NBER WORKING PAPER SERIES

RARE EVENTS, FINANCIAL CRISES, AND THE CROSS-SECTION OF ASSET RETURNS

Francesco Bianchi

Working Paper 21056
http://www.nber.org/papers/w21056

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2015

I am grateful to Chris Sims for his useful suggestions at the early stage of this work. I would like to thank Robert Barro, Markus Brunnermeier, John Campbell, Emmanuel Farhi, Jakub Jurek, Ralph Koijen, Howard Kung, Alisdair McKay, Kristoffer Nimark, Barbara Rossi, Motohiro Yogo, Adam Zawadowski, and seminar participants at the Annual Asset Pricing Retreat, Princeton University, Duke University, the Board of Governors of the FRS, the EEA-ESEM conference, the Bank of Italy, the University of Pisa, and the CEF Conference for helpful discussions and comments. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Rare Events, Financial Crises, and the Cross-Section of Asset Returns
Francesco Bianchi
NBER Working Paper No. 21056
March 2015
JEL No. C32,G01,G12

ABSTRACT

Similarities between the Great Depression and the Great Recession are documented with respect to the behavior of financial markets. A Great Depression regime is identified by using a Markov-switching VAR. The probability of this regime has remained close to zero for many decades, but spiked for a short period during the most recent financial crisis, the Great Recession. The Great Depression regime implies a collapse of the stock market, with small-growth stocks outperforming small-value stocks. This helps to explain the cross section of asset returns when risk is priced according to a version of the "Bad Beta, Good Beta" Intertemporal CAPM that allows for regime changes.

Francesco Bianchi
Department of Economics
Duke University
213 Social Sciences Building
Box 90097
Durham, NC 27708
and Cornell University, CEPR, and NBER
francesco.bianchi@duke.edu

1 Introduction

The recent financial crisis had pervasive consequences, leading the U.S. economy to its longest and most severe recession since World War II. Arguably, the crisis started with the end of a housing bubble that, in turn, led to the collapse of the subprime market. From there, in the span of a few months, it spread to the whole banking sector and then to the real economy. The decline in real economic activity accelerated in the fall of 2008 as the financial crisis unfolded. U.S. gross domestic product fell by 5% in an year, while the unemployment rate increased from less than 5% to 10%. The large contraction in real activity came with an equally dramatic decline in stock prices, with the S&P 500 index dropping by almost 57% from its October 2007 peak of $1,565$ to a mere 676.5 in March 2009. The possibility of a complete financial meltdown suddenly became a real concern and commentators and policymakers alike feared that the economy could be heading toward a second Great Depression (Krugman, 2009).

The Great Depression began with the devastating market crash, which occurred on October 29, 1929. It is difficult to establish whether the stock market crash was a cause or a consequence of the Great Depression. However, there is no doubt that the two events were closely related. In fact, the Great Depression can be considered an extreme example of how a financial crisis and a recession can negatively affect each other. Financial markets and economic institutions have evolved significantly since then, but the fear of a return to the Great Depression suggests that rare events might have a long lasting impact on the way agents think about the economy. In this paper, I document a series of similarities between these two rare events by examining the behavior of financial markets. I then show that these similarities are important to understand the cross section of asset returns.

In order to formally assess to what extent the Great Recession mirrored the Great Depression, I first estimate a Markov-switching vector autoregression (MS-VAR) that allows for both changes in the VAR coefficients and in the covariance matrix that characterizes the contemporaneous relations and volatilities of the disturbances. I include four key financial variables: the excess market return, the Term Yield spread, the Price Earnings ratio, and the Value spread. The excess market return captures the performance of the stock market with respect to a risk-free rate. The Term Yield spread measures the slope of the term structure of interest rates that in turn has predictive power for future real activity. The Price Earnings ratio can be considered a measure of market imbalance as it tends to be negatively correlated with future returns. Finally, the Value spread measures the difference between the log book-to-market ratios of small value and small growth stocks. Given that this last variable moves up when small growth stocks perform relatively better, it can be considered a proxy for the behavior of the cross section of asset returns.

A *Great Depression regime* emerges from the estimates. A central feature of this regime is that it implies a large collapse of the stock market with a contemporaneous large increase in the Value spread, suggesting that growth stocks perform relatively better than value stocks during financial crises. As implied by its name, this regime characterized the behavior of the stock market during the Great Depression, when the Price Earnings ratio and the Value spread touched the historical minimum and maximum, respectively. For the remainder of the sample, its probability has been close to zero until the early months of 2009. Therefore, the Great Recession shows a resurgence of this regime, even if for only two months. The probability of the Great Depression regime crossed the threshold of 50% in February 2009 for the first time since November 1948. However, it quickly returned to zero in March, arguably because of government interventions that were effective in preventing a financial meltdown and led to a reversal in the behavior of the stock market and the Value spread.

In order to reinforce this point, I use counterfactual simulations to show that since the starting of the Great Recession in mid-2008 until February 2009, financial markets were on a path consistent with the Great Depression regime: a persistent fall in the stock market paired with a contemporaneous increase in the Value spread. This opposite moving relation between the behavior of the stock market as a whole and the relative performance of growth stocks was absent during another important market decline: The end of the Information Technology (IT) bubble. In that case, the Value spread and the Price Earnings ratios were moving together. This suggests that market declines that are associated with financial crises might be inherently different and that monitoring the relative performance of growth and value portfolios during these events might be useful in understanding where markets are headed.

The similarities between the Great Depression and the Great Recession extend beyond the level dynamics that are implied by the VAR coefficients. Even the innovations present some interesting features. First, both periods were characterized by high volatility. More interestingly, both during the stock market crash that opened the Great Depression and the fall in the stock market that characterized the beginning of the Great Recession, shocks to market returns and the Value spread were negatively correlated. This has an important implication for asset pricing because it implies that during crises, innovations to the relative return of growth stocks move in an opposite direction with respect to stock market returns.

Given that the estimates point toward the existence of an interesting link between major financial crises and the cross section of asset returns, I devote the second part of the paper to a detailed analysis of the implications of the Great Depression and the Great Recession for the cross section of asset returns. I reconsider the *Bad Beta, Good Beta* Intertemporal CAPM (ICAPM) proposed by Campbell and Vuolteenaho (2004). The model is based on the idea that unexpected excess returns can be decomposed into news about future cash flows and news about future discount rates. Accordingly, the usual CAPM beta can be decomposed into two betas, one for each of the two types of news. The economically motivated

ICAPM predicts that the price of risk for the discount-rate beta should equal the variance of unexpected market returns, while the price of risk for the cash-flow beta should be γ times greater, where γ is the investor's coefficient of relative risk aversion.

As a first step, the VAR methodology used to derive the news is extended in order to reflect the possibility of regime changes. The ICAPM is then tested over different subsamples to highlight the importance of the two financial crises. Specifically, I use moving windows of 35 years starting from the late 1920s until the recent crisis. The results provide support for the idea that rare events play an important role in explaining the cross section of asset returns. During the early years of the sample, consistent with the results of Campbell and Vuolteenaho, the ICAPM performs well in explaining the 25 Fama-French portfolios sorted with respect to size and book-to-market ratios, but so does the traditional CAPM. However, as the data window moves away from the Great Depression, the explanatory power of ICAPM starts to slowly decline. By the 1990s, its R^2 starts moving around 30%, well below the 60% attained during the first half of the sample. However, as the window approaches the most recent financial crisis, the explanatory power of the ICAPM increases steeply, and the R^2 touches 60%, a value that was last reached at the end of 1978.

