yield curve is relatively flat (the annualized term premium is 1.9%). Historically this regime coincides with the stock market bubble of 1927-1929, the Great-Depression rebound of 1934-1937, and most of the WWII and immediate post-war years (the 'atomic' age). After one spike in the mid-1950s, the occurrences of this regime have been rather episodic, although some late periods in the tech bubble of 1999-2000 are captured by this state. Interestingly, in this regime the correlation between shocks to inflation and to real-short term rates is not statistically different from -1, i.e. inflation shocks are transmitted one-to-one to real interest rates.

Regime 2 is a low volatility regime characterized by substantial growth (2.9% per year) and moderate inflation (3.2%). This is a persistent regime (15 months on average) in which also equity risk premia are high (5.4%), although equity prices correspond to much higher multiples than in regime 1 (the dividend yield has an unconditional mean of 2.8%). As experienced in the 1990s, real short term rates are low and credit cheap, just in excess of 1% per year. In fact, regime 2 captures most of the booming years between the mid-1950s and 1974 (the interruptions simply correspond to officially dated NBER recessions, picked up by state 4). After capturing a portion of the 1980s (but with frequent switches in and out of regime 3), the 1990s are entirely characterized as regime 2 episodes; the same is true of the more recent 2002-2004 period. Interestingly − despite the debate on the so-called 'New Economy' of the 1990s − the experience of that decade does not appear different from other periods of sustained growth and moderate inflation, like the 1960s.[23]

Regime 3 describes the stages of the business cycle when the economy emerges from a trough. Accordingly, regime 3 is scarcely persistent (5 months on average). This state is dominated by high real growth (almost 15% per year, although this figure must be taken with caution, since the duration of the regime is less than 6 months), high equity premia (5.9%), and a clearly upward sloping yield curve. Figure 3 shows that the early 1980s and 1990s are captured by regime 3. Finally, regime 4 represents a classical bear/recession state, in which risk premia are small, dividend yields relatively high (as stock prices decline), and inflation and real growth negative. Consistently with this interpretation, the default spread is high in this state (25 basis points vs. a historical mean of 9 points), while the regime duration is moderate (4 months), coherently with the fact that recession and bear markets are generally short-lived. This state also implies high volatility: e.g. excess stock returns have an annual volatility of 41%, excess bond returns of 7%; even IP growth is relatively unstable, with a standard deviation of 11% per annum. Figure 3 shows that regime 4 picks up all major US recessions after WWII, in addition to a long period that matches the Great Depression.

Table 5 shows that the correspondence between sub-period means $E_{\tau_0 \to \tau_1}[\mathbf{y}_{t+1}]$ from the MSIH(4,0)-VAR(1) model and in the data is remarkable and obviously superior to the simple VAR(1). For instance, the four-state model recognizes that the highest historical excess returns in the US were produced in the 1920s and 30s and then again during the 1990s. Even when the values of $E_{\tau_0 \to \tau_1}[\mathbf{y}_{t+1}]$ depart from sub-sample means, the four state model always gets the ranking across sub-samples right. This is another indication that − albeit the structure of predictability is constant over time − the presence of time-variation in some of the parameters gives an essential contribution at tracking and predicting the variables under investigation. A visual inspection (see Guidolin and Ono, 2005, for details) of plots of the in-sample fitted values reinforces this conclusion. First, we notice that in many instances, the MSIH(4,0)-VAR(1) values are simply much more volatile (hence able to track the underlying series) than for the VAR(1), which is to be expected. Second, the four-state model does a superior job at matching the dynamics of most of the series: for six out of eight,

[23] Figure 3 shows that most of the 1990s (89%) are characterized as the low-volatility regime 2. This is fully consistent with the now widespread notion that an age of "great moderation" started in the early 1990s, see Stock and Watson (2002).

the correlation between actual values and fitted ones is higher under a MSIH(4,0)-VAR(1) than under a VAR(1).[24] In some instances, the distance is major: e.g. such correlations are 0.61 > 0.50 for real T-bill yields, 0.65 > 0.53 for CPI inflation, and 0.30 > 0.22 for real money growth.

6. Forecasting Performance

Ultimately, what matters of a model is not its ability to produce an accurate in-sample fit, but rather its out-of-sample performance. In fact, when models are flexible because of their rich parameterizations, accuracy of fit is relatively unsurprising. However, rich parameterizations are also known to introduce large amounts of estimation uncertainty which ends up deteriorating their out-of-sample performance (see e.g. Rapach and Wohar, 2005). In the predictability literature this has been stressed among the others by Chan, Karceski, and Lakonishok (1998) who − studying the out-of-sample predictability of stock returns − found that except for the term and default spreads, macroeconomic variables tend to perform poorly. Goyal and Welch (2003) report that the dividend yield is a good predictor of stock returns only when the forecasting horizon is longer than 5 to 10 years. Neely and Weller (2000) re-examine the findings of Bekaert and Hodrick (1992) using a predictive metric. They show that VAR models are outperformed by much simpler benchmarks. They suggest that the poor forecasting performance may be due to underlying structural changes.

In fact, Bossaerts and Hillion (1999) argue that even the best linear models contain no out-of-sample forecasting power even when the specification of the models is based on statistical criteria that should penalize over-fitting. They speculate that the parameters of the selected models may be changing over time so that the correct model ought to be nonlinear, possibly of a regime switching type. Recent papers (e.g. Guidolin and Timmermann (2005a) for excess stock and bond returns or Guidolin and Timmermann (2005b) for short-term interest rates) have found that regime switching models may prove extremely useful to forecast over intermediate frequencies, such as monthly data. One wonders if a similar result holds for the larger vector under investigation in this paper and when financial and macroeconomic variables are jointly modeled.

To assess whether a four-state model offers any useful prediction performance, we implement the following 'pseudo out-of-sample' recursive strategy. For a given model, we obtain recursive estimates over expanding samples starting with 1926:12 - 1985:01, 1926:12 - 1985:02, etc. up to 1926:12 - 2004:11. This gives a sequence of 239 sets of parameter estimates specific to each of the models. For instance, the regime switching model (3) generates 239 sets of regime-specific intercepts, covariance matrices, and transition matrices, as well as of regime-independent VAR(1) coefficients. At each final date in the expanding sample − i.e. on 1985:01, 1985:02, etc. up to 2004:11 − we calculate h-month ahead forecasts for each of the 8 variables under study, i.e. including both financial and macroeconomic variables. For concreteness, in what follows we focus on $h = 1, 4, 12$ months. We call $\hat{y}_t^{(M,h)}$ the forecast generated at time t, by model M at horizon h (for simplicity, we omit to use a variable index, $i = 1, ..., 8$). Finally, we evaluate the accuracy of the resulting forecasts, by calculating the resulting forecast errors defined as $e_t^{(M,h)} \equiv y_{t+h} - \hat{y}_t^{(M,h)}$.[25]

We compare the MSIH(4,0)-VAR(1) model with three alternative benchmarks:

1. The three state, MSIAH(3,1) model (with time-varying VAR coefficients) that turned out to provide an

[24] The exceptions are the IP growth rate and the dividend yield (a tie). However in the latter case the tie is reached at a correlation of 0.98, i.e. both models do an excellent job.

[25] We also make sure that over the expanding sequence of samples 1926:12 - 1985:01, 1926:12 - 1985:02, etc, there is full justification for using a four-state model instead of a simpler VAR(1). Guidolin and Ono (2005) present plots that show that both recursive LR tests and information criteria rankings consistently "reject" the null of a single-state VAR(1).

interesting in-sample fit in Section 5.1;

2. A Gaussian VAR(1) model, which represents the single-state version of (3) (therefore nested in this model), already used as a benchmark in panel A of Table 4;

3. Neely and Weller's (2000) benchmark, i.e. a simple model in which expected excess returns are assumed to be constant (and equal to the sample mean) and other predictors follow random walks.

Also the benchmark parameters (or the sample means, in the case of Neely and Weller's benchmarks) are recursively re-estimated according to the scheme illustrated above. Table 6 reports summary statistics concerning the quality of the relative forecasting performance. In particular, we report three statistics illustrating predictive accuracy: the root mean-squared forecast error (RMSFE),

$$RMSFE^{(M,h)} \equiv \sqrt{\frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h}\left(y_{t+h}-\hat{y}_t^{(M,h)}\right)^2},$$

the prediction bias

$$Bias^{(M,h)} \equiv \frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h}\left(y_{t+h}-\hat{y}_t^{(M,h)}\right),$$

and the standard deviation of forecast errors,

$$SD^{(M,h)} \equiv \sqrt{\frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h}\left[(y_{t+h}-\hat{y}_t^{(M,h)})-\frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h}\left(y_{t+h}-\hat{y}_t^{(M,h)}\right)\right]^2}.$$

Notice that the three statistics are not independent as it is well known that $MSFE^{(M,h)} \equiv \left[Bias^{(M,h)}\right]^2 + \left[SD^{(M,h)}\right]^2$, i.e. the MSFE can be decomposed in the contribution of bias and variance of the forecast errors. At a one-month horizon, the four-state model seems to be rather useful for forecasting purposes, as it displays the lowest MSFE for five out of eight variables. When compared to the other three benchmarks, the reduction in MSFE seems particularly important for excess stock returns and real money growth (the ratio of the four-state MSFE to the next best MSFE are 0.88 and 0.82, respectively). However for two variables simple benchmarks (recursive sample means for excess bond returns and the random walk for the default spread) outperform all other models, although the difference relative to the four-state model is small. Consistently with the results in Section 5.2, when $h = 1$ a richer three-state model with regime-dependent VAR coefficients is systematically dominated by the three-state model in terms of MSFE minimization; also a VAR(1) seems too simple to be able to produce useful forecasts.[26] It is also interesting to notice that in general Neely and Weller's benchmarks produce biases that are close to (or even lower than) those characterizing the four-state model, the difference being in the variances of the forecast errors: regime switching models produce higher biases (i.e. they often miss the actual point value of the forecast variable, but they seem to be able to systematically move in such a way that greatly reduces the variance of the errors. This is what one expects if the model's regimes correctly identify the economy's turning points in real time. This result is actually consistent with the remarks by Neely and Weller (2000) who suggest that the poor forecasting performance of

[26]The exceptions are the default spread, which is best modeled as a three-state process, and the inflation rate which is best predicted by a simple VAR(1).

more elaborate prediction models may be due to underlying time-varying coefficients, which is what regime switching models ought to capture.

At longer horizons, the superior performance of our four-state model tends to deteriorate slightly, while the three-state model acquires merit. For instance, at $h = 12$ the MSIAH(3,1) presents the lowest MSFE for half of the variables, although the three-state model still outperforms all benchmarks in predicting excess stock returns, the dividend yield, and real money growth. Interestingly, the family of regime switching models consistently outperform single-state (as well as naive) benchmarks at longer horizon: this is the case for five out of eight variables at $h = 4$, and for seven out of eight at $h = 12$. This is unsurprising in the light of the evidence we have offered in Section 5 of the fact that regime switching models generally perform well at fitting the multivariate density of the data: as h grows, forecasts from models with regimes rely less and less on the accurate identification of the most recent turning points of the economy, and increasingly on the overall properties of the unconditional density of the data. In this respect, it seems that both the three- and the four-state models offer an appreciable performance, while both single-state VARs and naive benchmarks appear to be grossly off for many variables (in particular, excess stock returns and real money growth).

6.1. *Testing Differential Predictive Accuracy*

Further examination of Table 6 reveals that – although rankings match our general comments – differences between models are often modest, which casts some doubt on how useful the three-state model may be in practice (e.g. see the small differences concerning real T-bill yields). This observation opens a further issue: are the out-of-sample performances different enough to allow us to draw any conclusions on the relative accuracy of alternative models? We start by implementing the forecast accuracy comparison tests proposed by Diebold and Mariano (1995). Let

$$dif_t^{(m,n,h)} \equiv \left(e_t^{(m,h)}\right)^2 - \left(e_t^{(n,h)}\right)^2$$

be the differential loss (in our case the standard square loss) of model m relative to the loss from model n. We can test the significance of the differences between two sets of forecast errors based on the statistic

$$DM_h^{(m,n)} = \frac{\frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h} dif_t^{(m,n,h)}}{\widehat{\sigma}(dif_t^{(m,n,h)})} \overset{a}{\sim} N(0,1), \tag{8}$$

where $\widehat{\sigma}(dif_{t,h}^{(m,n)})$ is an HAC estimate of the standard error of the loss function differential, i.e.

$$\widehat{\sigma}(dif_{t,h}^{(m,n)}) = \sum_{j=-h}^{h} Cov\left[dif_{t,h}^{(m,n)}, dif_{t+j,h}^{(m,n)}\right].$$

Giacomini and White (2004, henceforth GW) have recently argued that standard out-sample predictive ability tests are not necessarily appropriate for real-time forecast methods. For instance, both $e_t^{(m,h)}$ and $e_t^{(n,h)}$ are usually generated from parametric models that have to be recursively estimated over time, i.e. $e_{t,h}^{(m)}$ and $e_{t,h}^{(n)}$ have to be themselves estimated using $\hat{e}_{t,h}^{(m)}$ and $\hat{e}_{t,h}^{(n)}$. This means that $\widehat{dif}_t^{(m,n,h)} \equiv \left(\hat{e}_{t,h}^{(m)}\right)^2 - \left(\hat{e}_{t,h}^{(n)}\right)^2$ will be probably polluted by errors caused by estimation uncertainty concerning the parameters of the underlying models.[27] From a methodological point of view, GW shift the focus from the unconditional mean of differences

[27] The theory in Diebold and Mariano (1995) was developed for the baseline case of no parameter uncertainty. Exceptions exist: for instance, the random walk model does not require estimation of any parameters.

in loss functions (as in (8)) across prediction models to the conditional expectation of such differences across forecast methods, i.e. from the null

$$H_o : E\left[dif_t^{(m,n,h)}\right] = 0$$

under true parameter values (i.e. probability limits of parameter estimates), to

$$H_o' : E_t\left[dif_t^{(m,n,h)}\right] = 0$$

under the estimated parameters of models m and n. GW's approach delivers a few interesting payoffs, for instance conditional tests directly account for the effects of parameter uncertainty by expressing the null H_o' directly in terms of estimated parameters and fixed estimation windows.

In the case $h = 1$ Giacomini and White (2004) exploit the fact that the null is equivalent to stating that $\{dif_t^{(m,n,h)}\}$ is a martingale difference sequence, implying that for all measurable functions g_t in the information set at time t it should be $E\left[g_t \cdot dif_t^{(m,n,h)}\right] = 0$.[28] They show that given a set of q measurable functions $\mathbf{g}_t$, the null of equal conditional predictive ability (CPA) for a pair of models m, n can be tested using the statistic

$$GW_{\mathbf{g}}^{(m,n)}(h) \equiv (240\text{-}h)\left[\frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h}\mathbf{Z}_t^{(m,n)}(h)\right]'\left[\hat{\mathbf{\Omega}}(Z_t^{(m,n)}(h))\right]^{-1}\left[\frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h}\mathbf{Z}_t^{(m,n)}(h)\right] \qquad (9)$$

where

$$\mathbf{Z}_t^{(m,n)}(h) \equiv \mathbf{g}_t \cdot dif_t^{(m,n,h)} \qquad \hat{\mathbf{\Omega}}(Z_t^{(m,n)}(h)) \equiv \sum_{j=-h}^{h}\widehat{Cov}\left[\mathbf{Z}_t^{(m,n)}(h), \mathbf{Z}_{t+j}^{(m,n)}(h)\right]$$

Under regularity conditions, $GW_{\mathbf{g}}^{(m,n)}(h) \overset{a}{\sim} \chi^2_{(q)}$. The power properties of the tests obviously depend on the choice of test functions in $\mathbf{g}_t$, although it is also clear that rejections of H_o' with respect to some set of functions $\mathbf{g}_t$ may give indications as to ways in which the forecasting performance could be improved.

Table 7 illustrates the results of such tests for each possible pair of models involving the four-state MSIH(4,0)-VAR(1) and for $h = 1$, 12 (results for $h = 4$ are available upon request). The table is organized around benchmark models (in the flank of the display) and divided in two panels concerning alternative forecast horizons. Since the VAR(1) rarely displays superior performance, we omit it from the table.[29] For each variable and pair of forecast models, we report the values of the statistics $DM_h^{(m,n)}$ and $GW_{\mathbf{g}}^{(m,n)}(h)$. In the latter case, Table 7 show the results of CPA tests when $\mathbf{g}_t \equiv [1 \;\; \Delta dif_t^{(m,n,h)}]'$ ($q = 2$), as suggested in GW.[30] Values in parenthesis are p-values for the two statistics. Small p-values (say below 0.1) indicate superior predictive accuracy of the four-state model (more generally, of the model in the table's header, when compared to the benchmark in the flank) when either $DM_h^{(m,n)}$ is *negative* or $GW_{\mathbf{g}}^{(m,n)}(h)$ is reported with

[28] In the case $h \geq 2$, $\{dif_t^{(m,n,h)}\}$ is not a martingale difference sequence but $\forall g_t$ in the information set, $\{g_t \cdot dif_t^{(m,n,h)}\}$ should be "finitely correlated", i.e. uncorreled after a certain number of lags.

[29] From Table 6, the exception concerns inflation forecasts at $h = 1$. We therefore compute predictive accuracy tests for this case and find: $DM(1) = -0.89$ vs. the MSIH(3,0)-VAR(1), -0.11 vs. the MSIAH(4,1), and -0.69 vs. the random walk, i.e. values which fail to be stastically significant. The corresponding values of the $GM(1)$ test statistic are 3.90, 2.52, and 5.19, i.e. there is only some evidence that a VAR(1) might outperform the random walk.

[30] We also compute CPA tests when $\mathbf{g}_t \equiv [1 \;\; \Delta dif_t^{(m,n,h)} \;\; \Delta dif_{t-1}^{(m,n,h)} \;\; e_t^{(m,h)} \;\; e_{t-1}^{(m,h)} \;\; e_t^{(n,h)} \;\; e_{t-1}^{(n,h)}]'$, i.e. $q = 7$. Results are qualitatively similar (in general, more favorable to the four-state model) and therefore omitted. For instance, for the MSIH(4,0)-VAR(1) model applied to excess stock returns, $GW(1) = 22.55$ compared to the MSIAH(4,1) and $GW(1) = 11.56$ compared to Neely-Waller's benchmarks. Both statistics correspond to very small p-values.

a 'star';[31] for clarity, values associated with low p-values are boldfaced. For instance, when $h = 1$, Table 7 shows that the four-state model has superior predictive accuracy for excess stock returns than a three-state model according to both the DM and GW tests (with p-values of 0.02 and 0.07, respectively, i.e. both unconditionally and conditionally), and there is also some evidence of superior forecasting performance relative to Neely-Waller's benchmark (but only in the unconditional DM metric, with a p-value of 0.07).

The evidence of in favor of the four-state model is mixed at the one-month horizon: the only solid evidence is that MSIH(4,0)-VAR(1) outperforms all other models at forecasting the dividend yield and real money growth rate, as in these cases DM and GW tests give homogeneous indications; as already mentioned, there is also some evidence of superior performance in predicting excess stock returns, but this result is not robust to conditional, GW-type tests. A simple Neely and Weller's (2000) recursive sample mean shows a good performance when it comes to prediction of bond excess returns. On the contrary, the evidence in favor of regime switching models is unequivocal for long forecast horizons. In this case we have a "split" of relative performances, in the sense that both DM and GW tests show superior predictive accuracy of a four-state model in forecasting dividend yields and real money growth, and of the three-state model with respect to prediction of inflation and the default spread. The performance of the benchmarks remains unchallenged on the excess bond returns series. In summary, predictive accuracy tests show that differences in Table 6 could not be attributed entirely to chance. Consistently with much forecasting literature, at short horizons it remains true that outperforming the random walk or simple constant expected returns benchmarks is challenging, although some traces of over-performance could be found. At longer horizon, it seems clear that the flexibility of mixture (regime switching) models may be required to capture the salient features of the multivariate distribution of financial and macroeconomic variables, although some uncertainty remains as to the actual choice of the number of regimes.