In order to highlight why the Great Recession plays such an important role in improving the fit of the ICAPM, I show that the return of medium size growth stocks was visibly lower than the expected return implied by the ICAPM during the 1980s and 1990s. In other words, the return on these stocks was too low in light of a general increase in their risk level as captured by their discount rate and cash-flow betas. Symmetrically, returns on value stocks were quite high with respect to what was predicted by the ICAPM. In both cases, the anomalies were largely corrected during the Great Recession. This result suggests that the relative performance of these two classes of stocks during regular times might be compensated by their behavior during rare events. This is why the ICAPM is able to explain remarkably well the cross section of asset returns when financial crises are included in the sample. Barro (2006, 2009), following Rietz (1988), shows that rare disasters are potentially important in explaining the equity premium puzzle. The results presented here imply that rare events also play an important role for the cross section asset returns.

Furthermore, financial crises are also important in shaping agents' expectations. This conclusion can be inferred by comparing the explanatory power of the ICAPM under the benchmark case, in which *fully rational* agents form expectations taking into account the possibility of regime changes, with an alternative scenario in which agents form expectations disregarding the possibility of regime changes. This latter case corresponds to the case of anticipated utility: at each point in time, agents assume that the probabilities of the two regimes will not change in the future. I show that the benchmark case delivers substantially better results with an increase in the R^2 that fluctuates around 20%. Therefore, rare events are not only important because of their effect on realized returns, but also because of the way they shape agents' expectations. In other words, even in a sample in which no rare events have occurred, assets are priced to reflect the possibility of rare events.

As a methodological contribution, this paper proposes a simple algorithm to estimate a Markov-switching VAR in reduced form with Bayesian methods.[1] Practitioners that are more familiar with frequentist econometrics can make use of the results presented here and use a maximum likelihood approach. An MS-VAR allows for an analytical characterization of the news along the lines of the VAR approach proposed by Campbell (1991) to implement the present value decomposition of Campbell and Shiller (1988). The formulas presented in the paper are specific for the model of Campbell and Vuolteenaho (2004), but they can be easily modified to handle other models that make use of a present value decomposition to allow for the possibility of structural breaks. This approach, which formally isolates periods characterized by unusual dynamics, might also prove useful in explaining why the present value decomposition methodology is often sensitive to the sample choice. Furthermore, the Markov-switching extension can easily accommodate temporary non-stationary regimes as long as the system as a whole is stable, a feature particularly convenient when working with financial data.

The content of this paper can be summarized as follows. Section 2 presents a review of the related literature. In Section 3, I present the MS-VAR used to assess the presence of similarities between the Great Depression and the Great Recession. Section 4 reports the results for the MS-VAR estimates. Section 5 presents the implications of the two rare events for the cross section of asset returns. In Section 6, I conduct a robustness check exercise using real time estimates. In Section 7, I conclude.

2 Related Literature

This paper is related to several important contributions in macroeconomic and finance literature. Cogley and Sargent (2007) posit that agents update their beliefs according to Bayes' Law, but also that some rare events can arrest convergence to a rational expectations equilibrium, thereby initializing a new learning process. They argue that the Great Depression was one such event, and they use this to explain the high but declining equity premium. In this paper, I argue that the Great Depression is also important for the cross section of asset returns. However, I do this without relying on a deviation from rational expectations, but by showing that the Great Recession presented important similarities with the Great Depression.

The literature on the cross section of asset returns is very vast. Bansal, Dittmar, and Lundblad (2005), in the spirit of Bansal and Yaron (2004), focus on the aggregate consump-

[1] Sims and Zha (2006) provide methods to estimate a MS-VAR casted in its structural form.

tion risks embodied in cash flows. Using a neoclassical framework, Zhang (2005) shows that the value anomaly arises naturally due to costly reversibility and the countercyclical price of risk. Koijen, Lustig, and Nieuwerburgh (2014) provide evidence in favor of the idea that the business cycle is a priced state variable in stock markets. Campbell, Giglio, and Polk (2013) highlight that the 2007-2009 market fall was not offset by improving stock return forecasts as in the stock market downturn of 2000-2002, while Campbell, Giglio, Polk, and Turley (2014) extend the approximate closed-form intertemporal capital asset pricing model of Campbell (1993) to allow for stochastic volatility. They estimate the effects of volatility on the stock market by using a two-step procedure that allows them to include volatility of all shocks in the VAR. With respect to their work, I do not impose the restriction that all volatilities have to move in parallel, I allow the covariance structure of the disturbances to vary over time, and I model the possibility of regime changes in the VAR coefficients and, consequently, in the way agents map shocks into the news about future discount rates and future cash flows. To the best of my knowledge, this feature is new in the literature. Given that the primary interest of this paper is to assess the role of the Great Recession and the Great Depression, I do not price volatility when studying the cross section of asset returns. The possibility of merging the two approaches is an interesting path for future research.

This paper is related to the growing rare disasters literature. Barro (2006, 2009), following Rietz (1988), shows that rare disasters are potentially important in explaining several asset-pricing puzzles. In fact, Nakamura, Steinsson, Barro, and Ursua (2013), estimating an empirical model of consumption disaster, show that under Epstein-Zin-Weil preferences, rare disasters can rationalize a sizeable equity premium for modest values of risk aversion. Martin (2013a) studies a Lucas Orchard to point out that disasters can spread across assets, generating large risk premia even for assets with stable fundamentals, while Martin (2013b) shows how to use observable asset prices to make inferences that are robust to the details of the underlying consumption process. Wachter (2013) shows that by allowing for a time-varying probability of rare disasters, the model can explain at the same time the equity premium puzzle and the high stock market volatility. Gabaix (2012) extends Barro's results by allowing for a time-varying intensity of rare disasters to propose a solution to a series of finance puzzles. Bollerslev and Todorov (2011) show that the compensation for rare events accounts for a large fraction of the average equity and variance risk premia by using high-frequency data. Gourio (2012) embeds disasters into a production economy to jointly explain the equity premium and business cycles. Bai, Hou, Kung, and Zhang (2015) include disasters in an investment-based asset pricing model to argue value stocks are more exposed to disaster risk than growth stocks.

Chen, Joslin, and Tran (2012) introduce heterogenous beliefs and argue that disagreements generate strong risk-sharing motives, reducing the disaster risk premium. Julliard and Ghosh (2012) argue that rare events cannot account for the equity premium puzzle and the cross section of asset returns at the same time because an increase in the probability of rare events leads to a reduction in the cross-sectional dispersion of consumption risk compared to the cross-sectional variation of average returns. Backus, Chernov, and Martin (2011) use equity index options to quantify the distribution of consumption growth disasters and find that options imply smaller probabilities of extreme outcomes than have been estimated from macroeconomic data. The results of this paper suggest that it might be possible to reconcile these opposing views by using an ICAPM.