7. **Predicting Risk-Return Trade-Offs**

Although many decision makers might have an interest for the results so far and the general issue of whether the dynamic linkages between financial markets and macroeconomic factors have changed over time is of intrinsic interest, there is a class of economic agents that has a straightforward use for the model in Table 4: portfolio managers. In fact, while financial economists have been worrying about the implications of predictability for market efficiency and the theoretical properties of equilibrium asset prices, money managers have attempted to exploit statistical predictability patterns − including the reaction of asset returns to macroeconomic announcements − to improve the return-risk properties of their portfolios. In this sense, such decision makers would be mostly interested not in point forecasts of future asset returns (and possibly a few of the macroeconomic aggregates, such as inflation and real growth) or in the ability of (3) to approximate their conditional joint density, but in correctly forecasting (one-step ahead) Sharpe ratios,

$$SR_{t,t+1}^{(M,i)} \equiv \frac{E_t^{(M)}[y_{t+1}^i]}{\sqrt{Var_t^{(M)}[y_{t+1}^i]}},$$

where i indexes excess stock and bond returns, respectively. Sharpe ratios are the standard measure of the compensation per unit of risk used in the financial industry. Additionally, in simple (and myopic) mean-

[31] For instance, $DM_1^{(MSIH,VAR)} < 0$ indicates that on average $dif_t^{(MSIH,VAR,1)} \equiv (e_t^{(MSIH,1)})^2 - (e_t^{(VAR,1)})^2 < 0$, i.e. the four-state model forecasts better than the VAR.

variance asset allocation framework, the optimal weight to be assigned to some asset i would be $\omega_{t,t+1}^{(M,i)} = SR_{t,t+1}^{(M,i)}/\gamma$, where γ is a risk aversion parameter.[32]

Figure 4 shows 1-month ahead, predicted Sharpe ratios for both stocks and bonds. Such predictions are calculated under both the VAR(1) and the MSIH(4,0)-VAR(1) models and by recursively estimating the two models according to the same expanding-window format employed in Section 6. The VAR(1) model is selected to offer an example of the dangers of ignoring regimes in models in which one intends to exploit the existence of predictability of financial returns from macroeconomic variables. Once more, the VAR(1) model: (i) generates flat predictions which are hardly compatible with active portfolio management; (ii) misses the specificity of the 'tech bubble' of the late 1990s. In fact, a simple linear model generates roughly stable $SR_{t,t+1}^{(M,h)}$s between 0.25 and 0.30 for stocks and between 0 and 0.1 for bonds. Such values might have suggested roughly constant portfolio weights over time (probably tilted towards stocks), and quite unreasonable investment policies, with large and possibly increasing weight assigned to stocks throughout the 1990s, even at the peak of the bubble.

On the contrary, the four-state model gives reasonable risk-return insights. The equity Sharpe ratio fluctuates over time in a counter-cyclical manner, i.e. $SR_{t,t+1}^{(M,1)}$ is high during recessions and declines during economic booms. Therefore it suggests large commitment to equities in 1985, in the early 1990s, and recently during 2000-2001, when $SR_{t,t+1}^{(M,1)}$ exceeds 0.2 and achieves peaks of 0.3. Importantly, between 1998 and 2000, $SR_{t,t+1}^{(M,1)}$ strongly signals the presence of a bubble, i.e. of a modest compensation for risk, as the ratio declines below 0.1 and touches 0 in a few months. Under regime switching, $SR_{t,t+1}^{(M,2)}$ is stable for bonds, but still provides strong and possibly useful signals, as $SR_{t,t+1}^{(M,2)}$ becomes volatile and often exceeds 0.2 at several points between 1998 and 2002.

8. Conclusion

This paper has proposed to use multivariate regime switching models to study the possibility that the predictability patterns involving US asset returns and macroeconomic variables be time-varying. Using a long monthly data set (1926-2004) we find overwhelming evidence of regimes in the joint process for returns and macroeconomic factors, although the null of a stable set of dynamic predictability relationships cannot be rejected. In this sense, famous historical experiences concerning the linkages (or absence thereof) between financial markets and the real economy − for instance the Great Depression and the tech bubble of the 1990s − are not as heterogeneous as commonly thought. The good performance of our four-state model at fitting the entire density of the data and its useful forecasting performance stress that payoffs may exist in explicitly modeling the presence of regimes, although it is clear that when switches in intercepts and covariance matrices are accounted for, no need is left for explicitly time-varying predictability patterns.

Several extensions of this paper could be attempted. First of all, although we have tried to focus on a

[32] This is a very simplistic asset allocation strategy in which preferences are assumed to be constant, and investors need to actually care only about mean and variance. Additionally, such an investor would have to be myopic and ignore time-varying investment opportunities. We thank an anonymus referee for drawing our attention to the limitations of a direct mapping from Sharpe ratios to portfolio weights in this context. Notice that MMS models imply rich (as well as dynamic, as the state probabilities are updated) departures from standard zero skewness and constant kurtosis levels, which contradicts an assumption of mean-variance preferences. Ang and Bekaert (2002a) and Guidolin and Timmermann (2004, 2005a) explore the optimal asset allocation implications of regimes and predictability in the joint (and time-varying) distribution of excess asset returns. Notice that portfolio choice results should not be interpreted as suggestive of asset pricing implications because our analysis has taken asset returns as given. Our (partial equilibrium) results therefore concern an atomistic investor who could have chosen portfolio weights with no impact on observed asset prices.

selected number of macroeconomic predictors that had performed well so far at forecasting asset returns, the system might be expanded either to include other variables or to model different ones, to test the robustness of our conclusions. Second, in this paper we have pursued a theory-free approach that only focuses on the statistical aspects of predictability and on their implications for financial decisions. However, papers like Ang and Piazzesi (2003) and Wickens and Flavin (2001) have recently shown how predictability involving macroeconomic factors, no-arbitrage asset pricing, and optimal asset allocation can be brought together by imposing appropriate restrictions. Third, there is nothing compelling about using multivariate regime switching models to study time-varying linkages between financial markets and the macroeconomy. Other modeling approaches might prove useful. Among others, Bredin and Hyde (2005) have recently applied smooth transition regressions to study the nonlinear relationships between eight international stock returns and a number of macroeconomic variables. Finally, Section 7 only provides a brief and incomplete example of possible applications of models with regimes to real-time financial decisions. Ongoing research is currently investigating whether and how useful models with regimes may be for asset allocation purposes.

References

[1] Ang A., and G., Bekaert, 2002a, "International Asset Allocation with Regime Shifts", *Review of Financial Studies,* 15, 1137-1187.

[2] Ang, A., and G., Bekaert, 2002b, "Regime Switches in Interest Rates", *Journal of Business and Economic Statistics*, 20, 163-182.

[3] Ang, A., and M., Piazzesi, 2003, "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables", *Journal of Monetary Economics*, 50, 745-787.

[4] Balvers, R., T., Cosimano, and B., Mcdonald, 1990, "Predicting Stock Returns in an Efficient Market", *Journal of Finance*, 45, 1109-1128.

[5] Bekaert, G., and R., Hodrick, 1992, "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets", *Journal of Finance*, 47, 467-509.

[6] Berkowitz, J., 2001, "Testing Density Forecasts with Applications to Risk Management", *Journal of Business and Economic Statistics*, 19, 465-474.

[7] Bernanke, B., 1983, "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression", *American Economic Review*, 73, 257-276.

[8] Bernanke, B., and A., Blinder, 1992, "The Federal Funds Rate and the Channels of Monetary Transmission", *American Economic Review*, 82, 901-921.

[9] Bossaerts, P., and P., Hillion, 1999, "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?", *Review of Financial Studies*, 12, 405-428.

[10] Brandt, M., 1999, "Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach", *Journal of Finance*, 54, 1609-1645.

[11] Bredin, D., and S., Hyde, 2005, "Regime Changes in the Relationship Between Stock Returns and the Macroeconomy", mimeo.

[12] Campbell, J. 1987, "Stock Returns and the Term Structure", *Journal of Financial Economics*, 18, 373-399.

[13] Canova, F., and G., De Nicolo, 2000, "Stock Returns, Term Structure, Inflation and Real Activity: An International Perspective", *Macroeconomic Dynamics*, 4, 343-372.

[14] Chan, L., J., Karceski, and J., Lakonishok, 1998, "The Risk and Return from Factors", *Journal of Financial and Quantitative Analysis*, 33, 159-188.

[15] Cutler, D., J., Poterba, and L., Summers, 1989, "What Moves Stock Prices?', *Journal of Portfolio Management*, 15, 4-12.

[16] Davis, P., and G., Fagan, 1997, "Are Financial Spreads Useful Indicators of Future Inflation and Output Growth in EU Countries?", *Journal of Applied Econometrics*, 12, 701-714.

[17] Davies, R., 1977, "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative", *Biometrika*, 64, 247-254.

[18] Dempster, A., M., Laird, and D., Rubin, 1977, "Maximum Likelihood from Incomplete Data Via the EM Algorithm", *Journal of the Royal Statistical Society*, B 39, 1-22.

[19] Diebold, F., T., Gunther, and A., Tay, 1998, "Evaluating Density Forecasts", *International Economic Review*, 39, 863-883.

[20] Diebold, F. , and R., Mariano, 1995, "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, 253-263.

[21] Du, D., 2005, "Monetary Policy, Stock Returns and Inflation", *Journal of Economics and Business*, forthcoming.

[22] Emery, K., 1996, "The Information Content of the Paper-Bill Spread", *Journal of Economics and Business*, 48, 1-10.

[23] Estrella, A., and G., Hardouvelis, 1991, "The Term Structure as a Predictor of Real Economic Activity", *Journal of Finance*, 46, 555-576.

[24] Estrella, A., A., Rodrigues, and S., Schich, 2003, "How Stable Is the Predictive Power of the Yield Curve? Evidence from Germany and the United States", *Review of Economics and Statistics*, 85, 555-566.

[25] Fama, E., and K., French, 1988, "Dividend Yields and Expected Stock Returns", *Journal of Financial Economics*, 19, 3-29.

[26] Fama, E., and K., French, 1989, "Business Conditions and Expected Returns on Stocks and Bonds", *Journal of Financial Economics*, 25, 23-49.

[27] Fama, E., and K., French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics*, 33, 3-36.

[28] Fama, E., and W., Schwert, 1977, "Asset Returns and Inflation", *Journal of Financial Economics*, 5, 115-146.

[29] Fenton, V., and R., Gallant, 1996, "Qualitative and Asymptotic Performance of SNP Density Estimation", *Journal of Econometrics*, 74, 77-118.

[30] Fischer, S., and R., Merton, 1984, "Macroeconomics and Finance: The Role of the Stock Market", *Carnegie-Rochester Conference Series on Public Policy*, 21, 57-108.

[31] Flannery, M., and A., Protopapadakis, 2002, "Macroeconomic Factors Do Influence Aggregate Stock Returns", *Review of Financial Studies*, 15, 751-782.

[32] Garcia, R., 1998, "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", *International Economic Review*, 39, 763-788.

[33] Giacomini, R., and H., White, 2004, "Tests of Conditional Predictive Ability", mimeo, UCLA and UCSD.

[34] Goyal, A., and I., Welch, 2003, "Predicting the Equity Premium with Dividend Ratios", *Management Science*, 49, 639-654.

[35] Gray, S., 1996, "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process", *Journal of Financial Economics*, 42, 27-62.

[36] Guidolin, M., and S., Ono, 2005, "Are the Dynamic Linkages Between the Macroeconomy and Asset Prices Time-Varying?", mimeo, Federal Reserve Bank of St. Louis and University of York.

[37] Guidolin, M., and A., Timmermann, 2004, "Strategic Asset Allocation under Regime Switching", mimeo, Federal Reserve Bank of St. Louis and UCSD.

[38] Guidolin, M., and A., Timmermann, 2005a, "Economic Implications of Bull and Bear Regimes in UK Stock and Bond Returns", *Economic Journal*, 115, 111-143.

[39] Guidolin, M., and A., Timmermann, 2005b, "Optimal Forecast Combination Weights under Regime Shifts: An Application to US Interest Rates", mimeo, Federal Reserve Bank of St. Louis and UCSD.

[40] Guidolin, M., and A., Timmermann, 2006, "An Econometric Model of Nonlinear Dynamics in the Joint Distribution of Stock and Bond Returns", *Journal of Applied Econometrics*, 21, 1-22.

[41] Hamilton, J., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57, 357-384.

[42] Hamilton, J., 1993, "Estimation, Inference, and Forecasting of Time Series Subject to Changes in Regime", in Maddala, G., Rao, C., and Vinod, H., *Handbook of Statistics*, vol. 11., Amsterdam: North Holland.

[43] Hamilton, J., and R., Susmel, 1994, "Autoregressive Conditional Heteroskedasticity and Changes in Regime", *Journal of Econometrics*, 64, 307-333.

[44] Hardouvelis, G., 1987, "Macroeconomic Information and Stock Prices", *Journal of Economics and Business*, 39, 131-140.

[45] Homa, K., and D., Jaffee, 1971, "The Supply of Money and Common Stock Prices", *Journal of Finance*, 26, 1045-1066.

[46] Jaditz, T., L., Riddick, and C., Sayers, 1998, "Multivariate Nonlinear Forecasting: Using Financial Information to Forecast the Real Sector", *Macroeconomic Dynamics*, 2, 369-382.

[47] James, C., S., Koreisha, and M., Partch, 1985, "A VARMA Analysis of Casual Relations among Stock Returns, Real Output, and Nominal Interest Rates", *Journal of Finance*, 40, 1375-1384.

[48] Kaul, G., 1987, "Stock Returns and Inflation: the Role of the Monetary Sector', *Journal of Financial Economics*, 18, 253-276.

[49] Krolzig, H.-M., 1997, *Markov-Switching Vector Autoregressions*, Berlin, Springer-Verlag.

[50] Lee, B.-S., 1992, "Causal Relations among Stock Returns, Interest Rates, Real Activity, and Inflation", *Journal of Finance*, 47, 1591-1603.

[51] Leroux, B., 1992, "Maximum Likelihood Estimation for Hidden Markov Models", *Stochastic Processes and their Applications*, 40, 127-143.

[52] Neely, C., and P., Weller, 2000, "Predictability in International Asset Returns: Reexamination", *Journal of Financial and Quantitative Analysis*, 35, 601-620.

[53] Pesaran, H., and A., Timmermann, 1995, "Predictability of Stock Returns: Robustness and Economic Significance", *Journal of Finance*, 50, 1201-1228.

[54] Plosser, C., and G., Rouwenhorst, 1994, "International Term Structures and Real Economic Growth", *Journal of Monetary Economics*, 33, 133-156.

[55] Rapach, D., 2001, "Macro Shocks and Real Stock Prices", *Journal of Economics and Business*, 53, 5-26.

[56] Rapach, D., and M., Wohar, 2005, "In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining,", *Journal of Empirical Finance*, forthcoming.

[57] Rapach, D., M., Wohar, and J., Rangvid, 2005, "Macro Variables and International Stock Return Predictability", *International Journal of Forecasting*, 21, 137-166.

[58] Roeder, K., and L., Wasserman, 1997, "Practical Bayesian Density Estimation Using Mixtures of Normals", *Journal of the American Statistical Association*, 92, 367-377.

[59] Sola, M., and J., Driffill, 1994, "Testing the Term Structure of Interest Rates Using a Stationary Vector Autoregression with Regime Switching", *Journal of Economic Dynamics and Control*, 18, 601-628.

[60] Stock, J., and M., Watson, 2002, "Has the Business Cycle Changed and Why?", *NBER Macroeconomics Annual 2002*, MIT Press.

[61] Stock, J., and M., Watson, 2003, "Forecasting Output and Inflation: the Role of Asset Prices", *Journal of Economic Literature*, 41, 788-829.

[62] Turner, C., R., Starz, and C., Nelson, 1989, "A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market", *Journal of Financial Economics*, 25, 3-22.

[63] Wickens, M., and T., Flavin, 2001, "Macroeconomic Influences and Optimal Asset Allocation", mimeo, University of York.

[64] Wolfe, J., 1971, "A Monte Carlo Study of the Sampling Distribution of the Likelihood Ratio for Mixture of Multinormal Distributions", San Diego, NITS Research Laboratory working paper.

Table 1

Summary Statistics for Excess Stock and Bond Returns vs. Prediction Variables

The table reports a few summary statistics for monthly CRSP excess stock and (long-term government bond) return series, and a few macroeconomic variables employed as predictors of excess asset returns. Excess returns are calculated by difference with 30-day T-bill yields. The sample period is 1926:12 – 2004:12. In the case of equities, the CRSP universe spans stocks listed on the NYSE, the NASDAQ, and the AMEX. Data on bond returns refer to the CRSP 10-Year Treasury benchmark. All returns are expressed in monthly percentage terms. LB(j) denotes the j-th order Ljung-Box statistic.

Series	Mean	Median	St. Dev.	Skewness	Kurtosis	Jarque-Bera	LB(4)	LB(4)-squares
	Excess Asset Returns (Risk Premia)							
Value-weighted excess stock returns	0.6482	0.9900	5.4946	0.2133	10.6124	2269**	21.716**	166.87**
Excess bond returns (term premium)	0.1447	0.1400	1.8808	0.2447	5.5932	271.9**	5.1774	176.31**
	Prediction Variables							
12-month cumulated dividend yield	3.8132	3.6340	1.4987	0.9542	5.8183	452.3**	3334**	2829**
Real 1-month T-bill yield	0.0540	0.0700	0.5114	-1.9764	21.0381	13313**	542.13**	79.833**
Default spread	0.0943	0.0730	0.0600	2.4203	11.3805	3657**	3284**	2683**
CPI inflation rate	0.2498	0.2659	0.5279	1.1840	16.7930	7647**	596.9**	82.741**
Industrial production growth rate	0.2101	0.2270	2.0208	0.7663	13.2813	4219**	268.7**	372.7**
Real adj. monetary base growth rate	0.0381	0.1540	2.3031	1.7034	30.7269	30468**	34.722**	79.514**

* denotes 5% significance, ** significance at 1%.

Table 2

Model Selection Results

This table reports statistics used to select multivariate regime switching models of the form

$$\mathbf{y}_t = \boldsymbol{\mu}_{s_t} + \sum_{j=1}^{p} A_{js_t} \mathbf{y}_{t-j} + \Sigma_{s_t} \boldsymbol{\varepsilon}_t,$$

where $\mathbf{y}_t$ includes monthly excess stock and bond returns, as well as 6 prediction variables. The switching models are classified as MSIAH(k,p), where *I*, *A* and *H* refer to state dependence in the intercept, autoregressive terms and heteroskedasticity. *k* is the number of states and *p* is the autoregressive order. Models in the class MSI (*k*, 0)−VAR(*p*) have regime switching in the intercept but not in autoregressive coefficients. The sample period is 1926:12 – 2004:12.

Model [1]	Number of parameters [2]	Log-likelihood [3]	LR test for linearity [4]	AIC [5]	Hannan-Quinn [6]	BIC [7]	LR- test [8]
Base model: MSIA(1,0)							
MSIA (1,0)	44	24429.52	NA	-52.0502	-51.9635	-51.8228	
MSIA (1,1)	108	29708.90	NA	-63.2498	-63.0368	-62.6911	A: 0.000
MSIA (1,2)	172	29871.33	NA	-63.5280	-63.1885	-62.6375	A: 0.000
MSIA (1,3)	236	29945.19	NA	-63.6171	-63.1508	-62.3943	A: 0.000
Base model: MSIA(2,0)							
MSIA (2,0)	54	24819.30	779.57 (0.000)	-52.8608	-52.7544	-52.5818	
MSI (2,0) – VAR(1)	118	29778.58	139.36 (0.000)	-63.3773	-63.1446	-62.7669	A: 0.000
MSIA (2,1)	182	30213.37	1008.96 (0.000)	-64.1696	-63.8107	-63.2282	A: 0.000
MSIH (2,0)	90	26472.98	4086.94 (0.000)	-56.3137	-56.1364	-55.8486	H: 0.000
MSIAH (2,1)	218	31628.38	3838.98 (0.000)	-67.1162	-66.6863	-65.9886	A: 0.000 H: 0.000
MSIH (2,0) – VAR(1)	154	31542.04	3666.29 (0.000)	-67.0685	-66.7647	-66.2719	A: 0.000 H: 0.000
MSIH (2,0) – VAR(2)	218	31654.08	3565.48 (0.000)	-67.2430	-66.8126	-66.1144	A: 0.000
MSIAH (2,2)	346	31799.46	3856.24 (0.000)	-67.2801	-66.5971	-65.4889	A: 0.000
Base model: MSIA(3,0)							
MSIA (3,0)	66	25263.64	1668.24 (0.000)	-53.7836	-53.6536	-53.4425	
MSI (3,0) – VAR(1)	130	29938.81	459.84 (0.000)	-63.6941	-63.4377	-63.0216	A: 0.000
MSIA (3,1)	258	30734.08	2050.35 (0.000)	-65.1198	-64.6110	-63.7853	A: 0.000
MSIH (3,0)	138	26952.11	5045.18 (0.000)	-57.2340	-56.9620	-56.5207	H: 0.000
MSIAH (3,1)	330	32387.91	5358.03 (0.000)	-68.9998	-68.0490	-66.7928	A: 0.000 H: 0.000
MSIH (3,0) – VAR(1)	202	32235.27	5052.74 (0.000)	-68.4472	-68.0488	-67.4023	A: 0.000 H: 0.000
Base model: MSIA(4,0)							
MSIAH (4,1)	444	32771.52	6125.24 (0.000)	**-69.0159**	-68.2002	-66.6792	
MSIH (4,0) – VAR(1)	252	32547.82	5677.83 (0.000)	-69.0081	**-68.5111**	**-67.7046**	