The idea that rare events are important for the cross section of asset returns and to shape agents' expectation mechanism should be distinguished from the so called *peso phenomenon* that addresses agents' expectations of an economy wide disaster that has never materialized in the sample (Sandroni, 1998, Veronesi, 2004). Instead, the argument that I propose is that during rare events, assets might behave differently. As a result, including these rare events in the estimates is important to correctly judge their relative performance and to adequately model agents' expectation mechanism.

Markov-switching models are quite popular in financial econometrics. Lettau, Ludvigson, and Wachter (2008) estimate an univariate Markov-switching process to study the link between consumption risk and the decline in the equity premium. Bianchi, Lettau, and Ludvigson (2013) use a Markov-switching model to document infrequent shifts in the mean of the consumption-wealth variable cay_t (Lettau and Ludvigson, 2001). Ang and Bekaert (2002) examine the econometric performance of regime-switching models for interest rate data, while Pesaran, Pettenuzzo, and Timmermann (2006) use MS models in forecasting financial series. Gulen, Xing, and Zhang (2011) use an univariate MS model of stock returns to study the time-varying nature of the value premium. With respect to these contributions, I use a multivariate model with two separate processes controlling the VAR coefficients and the volatilities, while the literature often utilizes univariate processes in which a single chain controls all parameters of the model. Allowing for two separate chains is important because volatility changes would otherwise tend to dominate other regime breaks as pointed out by Ang and Timmermann (2012) and Sims and Zha (2006).

Lewellen, Nagel, and Shanken (2010) and Daniel and Titman (2006) have highlighted some drawbacks of the empirical methods used to test factor-model explanations of market anomalies. The results of this paper are robust with respect to their concerns given that the ICAPM imposes economically motivated restrictions on the premia. There are also some caveats about the VAR methodology used to retrieve cash-flow and discount rate news. Chen and Zhao (2005) argue that it is potentially misleading to obtain the two series with the discount-rate news being directly modeled and the cash-flow news calculated as the residual. Campbell, Polk, and Vuolteenaho (2010) and Engsted, Pedersen, and Tanggaard (2010) clarify the conditions under which VAR results are robust to the decision whether to forecast returns or cash flows.

3 The Model

In this section, I present the MS-VAR use to study the similarities between the Great Depression and the Great Recession.

3.1 A Markov-switching VAR

The variables of interest are assumed to evolve according to a Markov-switching VAR with one lag:

$$Z_t = c_{\xi_t^\Phi} + A_{\xi_t^\Phi} Z_{t-1} + \Sigma_{\xi_t^\Sigma}^{1/2} \omega_t \tag{1}$$

$$\Phi_{\xi_t^\Phi} = \left[c_{\xi_t^\Phi}, A_{\xi_t^\Phi}\right], \ \omega_t \sim N(0, I) \tag{2}$$

where Z_t is a $(n \times 1)$ vector of data. The unobserved states ξ_t^Σ and ξ_t^Φ can take on a finite number of values, $j^\Phi = 1, \ldots, m^\Phi$ and $j^\Sigma = 1, \ldots, m^\Sigma$, and follow two independent Markov chains.[2] This represents a convenient way to model heteroskedasticity and to allow for the possibility of changes in the dynamics of the state variables.[3] The probability of moving from one state to another is given by $P[\xi_t^\Phi = i|\xi_{t-1}^\Phi = j] = h_{ij}^\Phi$ and $P[\xi_t^\Sigma = i|\xi_{t-1}^\Sigma = j] = h_{ij}^\Sigma$.

Given $H^\Phi = [h_{ij}^\Phi]$ and $H^\Sigma = [h_{ij}^\Sigma]$ and a prior distribution for the initial state, we can compute the likelihood of the parameters of the model, conditional on the initial observation Z_0. The likelihood can then be combined with a prior probability for the parameters of the model to obtain their posterior probability. A by-product of the likelihood calculation are the filtered probabilities for Markov-switching states: $\pi_{t|t}^\Phi$ and $\pi_{t|t}^\Sigma$, where each element of the two vectors is defined by $\pi_{t|t}^{\Phi,i} = P[\xi_t^\Phi = i|Z^t, \Phi_{\xi_t^\Phi}, \Sigma_{\xi_t^\Sigma}, H^\Phi, H^\Sigma]$ and $\pi_{t|t}^{\Sigma,i} = P[\xi_t^\Sigma = i|Z^t, \Phi_{\xi_t^\Phi}, \Sigma_{\xi_t^\Sigma}, H^\Phi, H^\Sigma]$ for all i at each t where $Z^t = \{Z_s\}_{s=1}^t$. Therefore, the filtered estimates represent the probabilities assigned to the different regimes conditional on the model parameters and the data up to time t. These can be converted by a recursive algorithm to smoothed estimates: $\pi_{t|T}^\Phi$ and $\pi_{t|T}^\Sigma$, where each element of the two vectors is given by $P[\xi_t^\Phi = i|Z^T, \Phi_{\xi_t^\Phi}, \Sigma_{\xi_t^\Sigma}, H^\Phi, H^\Sigma]$ and $P[\xi_t^\Sigma = i|Z^T, \Phi_{\xi_t^\Phi}, \Sigma_{\xi_t^\Sigma}, H^\Phi, H^\Sigma]$. These are probabilities for the different regimes conditional on the model parameters and the whole dataset $Z^T = \{Z_s\}_{s=1}^T$.

[2] Note that the regime switch is modeled for a VAR in its reduced form. Sims and Zha (2006) work with the structural form of the VAR. However, it is not clear what kind of identifying restrictions could be imposed when dealing with four financial variables such as the ones that are included in the present model. Therefore, it seems more reasonable to proceed with this approach than to attempt to impose restrictions that are difficult to justify.

[3] An alternative way to model parameter instability consists of using smoothly time-varying parameters as in Cogley and Sargent (2006) and Primiceri (2005).

3.2 Dataset

The vector Z_t contains four state variables: The excess log return on the CRSP value-weighted index (ER_t), the Term Yield spread in percentage points (TY_t), measured as the yield difference between ten-year constant-maturity taxable bonds and short-term taxable notes, the log price earning ratio (PE_t), and the small-stock value-spread (VS_t), the difference in the log book-to-market ratios of small-value and small-growth stocks. The sample spans the period from December 1928 to June 2009.

The construction of the series follows Campbell and Vuolteenaho (2004). The excess market return is computed as the difference between the log return on the Center for Research in Securities Prices (CRSP) value-weighted stock index and the rate on three-month Treasury bills. The Term Yield spread is computed using data available on Global Financial Data by taking the yield difference between 10-year constant-maturity taxable bonds and short-term taxable notes, in percentage points. The Price Earnings ratio (Shiller, 2000) is the log of the ratio between the price of the S&P 500 index and a 10-year moving average of aggregate earnings of companies in the S&P 500 index. In line with the literature, earnings are averaged to avoid spikes in the Price Earnings ratio caused by cyclical fluctuations in earnings. The moving average is lagged by one quarter in order to ensure that all components of the time-t Price Earnings ratio are observable at time t.