Table 3 – part a

Density Specification Tests for Regime Switching Models

This table reports tests for the transformed z-scores generated by multivariate regime-switching models

$$\mathbf{y}_t = \boldsymbol{\mu}_{s_t} + \sum_{j=1}^{p} A_{js_t} \mathbf{y}_{t-j} + \Sigma_{s_t} \boldsymbol{\varepsilon}_t .$$

The tests are based on the principle that under the null of correct specification, the probability integral transform of the one-step-ahead standardized forecast errors should follow an IID uniform distribution over the interval (0,1). A further Gaussian transform described in Berkowitz (2001) is applied to perform Likelihood ratio tests of the null that (under correct specification) the transformed z-scores, z^*_{t+1}, are IIN(0,1) distributed. In particular, given the transformed z-score model

$$z^*_{t+1} = \alpha + \sum_{j=1}^{q} \sum_{i=1}^{l} \beta_{ij} (z^*_{t+1-i})^j + \sigma u_{t+1} ,$$

the Jarque-Bera statistic tests the hypothesis of normality, LR_2 tests the hypothesis of zero mean and unit variance under the restriction $q = l = 0$; LR_3 tests the joint hypothesis of zero mean, unit variance, and $\rho_{11}= 0$ under $q = l = 1$; LR_6 tests the joint null of zero mean, unit variance, and $\beta_{11}= \beta_{12}= \beta_{21}= \beta_{22}= 0$ with $q = l = 2$. Boldfaced statistics indicate that the null of no misspecification should be rejected.

Model	Number of parameters	Jarque-Bera test	LR_2	LR_3	LR_6
Value-weighted excess stock returns					
Linear (VAR(1))	108	**3412.4 (0.000)**	**110.22 (0.000)**	**113.66 (0.000)**	**121.64 (0.000)**
MSIAH (3,1)	330	**1412.9 (0.000)**	**47.88 (0.000)**	**49.4 (0.000)**	**76.9 (0.000)**
MSIH(4,1) – VAR(1)	252	1.46 (0.483)	5.06 (0.080)	7.10 (0.069)	**13.52 (0.035)**
Excess 10-Year bond returns (term premium)					
Linear (VAR(1))	108	**14.31 (0.001)**	**80.62 (0.000)**	**83.20 (0.000)**	**108.46 (0.000)**
MSIAH (3,1)	330	**90.21 (0.000)**	**61.30 (0.000)**	**77.96 (0.000)**	**93.34 (0.000)**
MSIH(4,1) – VAR(1)	252	**8.96 (0.011)**	5.30 (0.071)	7.74 (0.052)	10.20 (0.116)
12-month cumulated dividend yield					
Linear(VAR(1))	108	**13752 (0.000)**	**202.56 (0.000)**	**204.96 (0.000)**	**260.90 (0.000)**
MSIAH (3,1)	330	**611.7 (0.000)**	**130.90 (0.000)**	**135.16 (0.000)**	**168.98 (0.000)**
MSIH(4,1) – VAR(1)	252	**7.63 (0.022)**	1.84 (0.399)	6.18 (0.103)	10.26 (0.114)
Real 1-month T-bill Yield					
Linear(VAR(1))	108	**2674.6 (0.000)**	**96.34 (0.000)**	**98.52 (0.000)**	**110.10 (0.000)**
MSIAH (3,1)	330	**30.05 (0.000)**	5.10 (0.078)	**15.62 (0.001)**	**44.14 (0.000)**
MSIH(4,1) – VAR(1)	252	**9.90 (0.007)**	**7.22 (0.027)**	**13.16 (0.004)**	**18.52 (0.005)**

Table 3 – part b

Density Specification Tests for Regime Switching Models

Model	Number of parameters	Jarque-Bera test	LR_2	LR_3	LR_6
Default spread					
Linear (VAR(1))	108	**46162 (0.000)**	**304.94 (0.000)**	**308.16 (0.000)**	**313.46 (0.000)**
MSIAH (3,1)	330	**2493.3 (0.000)**	**57.90 (0.000)**	**88.78 (0.000)**	**91.90 (0.000)**
MSIH(4,1) – VAR(1)	252	**19.41 (0.000)**	**9.90 (0.007)**	**30.88 (0.000)**	**56.72 (0.000)**
CPI inflation rate					
Linear (VAR(1))	108	**21.66 (0.000)**	**95.34 (0.000)**	**102.80 (0.000)**	**118.38 (0.000)**
MSIAH (3,1)	330	**120.22 (0.000)**	**23.38 (0.000)**	**54.64 (0.000)**	**72.06 (0.000)**
MSIH(4,1) – VAR(1)	252	1.58 (0.454)	5.04 (0.080)	**9.54 (0.023)**	**13.34 (0.038)**
Real industrial production growth rate					
Linear (VAR(1))	108	**208.11 (0.000)**	**524.32 (0.000)**	**532.12 (0.000)**	**550.96 (0.000)**
MSIAH (3,1)	330	**842.65 (0.000)**	**494.16 (0.000)**	**550.64 (0.000)**	**562.84 (0.000)**
MSIH(4,1) – VAR(1)	252	5.41 (0.067)	5.56 (0.062)	**10.54 (0.014)**	**18.34 (0.005)**
Real adjusted monetary base growth rate					
Linear (VAR(1))	108	**953.52 (0.000)**	**297.92 (0.000)**	**327.22 (0.000)**	**343.62 (0.000)**
MSIAH (3,1)	330	**2382.3 (0.000)**	**149.68 (0.000)**	**310.82 (0.000)**	**482.84 (0.000)**
MSIH(4,1) – VAR(1)	252	3.55 (0.169)	3.68 (0.159)	**10.90 (0.012)**	**25.62 (0.000)**

Table 4 – part a

Estimates of a Four-State Switching Model with Time-Invariant VAR(1) Matrix

Panel A – Single State VAR(1) Model

	Stock	Bond	Div. yield	T-bill	Default	Inflation	Growth	Money
1. Intercept	-0.3878	-0.0376	0.1022***	-0.0425	0.0001	0.0605	0.1356	0.5340*
	(0.6076)	(0.2092)	(0.0331)	(0.0522)	(0.0013)	(0.0518)	(0.1910)	(0.3037)
2. VAR(1) Matrix								
Stock excess returns	0.1046***	0.1108	0.2653*	-0.2514	0.5962	-0.3426	0.0653	0.1257**
	(0.0333)	(0.0960)	(0.1430)	(0.7425)	(3.5335)	(0.7735)	(0.0974)	(0.0586)
Bond excess returns	-0.0260**	0.0626*	-0.0020	0.0522	2.2183*	-0.0485	0.0115	-0.0104
	(0.0115)	(0.0331)	(0.0492)	(0.2557)	(1.2168)	(0.2664)	(0.0335)	(0.0202)
Dividend yield	-0.0060***	-0.0011	0.9835***	-0.0403	-0.3175*	-0.0285	-0.0029	-0.0052*
	(0.0018)	(0.0052)	(0.0078)	(0.0405)	(0.1925)	(0.0421)	(0.0053)	(0.0032)
T-bill real yield	-0.0011	0.0013	-0.0249*	0.5768***	1.2793***	0.1068*	-0.0518***	-0.0027
	(0.0029)	(0.0083)	(0.0123)	(0.0638)	(0.3037)	(0.0605)	(0.0084)	(0.0050)
Default spread	-0.0009***	0.0007***	0.0007**	0.0021	0.9628***	0.0002	-0.0005**	0.0000
	(0.0001)	(0.0002)	(0.0003)	(0.0016)	(0.0077)	(0.0017)	(0.0002)	(0.0001)
Inflation	0.0013	-0.0087	0.0232*	0.3867***	-1.2891***	0.8632***	0.0526***	0.0026
	(0.0028)	(0.0082)	(0.0122)	(0.0633)	(0.3011)	(0.0659)	(0.0083)	(0.0050)
IP real growth	0.0760***	-0.0372	-0.0854*	-0.9677***	3.3244***	-0.9506***	0.4008***	0.0396**
	(0.0105)	(0.0302)	(0.0449)	(0.2334)	(1.1108)	(0.2432)	(0.0306)	(0.0184)
Money real growth	-0.0025	0.0219	-0.2341***	-1.3385***	10.338***	-2.3610***	-0.0316	-0.1880***
	(0.0188)	(0.0543)	(0.0809)	(0.4201)	(1.9992)	(0.4377)	(0.0551)	(0.0332)
3. Correlations/Volatilities								
Stock excess returns	0.0542***							
Bond excess returns	0.1377**	0.0187***						
Dividend yield	-0.8830***	-0.1368**	0.0030**					
T-bill real yield	-0.0333	0.0561	0.0404	0.0047***				
Default spread	-0.2596***	0.0712	0.3324***	0.0371	0.0001*			
Inflation	0.0221	-0.0582	-0.0313	-0.9909***	-0.0298	0.0046***		
IP real growth	0.1128**	-0.0104	-0.1012**	0.4401***	-0.1744**	-0.4434***	0.0170***	
Money real growth	-0.0034	0.0594*	-0.0027	0.2329**	0.0677*	-0.2332**	0.0259	0.0031**

Panel B – Four State Model

	Stock	Bond	Div. yield	T-bill	Default	Inflation	Growth	Money
1. Intercept								
Bull-rebound	0.8945**	0.2530***	0.0382*	-0.2271***	0.0022	0.2321**	0.1633*	-0.9862**
	(0.4420)	(0.1031)	(0.0185)	(0.0781)	(0.0046)	(0.1181)	(0.1090)	(0.3959)
Stable-growth	0.6454**	0.2955***	0.0169	-0.0008	0.0009	0.0225	0.2330*	-0.0108
	(0.2707)	(0.0957)	(0.0173)	(0.0416)	(0.0019)	(0.0418)	(0.1336)	(0.2049)
Expansion-peak	0.5782***	1.0141	0.0408*	0.0015	0.0032	0.0399	-0.0207	0.0478*
	(0.0874)	(0.5851)	(0.0206)	(0.0776)	(0.0019)	(0.0777)	(0.2502)	(0.0261)
Bear-recession	-1.9266***	-0.2598***	0.1301***	0.1575***	0.0089*	-0.1582**	0.4802***	0.4399***
	(0.4820)	(0.0570)	(0.0095)	(0.0191)	(0.0052)	(0.0793)	(0.1830)	(0.0836)
2. VAR(1) Matrix								
Stock excess returns	-0.0009	0.1057	0.0870	-1.3822***	4.7186***	-2.1260***	-0.0433**	0.0821*
	(0.0287)	(0.0719)	(0.1220)	(0.4207)	(0.0805)	(0.7250)	(0.0207)	(0.0450)
Bond excess returns	-0.0341*	0.0167	-0.0401*	-0.9605***	2.9146***	-1.0749***	-0.0079	0.0159
	(0.0173)	(0.0324)	(0.0242)	(0.2579)	(1.0360)	(0.3645)	(0.0204)	(0.0137)
Dividend yield	-0.0003	-0.0034	0.9940***	0.0343	-0.3475***	0.0619*	0.0032	-0.0044*
	(0.0011)	(0.0025)	(0.0044)	(0.0298)	(0.0753)	(0.0302)	(0.0038)	(0.0024)
T-bill real yield	0.0004	0.0045*	0.0011	0.4529***	0.3641**	0.0811	-0.0187	-0.0025
	(0.0025)	(0.0024)	(0.0102)	(0.0651)	(0.1875)	(0.0669)	(0.0092)	(0.0055)
Default spread	-0.0003*	0.0003*	0.0001	0.0034*	0.9640***	0.0031*	0.0000	-0.0000
	(0.0002)	(0.0002)	(0.0004)	(0.0022)	(0.0043)	(0.0020)	(0.0001)	(0.0001)
Inflation	-0.0006	-0.0092	-0.0008	0.4872***	-0.3839**	0.8599***	0.0190	0.0028*
	(0.0024)	(0.0054)	(0.0102)	(0.0655)	(0.1674)	(0.0672)	(0.0142)	(0.0016)
IP real growth	0.0235*	-0.0089	-0.0385	-0.8952	0.9575**	-0.9208***	0.3861***	0.0185
	(0.0120)	(0.0183)	(0.0362)	(0.2186)	(0.4569)	(0.2189)	(0.0326)	(0.0199)
Money real growth	0.0143	0.0233	0.0394	-1.4162***	6.8721***	-2.3985***	-0.0294	-0.1786***
	(0.0118)	(0.0253)	(0.0491)	(0.2968)	(1.9272)	(0.3015)	(0.0464)	(0.0324)

* denotes 10% significance, ** significance at 5%, *** significance at 1%.

Table 4 – part b

Estimates of a Four-State Switching Model with Time-Invariant VAR(1) Matrix

	Panel B – Four State Model							
	Stock	Bond	Div. yield	T-bill	Default	Inflation	Growth	Money
3. Correlations/Volatilities								
Regime 1 (Bull-rebound):								
Stock excess returns	0.0385***							
Bond excess returns	0.1494**	0.0068***						
Dividend yield	-0.8806***	-0.1687***	0.0021***					
T-bill real yield	0.0438	0.0522	-0.0286	0.0071***				
Default spread	-0.1434**	-0.0127	0.1049**	-0.0593	3.9e-05*			
Inflation	-0.0429	-0.0522	0.0270	-0.9990***	0.0589	0.0071***		
IP real growth	-0.0918**	0.0388	0.0678*	0.5054***	-0.1040*	-0.5051***	0.0256***	
Money real growth	0.1742**	0.0220	-0.1864**	0.3201***	-0.0891*	-0.3211***	0.1508**	0.0351***
Regime 2 (Stable-growth):								
Stock excess returns	0.0365***							
Bond excess returns	0.1433**	0.0180***						
Dividend yield	-0.8999***	-0.1807***	0.0011***					
T-bill real yield	0.0784	0.0593	-0.0721	0.0023***				
Default spread	-0.0271	-0.0144	0.0204	0.0989*	3.98e-05*			
Inflation	-0.0917*	-0.0554	0.0884*	-0.9857***	-0.1034**	0.0023***		
IP real growth	0.0953*	-0.0492	-0.1158**	0.5290***	-0.0144	-0.5406***	0.0083***	
Money real growth	0.0539	0.0009	-0.0449	0.3928***	0.0472	-0.3953***	0.2029**	0.0108***
Regime 3 (Expansion-peak):								
Stock excess returns	0.0606***							
Bond excess returns	0.1965***	0.0277***						
Dividend yield	-0.9691***	-0.2080***	0.0026***					
T-bill real yield	0.0063	0.0896	-0.0427	0.0043***				
Default spread	0.0773*	0.3746***	-0.0809**	0.0494	0.0001**			
Inflation	-0.0665	-0.1141*	0.1161**	-0.9535***	-0.0045	0.0041***		
IP real growth	0.1126**	-0.0253	-0.1372**	0.5943***	-0.0540	-0.5789***	0.0125***	
Money real growth	-0.0977**	0.0234	0.0702*	0.3791***	-0.0690	-0.4093***	0.2576**	0.0198***
Regime 4 (Bear-recession):								
Stock excess returns	0.1196***							
Bond excess returns	0.1040*	0.0197***						
Dividend yield	-0.9225***	-0.1873***	0.0082***					
T-bill real yield	-0.1698**	0.1577**	0.1597**	0.0067***				
Default spread	-0.4836***	-0.1435**	0.4864***	0.0744	0.0003***			
Inflation	0.1727**	-0.1605**	-0.1631**	-0.9981***	-0.0797*	0.0067***		
IP real growth	0.3401***	0.0683	-0.2432***	0.1544**	-0.4347***	-0.1601**	0.0317***	
Money real growth	-0.0456	0.2287***	0.0456	0.0262	0.1042*	-0.0180	-0.1768**	0.0769***

4. Transition probabilities	Bull-rebound	Stable-peak	Expansion	Bear-recession
Bull-rebound	0.8939***	0.0237*	0.0207	0.0616
	(0.1560)	(0.0128)	(0.0249)	
Stable-growth	0.0100	0.9311***	0.0523**	0.0066
	(0.0349)	(0.2305)	(0.2410)	
Expansion-peak	0.0230	0.1517***	0.8030***	0.0381
	(0.2340)	(0.0555)	(0.1159)	
Bear-recession	0.1683***	0.0340*	0.0381**	0.7596
	(0.0200)	(0.0183)	(0.0160)	

* denotes 10% significance, ** significance at 5%, *** significance at 1%.

Table 5

Implied Monthly Means from a Four-State Switching Model

This table reports estimates for a single state and a four-state VAR(1) regime switching model (MSIH(4,0)-VAR(1)):

$$\mathbf{y}_t = \boldsymbol{\mu}^*_{s_t} + \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\Sigma}^*_{s_t}\,\boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t \sim$ I.I.D. $N(0, \mathbf{I}_8)$ is an unpredictable return innovation. The sample period is 1926:12 – 2004:12.

	Stock	Bond	Div. yield	T-bill	Default	Inflation	Growth	Money
	Panel A – Data							
Overall mean	0.6482	0.1447	3.8132	0.0540	0.0943	0.2498	0.2101	0.0381
Mean 1926-1946	0.7174	0.2699	4.9296	0.0100	0.1472	0.0621	0.1767	0.6411
Mean 1947-1966	0.8997	-0.0143	4.2120	0.0029	0.0520	0.1805	0.2503	0.1248
Mean 1967-1986	0.3232	0.0775	3.7191	0.0924	0.1029	0.5080	-0.8127	-0.4029
Mean 1987-2004	0.6528	0.2563	2.2291	0.1170	0.0728	0.2494	0.2616	0.2510
	Panel B – Single state VAR(1) Model							
Overall mean	0.6457	0.1424	3.8097	0.0538	0.0933	0.2521	0.2042	0.0376
Unconditional mean	0.5683	0.0904	3.9713	0.0398	0.0730	0.3432	0.4883	0.4001
Mean 1926-1946	1.1772	0.2696	4.8931	0.0287	0.1451	0.0497	0.2823	0.6247
Mean 1947-1966	0.7784	0.0339	4.2129	-0.0632	0.0521	0.2529	0.1367	-0.3354
Mean 1967-1986	0.4009	0.1527	3.7089	0.1492	0.1019	0.4427	-0.7283	-0.4934
Mean 1987-2004	0.1796	0.1101	2.2623	0.1058	0.0720	0.2644	-0.1372	0.6391
	Panel C – Four State Model							
Overall mean*	0.6302	0.1446	3.8132	0.0475	0.0943	0.2547	0.2110	0.0428
Mean 1926-1946*	1.0211	0.2878	4.9202	-0.0239	0.1469	0.0912	0.2134	0.5078
Mean 1947-1966*	0.9044	0.0612	4.2089	-0.0250	0.0521	0.2177	0.0375	0.1759
Mean 1967-1986*	0.1822	0.1544	3.7254	0.1347	0.1029	0.4541	-0.7948	-0.3829
Mean 1987-2004*	0.5890	0.0670	2.2413	0.1107	0.0731	0.2561	0.3266	0.1286
Unconditional mean	0.6137	0.1569	3.3521	0.0542	0.0890	0.2521	0.1773	0.0583
Regime 1 – unc. mean	1.2129	0.0607	5.2311	-0.2982	0.0590	0.3832	0.0679	-0.6800
Regime 2 – unc. mean	0.4514	-0.0522	2.8230	0.0889	0.0611	0.2733	0.2394	0.2254
Regime 3 – unc. mean	0.4903	0.7252	2.5103	0.1992	0.1503	0.3964	1.2231	0.0098
Regime 4 – unc. mean	0.2066	0.5220	2.1770	0.3442	0.2482	-0.4500	-0.8355	2.4221

* Based on smoothed probabilities.