The small-stock Value spread is constructed by using the six "elementary" portfolios available on Professor French's website. These elementary portfolios, which are constructed at the end of each June, are the intersections of two portfolios based on size (market equity, ME) and three portfolios formed on the ratio of book equity to market equity (BE/ME). The size breakpoint for year t is the median NYSE market equity at the end of June of year t. The book-to-market ratio for June of year t is the book equity for the last fiscal year end in $t-1$ divided by ME for December of $t-1$. The BE/ME breakpoints are the 30th and 70th NYSE percentiles.

At the end of June of year t, the small-stock Value spread is given by the difference between the $\ln(BE/ME)$ of the small high-book-to-market portfolio and the $\ln(BE/ME)$ of the small low-book-to-market portfolio. For months July through May, the small-stock Value spread is updated by adding the cumulative log return from the previous June on the small low-book-to-market portfolio minus the cumulative log return on the small high-book-to-market portfolio to the end-of-June small-stock Value spread. Therefore, an increase in the Value spread reflects the fact that small-growth stocks are outperforming small-value stocks.

3.3 Estimation algorithm

The model is estimated with Bayesian methods. Proper priors are put on all the parameters in the model. The priors for all parameters are very loose and identical across the different regimes. This implies that the features of the regimes are not restricted and that differences will arise only because of the data. Appendix A describes the priors in detail. I also impose covariance stationarity by adopting the concept of mean square stability. An MS model is mean square stable if both the first and second moments converge.[4]

The posterior is obtained combining the likelihood with the priors. Appendix B describes how to compute the likelihood and the regime probabilities for a given set of parameters. I first search for the posterior mode maximizing the sum of the logarithm of the priors and the log-likelihood. This is an important step because MS models tend to have multiple peaks. I then employ a Gibbs sampling algorithm to draw from the posterior distribution. Here I briefly summarize the steps of the Gibbs sampling that is described in detail in Appendix C:

1. Sampling $\xi^{\Phi,T}$ and $\xi^{\Sigma,T}$: Following Kim and Nelson (1999a) I use a Multi-Move Gibbs sampling to draw ξ_t^{Φ} from $f(\xi_t^{\Phi}|Z^T, \Phi_{\xi_t^{\Phi}}, \Sigma_{\xi_t^{\Sigma}}, H^{\Phi}, H^{\Sigma}, \xi_t^{\Sigma})$ and ξ_t^{Σ} from $f(\xi_t^{\Sigma}|Z^T, \Phi_{\xi_t^{\Phi}}, \Sigma_{\xi_t^{\Sigma}}, H^{\Phi}, H^{\Sigma}, \xi_t^{\Sigma})$.

2. Sampling $\Sigma_{\xi_t^{\Sigma}}$ given $\Phi_{\xi_t^{\Phi}}$, $\xi^{\Phi,T}$, $\xi^{\Sigma,T}$: Given $\Phi_{\xi_t^{\Phi}}$ and the regime sequence $\xi^{\Phi,T}$ we can compute the residuals of the MS-VAR at each point in time. Then, given $\xi^{\Sigma,T}$, we can group all the residuals that pertain to a particular regime. Therefore, $\Sigma_{\xi_t^{\Sigma}}$ can be drawn from an inverse Wishart distribution for $\xi_t^{\Sigma} = 1...m^{\Sigma}$.

3. Sampling $\Phi_{\xi_t^{\Phi}}$ given $\Sigma_{\xi_t^{\Sigma}}$, $\xi^{\Phi,T}$, $\xi^{\Sigma,T}$: When drawing the VAR coefficients, we need to take into account the heteroskedasticity implied by the switches in $\Sigma_{\xi_t^{\Sigma}}$. This can be done grouping all the observations that pertain to a specific regime and then applying the formulas for the posterior of the VAR coefficients recursively.

4. Sampling H^{Φ} and H^{Σ}: Given the draws for the state variables $\xi^{\Phi,T}$ and $\xi^{\Sigma,T}$, the transition probabilities are independent of Z^T and the other parameters of the model. Therefore, they can be drawn using a Dirichlet distribution.

5. If the algorithm has reached the desired number of iterations, stop. Otherwise, go back to step 1.

I use 1,000,000 Gibbs sampling iterations of which one every 100 are retained. Convergence is checked using the methods suggested by Geweke (1992) and Raftery and Lewis

[4]Mean square stability holds if and only if all the eigenvalues of the matrix $\Xi \equiv bdiag(A_1 \otimes A_1, ..., A_m \otimes A_m)(H^{\Phi} \otimes I_{n^2})$ are inside the unit circle where *bdiag* is a matrix operator that takes a sequence of matrices and construct a block diagonal matrix. Please refer to Costa, Fragoso, and Marques (2004) and Bianchi (2014) for more details.

(1992).[5]

4 The Great Depression and The Great Recession

In what follows I highlight the similarities and differences between the Great Depression and the Great Recession.

4.1 Parameter estimates and regime probabilities

This subsection reports parameter estimates and regime probabilities for the MS-VAR described above. The number of regimes for the VAR coefficients is equal to two, $m^{\Phi} = 2$, while the number of regimes for the covariance matrix is equal to three, $m^{\Sigma} = 3$. Therefore, we have a total of six possible regime combinations. Figure 1 shows the smoothed and filtered probabilities of Regime 1 for the VAR coefficients ($\xi_t^{\Phi} = 1$) at the posterior mode, while Figure 2 reports the smoothed and filtered probabilities for Regime 1 and Regime 3 for the covariance matrices ($\xi_t^{\Sigma} = 1$ and $\xi_t^{\Sigma} = 3$). Table 1 reports posterior mode and 68% error bands for the parameters of the Markov-switching VAR.

I shall start by analyzing the results for the VAR coefficients. The upper panel of Figure 1 contains the filtered and smoothed probabilities of Regime 1 for the VAR coefficients ($\xi_t^{\Phi} = 1$) together with the evolution of the Price Earnings ratio and the Value spread, where the variables have been normalized to fit in the graph. I report both the filtered and smoothed probabilities because they convey different information. We can think about the filtered probability as the probability that would be attached to a particular regime by an agent that was aware of all parameters of the model except for the regime in place at time t. In other words, this is the probability that an agent would attach to Regime 1 at time t if she knew the VAR coefficients, the covariance matrices, the transition matrices, and only the data up to time t. Instead, the smoothed probabilities reflect all the information contained in the dataset. This is the probability that an agent would attach to Regime 1 at time t if she knew the VAR coefficients, the covariance matrices, the transition matrices, and the whole dataset up to time T.

In order to facilitate the interpretation of the results, the second row of the figure focuses on three key events: The Great Depression, the IT bubble, and the Great Recession. Regime 1 clearly dominates the first decade, a period characterized by large market crashes and an unusually high level for the Value spread. The behavior of the Value spread and the price earning ratio in the early 1930s is worth noting. The largest stock market crash of U.S.

[5] For Raftery and Lewis (1992) I target 90% credible sets, with a 1% accuracy to be achieved with a 95% minimum probability. The required number of draws never exceeds $2,000$. This value is well below the number of draws retained to analyze the model ($10,000$).