Table 6

Out-of-Sample, Recursive Predictive Performance

The table reports the root-mean-square forecast error, the predictive bias, and the forecast error variance for four models: a four-state regime switching model with constant VAR coefficients (MSIH(4,0)-VAR(1)), a three-state switching model with regime-specific VAR coefficients (MSIAH(4,1)), a single-state Gaussian VAR(1), and a Neely and Weller's (2000) benchmarks (recursive sample mean for excess returns and random walk otherwise). The (pseudo) out-of sample period is 1985:01 – 2004:11. The models are recursively estimated on expanding windows 1926:12 – 1985:01, 1926:12 – 1985:02, up to 1926:12 – 2004:11. The three panels consider 1-, 4-, 12-month forecasts, respectively.

		Stock	Bond	Div. yield	T-bill	Default	Inflation	Growth	Money
One-month horizon		**Four-state MSIH(4,0)-VAR(1)**							
	Root-MSFE	3.773	1.285	1.443	0.880	0.152	0.277	1.385	2.421
	Bias	0.369	0.218	-1.428	-0.128	0.113	-0.224	0.359	0.561
	St. dev.	3.755	1.266	0.208	0.871	0.102	0.163	1.338	2.355
		Three-state MSIAH(4,1)							
	Root-MSFE	4.310	1.287	3.120	0.888	0.115	0.927	1.422	5.570
	Bias	0.883	0.287	2.533	-0.145	0.071	-0.167	0.318	0.266
	St. dev.	4.219	1.255	1.821	0.876	0.090	0.912	1.386	5.564
		Single-state VAR(1)							
	Root-MSFE	5.256	1.257	1.580	0.900	0.115	0.253	2.442	2.964
	Bias	0.631	0.227	-1.333	-0.124	0.071	-0.080	0.408	0.282
	St. dev.	5.218	1.236	0.848	0.891	0.090	0.240	2.408	2.951
		Neely and Weller's benchmarks							
	Root-MSFE	6.138	1.238	3.163	0.891	0.115	0.923	3.438	5.881
	Bias	0.036	0.146	2.578	-0.144	0.071	-0.169	0.456	0.599
	St. dev.	6.138	1.229	1.832	0.879	0.091	0.907	3.408	5.850
Four-month horizon		**Four-state MSIH(4,0)-VAR(1)**							
	Root-MSFE	4.189	1.259	2.874	0.881	0.114	0.930	3.411	4.330
	Bias	0.426	0.201	2.318	-0.116	0.071	-0.198	0.300	0.627
	St. dev.	4.167	1.243	1.699	0.873	0.090	0.909	3.398	4.284
		Three-state MSIAH(4,1)							
	Root-MSFE	5.942	1.282	2.988	0.876	0.113	0.900	3.380	5.338
	Bias	0.787	0.322	2.383	-0.187	0.071	-0.134	0.167	0.369
	St. dev.	5.890	1.241	1.802	0.855	0.088	0.900	3.376	5.325
		Single-state VAR(1)							
	Root-MSFE	6.245	1.946	3.261	1.479	0.113	1.258	4.743	5.868
	Bias	-1.483	1.402	2.668	-1.225	0.068	0.951	-3.348	-2.429
	St. dev.	6.251	1.350	1.875	0.830	0.090	0.823	3.360	5.341
		Neely and Weller's benchmarks							
	Root-MSFE	8.187	1.236	3.164	0.876	0.116	0.910	3.473	5.770
	Bias	0.026	0.135	2.572	-0.161	0.072	-0.158	0.433	0.596
	St. dev.	8.187	1.229	1.842	0.860	0.091	0.897	3.446	5.739
One-year horizon		**Four-state MSIH(4,0)-VAR(1)**							
	Root-MSFE	5.305	1.280	2.964	0.884	0.111	0.913	3.449	4.411
	Bias	0.383	0.222	2.371	-0.102	0.071	-0.213	0.339	0.647
	St. dev.	5.291	1.260	1.779	0.878	0.086	0.888	3.432	4.363
		Three-state MSIAH(4,1)							
	Root-MSFE	6.325	1.315	3.034	0.880	0.106	0.853	3.396	5.367
	Bias	0.227	0.392	2.040	-0.252	0.067	-0.068	0.074	3.255
	St. dev.	6.321	1.255	2.246	0.843	0.082	0.851	3.395	5.357
		Single-state VAR(1)							
	Root-MSFE	7.376	1.818	3.694	1.294	0.120	1.128	4.553	5.765
	Bias	-1.122	1.264	3.145	-0.976	0.077	0.745	-3.043	-2.118
	St. dev.	7.290	1.307	1.938	0.850	0.092	0.847	3.387	5.362
		Neely and Weller's benchmarks							
	Root-MSFE	8.288	1.253	3.168	0.898	0.117	0.924	3.476	6.130
	Bias	-0.039	0.137	2.559	-0.185	0.074	-0.151	0.466	0.574
	St. dev.	8.288	1.245	1.867	0.879	0.091	0.911	3.445	6.103

Table 7

Predictive Accuracy Tests

The table reports differential predictive accuracy tests of the four-state MSIH(4,0)-VAR(1) model vs. three alternative models: a three-state switching model with regime-specific VAR coefficients (MSIAH(3,1)), a single-state Gaussian VAR(1), and a Neely and Weller's (2000) benchmarks (recursive sample mean for excess returns and random walk otherwise). The (pseudo) out-of sample period is 1985:01 – 2004:11. The models are recursively estimated on expanding windows 1926:12 – 1985:01, 1926:12 – 1985:02, up to 1926:12 – 2004:11. DM tests are unconditional in nature, while GW tests are conditional and employ simple test functions [1 $diff_{t-1}$]. p-values are computed using HAC standard errors in both cases. Negative values indicate that the model in the table's header has higher predictive accuracy than the benchmark in the flank; an '*' has the same meaning in Giacomini-White's (2004) tests.

Benchmark		Stock	Bond	Div. yield	T-bill	Default	Inflation	Growth	Money
Forecast Horizon: 1 month									
		Four-state MSIH(4,0)-VAR(1)							
Three-state MSIAH(3,1)	**DM test**	-2.255	-0.125	-5.314	-1.161	0.287	-0.209	-0.238	-3.761
		(0.024)	(0.901)	(0.000)	(0.246)	(0.774)	(0.835)	(0.812)	(0.000)
	GW test	5.340*	3.188*	21.905*	1.488*	0.000	3.878*	0.174*	3.195*
		(0.069)	(0.203)	(0.000)	(0.475)	(0.999)	(0.144)	(0.917)	(0.202)
		Neely and Weller's benchmarks							
	DM test	0.998	-1.967	4.941	0.274	2.347	-0.375	0.158	1.308
		(0.319)	(0.049)	(0.000)	(0.784)	(0.019)	(0.708)	(0.875)	(0.191)
	GW test	4.562	18.685*	33.248	8.150	0.000	7.084*	2.985	1.853
		(0.102)	(0.000)	(0.000)	(0.017)	(0.999)	(0.029)	(0.225)	(0.396)
		Four-state MSIH(4,0)-VAR(1)							
Neely-Weller	**DM test**	-1.817	4.294	-3.018	-0.950	1.775	-0.326	-1.193	-3.536
		(0.069)	(0.000)	(0.003)	(0.342)	(0.076)	(0.745)	(0.233)	(0.000)
	GW test	1.689*	4.767	10.522*	7.645*	0.001	6.629*	3.180*	10.449*
		(0.430)	(0.092)	(0.005)	(0.022)	(0.999)	(0.036)	(0.204)	(0.005)
Forecast Horizon: 1 year									
		Four-state MSIH(4,0)-VAR(1)							
Three-state MSIAH(3,1)	**DM test**	-0.637	-3.586	-4.800	-0.517	3.452	2.896	2.356	-2.008
		(0.524)	(0.000)	(0.000)	(0.605)	(0.001)	(0.004)	(0.018)	(0.047)
	GW test	1.567*	5.175*	6.707*	5.257*	6.293	5.158	5.199	1.274*
		(0.457)	(0.075)	(0.035)	(0.072)	(0.043)	(0.076)	(0.074)	(0.529)
		Neely and Weller's benchmarks							
	DM test	1.509	-4.854	5.229	1.887	4.929	3.511	1.439	2.722
		(0.131)	(0.000))	(0.000)	(0.059)	(0.000)	(0.001)	(0.150)	(0.006)
	GW test	2.589	6.957*	8.363	5.257	7.369	5.807	6.906	5.089
		(0.274)	(0.031)	(0.015)	(0.072)	(0.025)	(0.055)	(0.032)	(0.079)
		Four-state MSIH(4,0)-VAR(1)							
Neely-Weller	**DM test**	-0.836	3.850	-3.678	-1.934	-4.799	-0.066	-0.470	-2.464
		(0.403)	(0.000)	(0.000)	(0.053)	(0.000)	(0.947)	(0.640)	(0.014)
	GW test	0.359	5.854	5.440*	4.341*	7.046*	3.467*	4.575*	4.355*
		(0.836)	(0.054)	(0.066)	(0.114)	(0.030)	(0.177)	(0.101)	(0.113)

Figure 1

Distribution of Transformed (Generalized) z-Scores from Three-State VAR(1) Switching Model