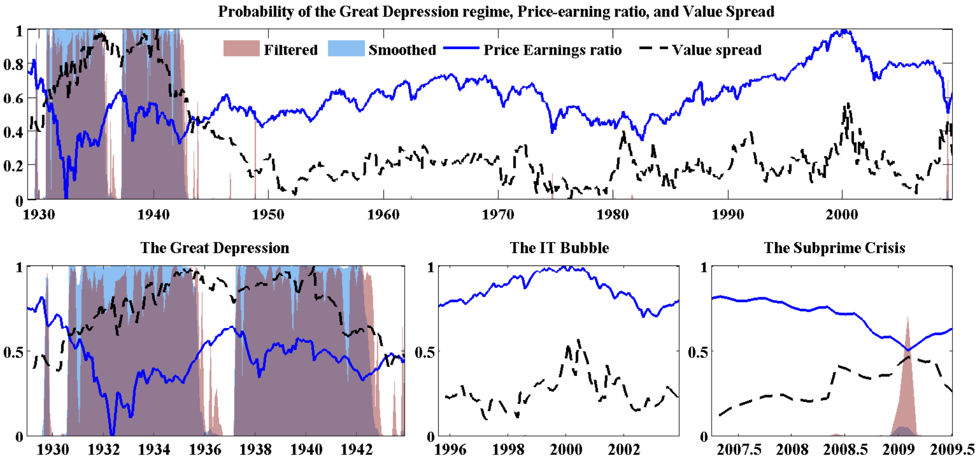

Figure 1: **Regime probabilities for VAR coefficients.** The first panel reports the filtered (red/dark gray area) and smoothed (blue/light gray area) probabilities of the Great Depression regime together with the evolution of the price-earnings ratio and the value spread. The lower three panels zoom on three key events: The Great Depression, the end of the Information Technology bubble, and the Great Recession.

history came with a substantial increase in the Value spread that reached historic heights. In other words, during the most severe recession that the U.S. has ever experienced, growth stocks were outperforming value stocks, and this situation of disequilibrium lasted for more than a decade. The probability of this regime went down only around 1942, when the U.S. started winning WWII. A rational agent who is trying to hedge against risk is likely to find this pattern extremely informative. From here forward I will refer to Regime 1 as the *Great Depression regime*, while I will name Regime 2 the *Regular times regime*.

After the 1930s, the probability of the *Great Depression regime* has generally been close to zero. However, a visible increase in the probability occurred in correspondence with the recent financial crisis. The increase is much larger for the filtered probability than for the smoothed probability. This implies that an agent that had been observing the market in real time would have attached a much larger probability to entering a depression-like regime, while ex-post, with the benefit of the hindsight, the same agent would have concluded that the probability of having observed a manifestation of the Great Depression regime was in fact much smaller. However, even in this latter case, in which the entire dataset is used to infer the smoothed probabilities, we cannot rule out the possibility that during the first two months of 2009, financial markets' behavior was in line with what occurred during the dawn of the Great Depression.

While there are other periods of time during which we observe an increase in the filtered probability of the Great Depression regime, the second month of 2009 was the first time that such a probability crossed 50% since November 1948, a period marked by the rise of the Cold War, the first Israeli-Arab war, and the unexpected presidential election victory of the incumbent President Truman over the Republican candidate, Thomas E. Dewey. Similarly, in the first two months of 2009, the smoothed probability crossed the 5% value for the first time since September 1942, i.e., since World War II. Finally, it is worth emphasizing that these results are even stronger if we were to recursively estimate the model. In this case, the probability assigned to the Great Depression regime would be larger than 80% in February 2009. Section 6 reports results for this alternative approach.[6]

Later, I will investigate more in depth what could explain the increase in the probability of the Great Depression regime at the beginning of 2009. For now, it is enough to point out that the spike in the probability of the Great Depression regime at the beginning of 2009 coincides with a deep decline in the Price Earnings ratio combined with a substantial increase in the Value spread. In other words, the price earning ratio and the Value spread are moving in opposite directions in a way that is very similar to what occurred during the early years of the Great Depression. In this respect, it is quite instructive to compare the Great Recession stock market decline with the end of the IT bubble. In this second case, the Value spread and the Price Earnings ratio were moving in parallel. Recall that the Value spread tends to rise when growth stocks perform relatively better than value stocks. Given that the rise and burst of the IT bubble were mostly driven by IT stocks, it is not surprising that the two variables were moving together. Nevertheless, this evidence suggests that stock market crashes that are associated with financial crises might have very different implications for the relative performance of growth and value stocks. This is why the probability of the Great Depression regime does not increase every time that the Price Earnings ratio falls, but it is more likely to do so if such a fall is associated with a contemporaneous increase in the Value spread.

The two regimes are strongly identified and the parameter estimates present some distinctive features.[7] First of all, the autoregressive component for excess returns is substantially larger under the Great Depression regime ($\xi_t^\Phi = 1$), while the autoregressive components for the Term Yield spread, the Price Earnings ratio, and the Value spread are substantially smaller. A high price earning ratio predicts low stock market returns in both regimes, but the effect is significantly stronger under the Great Depression regime. The Value spread enters the excess return and Price Earnings ratio equations with a positive sign in both regimes, but the coefficients are substantially larger under the Great Depression regime. Finally, the coefficients of the Term Yield spread and of the Price Earnings ratio in the Value spread equation are positive under the Great Depression regime, while they are smaller and negative in the Regular times regime ($\xi_t^\Phi = 2$). Before proceedings, it is worth emphasizing that the

[6] I focus on the estimates obtained using the entire sample to reflect the idea that agents in the economy have in fact more information than the econometrician. Therefore, the econometrician uses all the available information to learn the parameters that are known to investors. This approach is common in the asset pricing literature.

[7] Appendix D contains results for a formal pairwise comparison of the VAR coefficients across the two regimes.

$\xi_t^\Phi = 1$	ER_t	TY_t	PE_t	VS_t	$constant$
ER_{t+1}	0.1650 (0.0509,0.1822)	−0.0184 (−0.0255,−0.003)	−0.1225 (−0.1232,−0.0937)	0.1275 (0.0894,0.1341)	0.0423 (0.0065,0.0608)
TY_{t+1}	−0.1492 (−0.4170,0.1408)	0.8874 (0.8468,0.9644)	−0.0332 (−0.1030,0.0761)	0.0804 (−0.0958,0.1800)	0.1460 (−0.0512,0.3517)
PE_{t+1}	0.1657 (0.0480,0.1832)	−0.0175 (−0.0247,−0.0007)	0.8732 (0.8720,0.9029)	0.1444 (0.1022,0.1498)	0.0050 (−0.0266,0.0288)
VS_{t+1}	0.0062 (−0.0438,0.0587)	0.0450 (0.0307,0.0490)	0.0563 (0.0365,0.0617)	0.9000 (0.8905,0.9299)	0.0020 (−0.0228,0.0373)

$\xi_t^\Phi = 2$	ER_t	TY_t	PE_t	VS_t	$constant$
ER_{t+1}	0.0563 (0.0251,0.0939)	0.0010 (0.0001,0.0029)	−0.0134 (−0.0176,−0.0093)	0.0157 (0.0088,0.0216)	0.0234 (0.0094,0.0380)
TY_{t+1}	0.2591 (0.0351,0.4762)	0.9638 (0.9580,0.9756)	−0.0171 (−0.0416,0.0118)	0.0729 (0.0349,0.1130)	−0.0176 (−0.1263,0.0669)
PE_{t+1}	0.0187 (−0.0125,0.0575)	0.0016 (0.0006,0.0034)	0.9913 (0.9869,0.9952)	0.0193 (0.0121,0.0250)	−0.0019 (−0.0150,0.0138)
VS_{t+1}	−0.0024 (−0.0286,0.0263)	−0.0027 (−0.0033,−0.0010)	−0.0037 (−0.0067,0.0002)	0.9765 (0.9698,0.9812)	0.0465 (0.0331,0.0582)

$\xi_t^\Sigma = 1$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0653 (0.0679,0.0840)	0.0033 (−0.0046,0.0125)	0.0036 (0.0039,0.0061)	0.0015 (0.001,0.0021)
u_{TY}	0.0617 (−0.0657,0.1723)	0.8067 (0.8449,1.0257)	0.0026 (−0.0051,0.0109)	−0.005 (−0.0140,−0.0022)
u_{PE}	0.9189 (0.8864,0.9343)	0.0520 (−0.0795,0.1621)	0.0608 (0.0633,0.0788)	0.0014 (0.0009,0.0020)
u_{VS}	0.4559 (0.256,0.4975)	−0.1212 (−0.2739,−0.0463)	0.457 (0.2596,0.5016)	0.0506 (0.0482,0.0585)

$\xi_t^\Sigma = 2$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0363 (0.0367,0.0389)	−0.0005 (−0.0009,−0.0002)	0.0013 (0.0013,0.0015)	0.0001 (0.0001,0.0002)
u_{TY}	−0.0688 (−0.1056,−0.0259)	0.2133 (0.2233,0.2470)	−0.0005 (−0.0009,−0.0002)	0.0002 (−0.0001,0.0004)
u_{PE}	0.9474 (0.9473,0.955)	−0.0592 (−0.0955,−0.0169)	0.0369 (0.0371,0.0394)	0.0001 (0.0001,0.0002)
u_{VS}	0.1194 (0.1152,0.2006)	0.0285 (−0.0181,0.0604)	0.1001 (0.0944,0.1785)	0.0279 (0.0284,0.0305)

$\xi_t^\Sigma = 3$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.1040 (0.1068,0.1270)	−0.0078 (−0.0132,−0.0026)	0.0110 (0.0116,0.0165)	−0.0026 (−0.0051,−0.0016)
u_{TY}	−0.2353 (−0.3151,−0.0651)	0.3183 (0.3083,0.3922)	−0.0085 (−0.0143,−0.0033)	−0.0025 (−0.0076,0.0022)
u_{PE}	0.9611 (0.9521,0.9694)	−0.2411 (−0.3240,−0.0800)	0.1102 (0.1134,0.1348)	−0.0029 (−0.0057,−0.0019)
u_{VS}	−0.2206 (−0.3286,−0.1158)	−0.0684 (−0.1714,0.0509)	−0.2346 (−0.3425,−0.1305)	0.1139 (0.1161,0.1387)

H^Φ	$\xi_t^\Phi = 1$	$\xi_t^\Phi = 2$
$\xi_{t+1}^\Phi = 1$	0.9778 (0.9561,0.9831)	0.0050 (0.0038,0.0099)
$\xi_{t+1}^\Phi = 2$	0.0222 (0.0169,0.0439)	0.9950 (0.9901,0.9962)

H^Σ	$\xi_t^\Sigma = 1$	$\xi_t^\Sigma = 2$	$\xi_t^\Sigma = 3$
$\xi_{t+1}^\Sigma = 1$	0.7959 (0.7343,0.8532)	0.0224 (0.0075,0.0232)	0.0526 (0.0353,0.1037)
$\xi_{t+1}^\Sigma = 2$	0.1685 (0.0865,0.2030)	0.9243 (0.9182,0.9419)	0.3666 (0.3667,0.5150)
$\xi_{t+1}^\Sigma = 3$	0.0356 (0.0288,0.0940)	0.0533 (0.0443,0.0652)	0.5807 (0.4152,0.5636)

Table 1: Parameter estimates. The three sets of tables contain modes and 68% error bands for the posterior distribution of the parameters of the Markov-switching VAR. The first two panels report the estimates for the VAR coefficients. The second set of panels contains the standard deviations of the shocks on the main diagonal, the correlations of the shocks below the main diagonal, and the covariances above the main diagonal. The last tables contain the estimates of the transition matrices.

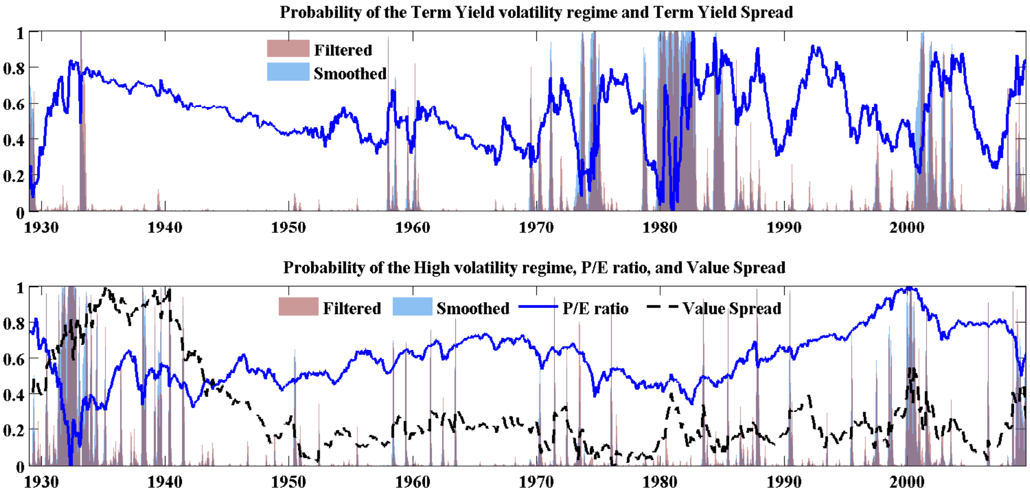

Figure 2: **Probabilities of the volatility regimes.** The figure reports the filtered and smoothed probabilities of Term Yield volatility regime (top panel) and the High volatility regime (lower panel). These two regimes correspond to Regime 1 and Regime 3 for the covariance matrix. The first panel also reports the evolution of the Term Yield Spread, while the second panel contains the Price-earnings ratio and the Value Spread. All variables are rescaled to fit in the 0-1 scale.

dynamic properties of the model do not only depend on the VAR coefficients: Regime changes can also induce strong commovements between the variables of interest. These aspects will be analyzed in the next subsection.