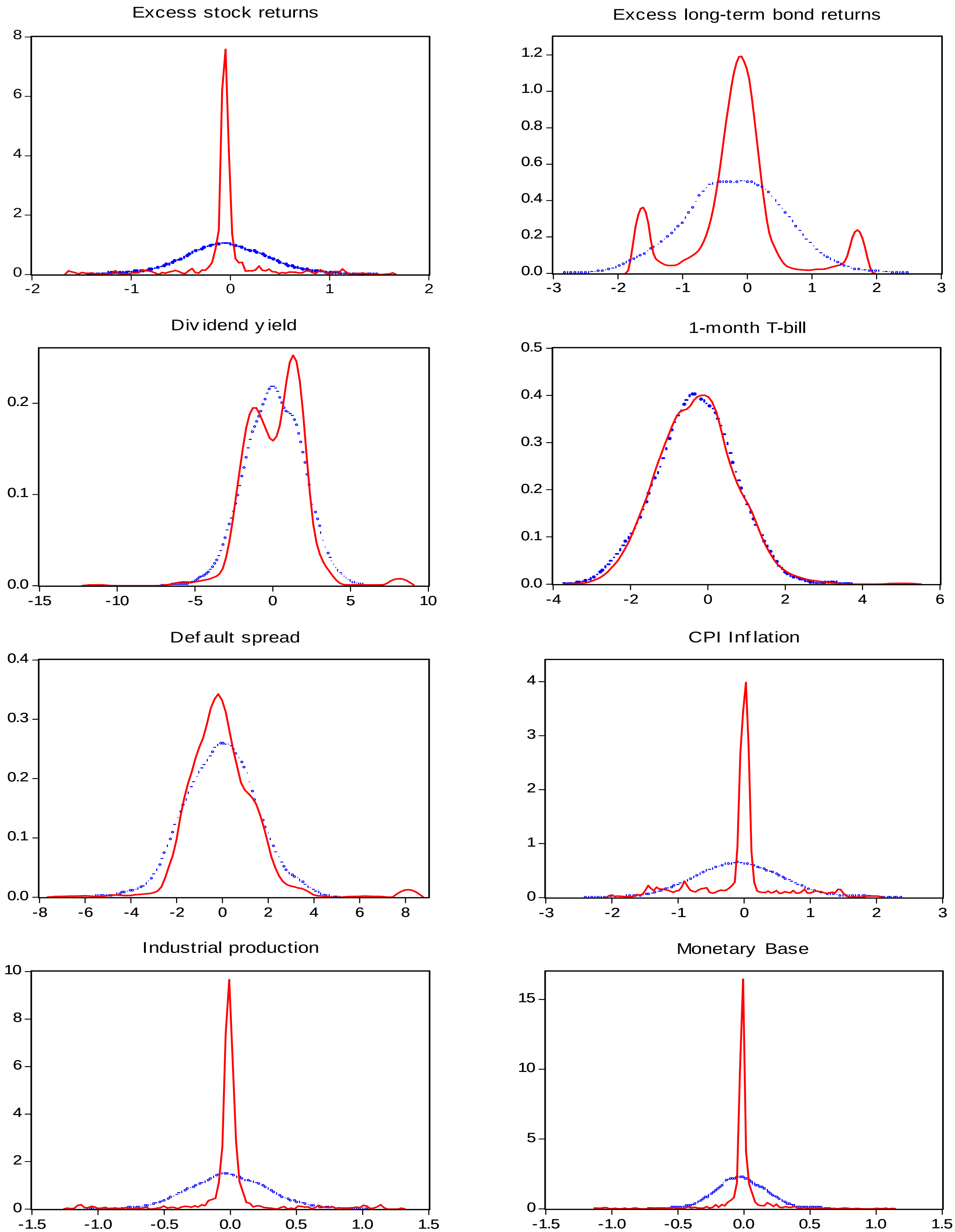

Figure 2

Distribution of Transformed (Generalized) z-Scores from Four-State Model with Time-Invariant VAR(1) Matrix

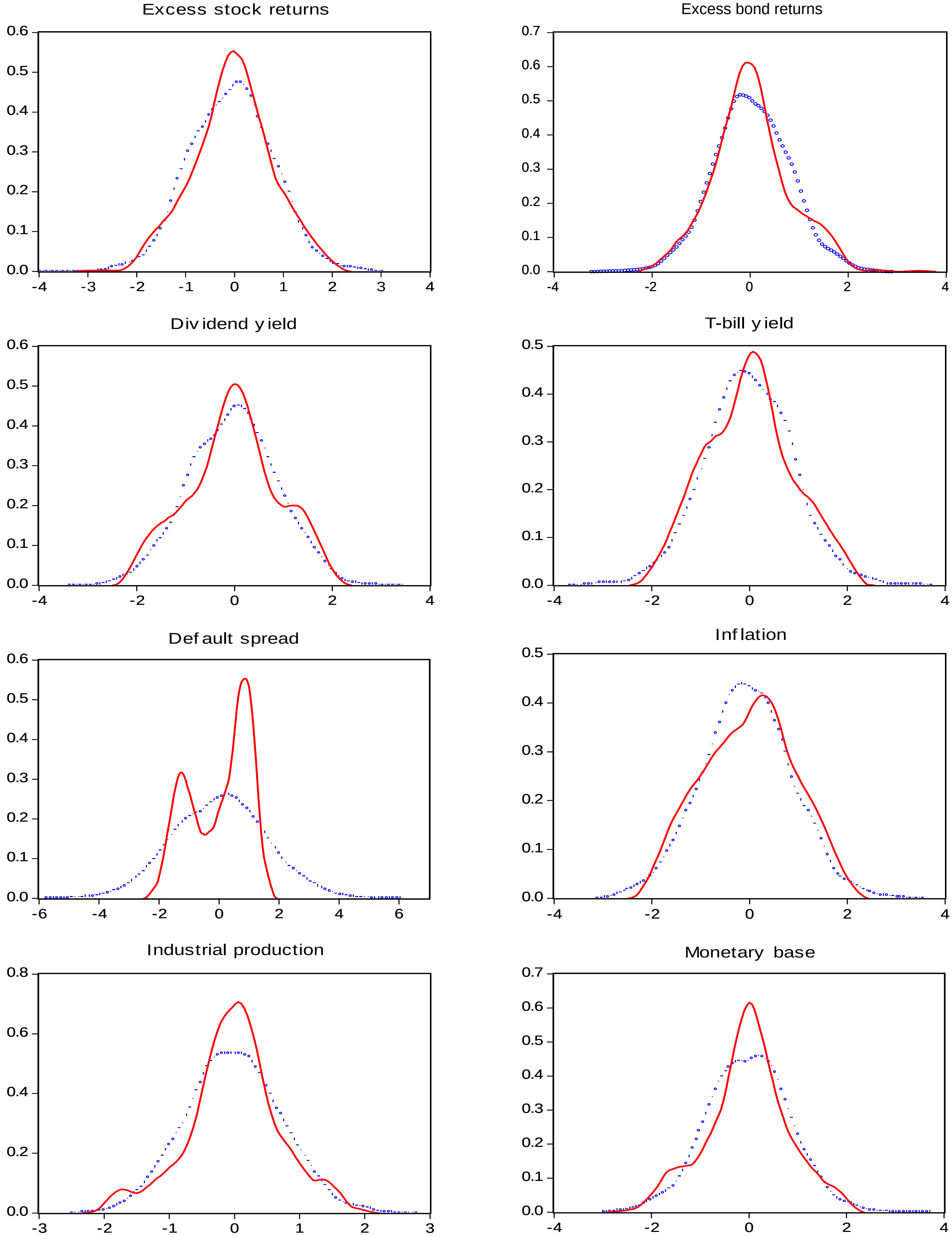

Figure 3

Smoothed Probabilities from Four-State Model

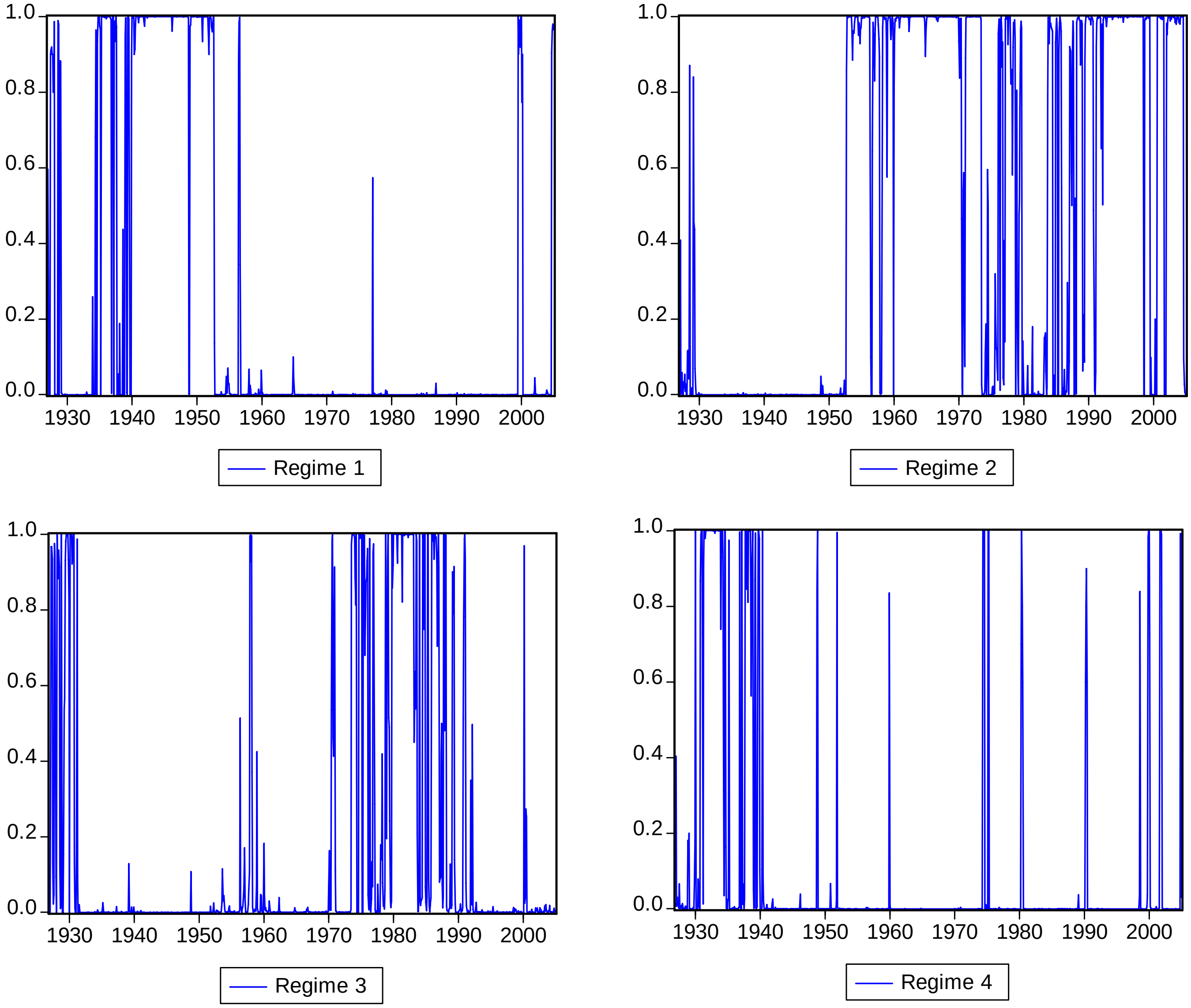

Figure 4
Implied Monthly Predicted Sharpe Ratios from a Four-State Model