Some interesting patterns emerge from the analysis of the covariance matrix estimates and their corresponding probabilities. Regime 2 ($\xi_t^\Sigma = 2$) can be regarded as the *Low volatility regime*, showing the lowest values for the standard deviations of all innovations. Regime 1 ($\xi_t^\Sigma = 1$) presents an increase of the magnitude for all shocks, but especially for the innovations to Term Yield spread. Looking at Figure 2, we can see that this regime mostly prevails during the early years of the Volcker chairmanship when the Federal Reserve was targeting reserves with the result of generating high volatility in the FFR and, consequently, the yield spread. I will refer to this regime as the *Term Yield volatility regime.* Regime 3 ($\xi_t^\Sigma = 3$) is instead characterized by a more modest increase in the volatility of the Term Yield spread innovations, but a much larger increase in the volatility of the other shocks. Interestingly, Regime 3 prevails for extended periods of time during the 1930s, the 2001 IT bubble burst, and the 2008/9 financial crisis. So it can be considered an *High Uncertainty regime* across several dimensions. The correlation structure of the innovations is also worth noting. Under the High Uncertainty regime, innovations to the Value spread are strongly negatively correlated with innovations to the excess return and Price Earnings equations. This is in sharp contrast with the positive sign that prevails under the other two regimes and implies that small growth stocks tend, in relative terms, to move against the market. Similarly, the correlation of Term Yield innovations with Price Earnings ratio and excess return innovations is strongly negative under the High Uncertainty regime, while under the other two regimes the correlation is slightly negative (Low volatility regime) or centered on zero (Term Yield volatility regime).

Finally, the regime persistences reported at the bottom of Table 1 show that for the VAR coefficients, the Great Depression regime is significantly less persistent and frequent than the Regular times regime, consistent with the idea that the Great Depression was a rare event. As for the transition matrix of the innovation covariance matrix, Regime 2, the Low volatility regime, is the most persistent, followed by the Term Yield volatility regime and the High volatility regime. Their unconditional probabilities are 77.8%, 11.3%, and 10.9%. These estimates imply that the low volatility regime prevails most of the time with relatively frequent but short lasting deviations to the Term Yield volatility regime and the High volatility regime.

4.2 Entering the Great Depression

As mentioned above, regime changes also play a key role in shaping the dynamic properties of the model. In fact, regime changes can be regarded as shocks themselves and can have fairly long lasting consequences. In order to understand the role played by regime changes and at the same time capture the salient features of the Great Depression, Figure 3 reports a simulation in which all Gaussian shocks are set to zero, and regimes follow their most likely path based on the smoothed probabilities at the posterior mode. The initial values coincide with the actual data. The simulated series are reported with a solid blue line, while the red dashed line corresponds to the actual data. The two horizontal lines mark the regime-specific conditional steady states. These are the values to which the variables would converge if a regime were in place forever.

The first aspect that emerges from this simulation is that a change from the Regular times regime to the Great Depression regime determines a sharp drop in the stock market and a contemporaneous increase in the Value spread and the Term Yield spread. The drop in the stock market tends to be very large, and it overshoots with respect to the conditional steady state of the Great Depression regime. Therefore, after a dramatic fall, the stock market partially recovers, while the Value spread and Term Yield spread keep moving toward their corresponding conditional steady states. Notice that the short break in the realization of the Great Depression regime that is identified in the estimates coincides with a partial recovery in the stock market and a contemporaneous fall in the Value spread. However, once the model returns to the Great Depression regime the variables tend to follow a path similar to the one that was prevailing before the break. Overall, during the 1930s the regime sequence plays an important role in tracking the behavior of the three variables, implying that the Great Depression regime captures some salient features of the Great Depression.

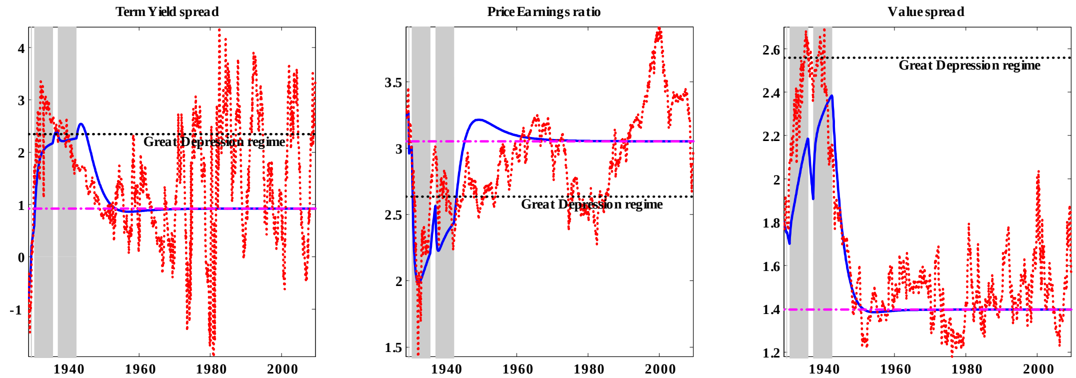

Figure 3: **The Great Depression.** The figure reports a simulation in which all the Gaussian shocks are set to zero, and regimes follow their most likely path based on the smoothed probabilities at the posterior mode. The initial values coincide with the actual data. The simulated series are reported with a solid blue line, while the red dashed line corresponds to the data. The two horizontal lines mark the regime-specific conditional steady states. These are the values to which the variables would converge if a regime were in place forever.

Once the economy returns to the Regular times regime, the model predicts a quick fall in the Value spread and the Term Yield spread. The stock market moves in the opposite direction, showing a steady increase and converging to the higher Regular times conditional steady state. It is also important to emphasize that the conditional steady state for the Great Depression regime is never really reached by the Value spread and the Price Earnings ratio. This is because both in the estimation and in the simulation, the Great Depression regime is not in place long enough to allow for convergence to the conditional steady state.[8]

Figure 4 focuses on the last months of the sample to better understand the similarities and the differences between the Great Depression and the Great Recession. The figure reports two simulations in which all Gaussian shocks have been set to zero starting from February 2009, the month in which the filtered probability of the Great Recession regime spiked. In the first simulation (solid blue line), a counterfactual regime sequence is assumed: starting from February 2009, the Great Depression regime prevails until the end of the sample. In the second simulation, the actual regime sequence is assumed to be in place. The red dotted line corresponds to the data. Notice that until February 2009, the three series coincide. Therefore, the two simulations can be used to understand why the probability of the Great Depression regime increased in the very beginning of 2009, but then quickly fell in March 2009. Furthermore, the simulations shed light on what agents were likely to expect in the moment that the probability of the Great Depression regime spiked.

As already noted above, since the end of 2008 and until February 2009, the stock market experienced a prolonged fall associated with a contemporaneous increase in the Value

[8] In Markov-switching models this is a fairly common finding. If the variables converge or not to their conditional steady states depends on the interaction between the persistence of the regime and the persistence of the variables under such a regime.

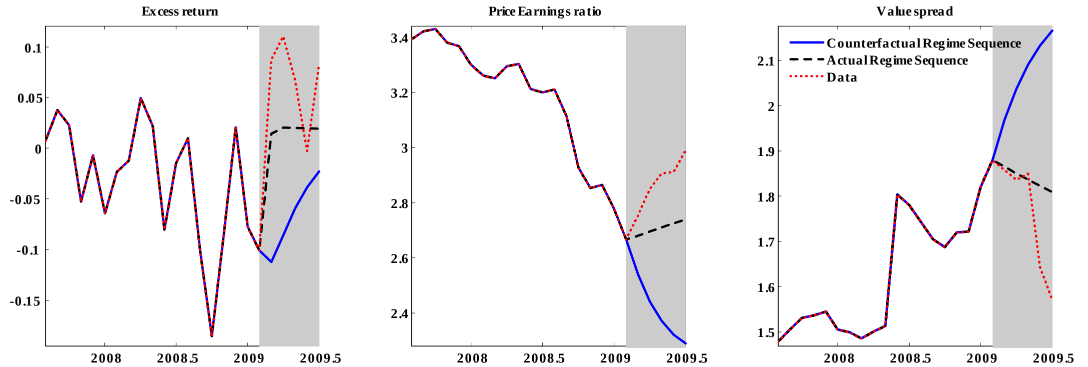

Figure 4: **The Great Recession.** The figure reports two simulations in which all Gaussian shocks have been set to zero starting from March 2009. In the first simulation (solid blue line) a counterfactual regime sequence is assumed: Starting from March 2009 the Great Depression regime prevails until the end of the sample. In the second simulation the actual regime sequence is assumed to be in place. The red dotted line corresponds to the data.

spread. This behavior is remarkably similar to what is observed in the beginning of the Great Depression. The solid blue line shows what would have happened if starting February 2009 the economy had in fact entered the Great Depression regime: The Price Earnings ratio and the Value spread would have kept moving in exactly the same fashion, while excess returns would have stayed negative. In other words, the counterfactual simulation highlights that until February 2009 financial markets were in fact on a path very similar to what implied by the Great Depression regime. However, in March 2009, these dynamics reverted. Excess stock market returns increased and became positive, the Price Earnings ratio recovered, and the Value spread started declining. The black dashed line shows that the return to the Regular times regime captures these changes, even if in the data the movements were somehow more pronounced. Recall that this is a period of high volatility, so the discrepancy between the "regime only" simulation and the actual data should not be surprising.

In light of these findings, it is then interesting to review the main events that characterized the beginning of the Great Recession. An early flag emerged in June 2007, with the collapse of two hedge funds owned by Bear Stearns. Less than one year later, in March 2008, the Federal Reserve had to intervene in order to prevent the Bear Stearns bankruptcy by assuming $30 billion in liabilities and engineering a sale to JPMorgan Chase. From that moment on, the crisis accelerated with the Treasury Department taking over Fannie Mae and Freddie Mac on September 7, Lehman brothers filing for the largest bankruptcy case in U.S. history one week later (September 15), and the Federal Reserve bailing out AIG. In December 2008, unemployment reached its highest value in 15 years and the Federal Reserve cut the FFR to zero. Over the same period of time, the Price Earnings ratio kept moving down while the Value spread increased.

President Obama took office in January 2009, and Wall Street experienced the worst Inauguration Day drop ever (I am not claiming a causal relation between the two events). At this point, fears that the U.S. might be heading toward a second Great Depression became widespread (Krugman, 2009). On February 10, the secretary of the Treasury Geithner outlined the plan for the expansion of the government bank rescue effort. The plan was received with some skepticism by financial markets, arguably because it was lacking many important details (Solomon, 2009). As a result, the market experienced a fall of almost 5%. A few days later, President Obama signed into law a $787 billion stimulus package that included tax cuts and money for infrastructure, schools, health care, and green energy. Even in this case, some commentators worried that the government intervention might not be enough. In the meantime, the stock market experienced two weeks of declines, reaching its lowest level in 12 years. Notice that it is in February that the probability of the Great Depression regime crossed 50%. However, in March 2009, more encouraging economic data were released and details of the rescue plan were disclosed. Arguably, this had a positive effect of the stock market that turned around. At the same time, the Value spread started moving down and the probability of the Great Depression regime went back to values close to zero.

In summary, the estimates and the counterfactual simulations suggest that during the second half of 2008 and until February 2009, financial markets were on a path consistent with a switch to the Great Depression regime: a falling Price Earnings ratio and an increasing Value spread. This explains the increase in the probability of the Great Depression regime. These patterns came to a stop in March 2009 when the government increased its effort to prevent a financial meltdown and to facilitate an economic recovery. This explains why in the estimates the filtered probability assigned to the Great Depression regime increased significantly at the beginning of 2009, but it quickly went back to zero: The economy did not enter a Great Depression, at least in terms of the behavior of financial markets.

5 The Cross Section of Asset Returns

The results shown so far have highlighted a series of properties that are quite informative regarding the similarities between the Great Depression and the Great Recession. Two aspects seem particularly relevant. First, during the Great Depression and at the beginning of the Great Recession, the Price Earnings ratio and the Value spread were moving in opposite directions. Second, innovations to the Price Earnings ratio and the Value spread were often negatively correlated during these events. Both results suggest that financial crises imply important changes in the behavior of the cross section of asset returns, with small growth stocks performing relatively better. In order to further explore this idea, I make use of Campbell and Vuolteenaho's ICAPM. Consistent with the Markov-switching model described above, it is important to model the possibility of regime changes when describing agents' expectations formation mechanism. In order to keep the paper self-contained, I will briefly present the ICAPM proposed by Campbell and Vuolteenaho (2004), and then I will explain how to extend their approach to allow for regime changes.

5.1 ICAPM

Fama and French (1992, 1993) show that the CAPM fails to describe average realized stock returns since the early 1960s, when a value-weighted equity index is used as a proxy for the market portfolio. This failure is most apparent for the price of small stocks and value stocks. Those stocks have experienced average returns that cannot be explained through their market betas. Ang and Chen (2007) argue that this failure is specific to the post-1963 sample.

In order to solve the small-value puzzle, Campbell and Vuolteenaho (2004) start from the premise that the returns on the market portfolio can be split into two components. An unexpected change in excess returns can be determined by news about future cash flows or by a change in the discount rate that investors apply to these cash flows. While a fall in expected cash flows is simply bad news, an increase in discount rates implies at least an improvement in future investment opportunities. Therefore, the single CAPM beta can be decomposed into two sub-betas: one reflecting the covariance with news about future cash flows (bad beta), the other linked to news about discount rates (good beta). The previous argument suggests that given two assets with the same CAPM beta, the one with the highest cash-flow beta should have a larger return.

Using the loglinear approximation for returns introduced by Campbell and Shiller (1988), unexpected excess returns can be approximated by:

$$r_{t+1} - \mathbb{E}_t r_{t+1} = N_{CF,t+1} - N_{DR,t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \tag{3}$$

where r_{t+1} is a log stock market return, d_{t+1} is the log dividend paid by the stock, Δ denotes a one period change, $\mathbb{E}_t$ denotes a rational expectation formed at time t, and ρ is the discount coefficient that is set to 0.95 per annum. $N_{CF,t+1}$ and $N_{DR,t+1}$ represent news about the future market cash flows and news about the future market discount returns, respectively.

The VAR methodology introduced by Campbell (1991) provides an estimate for the terms $\mathbb{E}_t r_{t+1}$ and $N_{DR,t+1}$. Then $N_{CF,t+1}$ is derived from (3) as a residual. Specifically, consider a VAR in companion form:

$$Z_{t+1} = c + AZ_t + u_{t+1} \tag{4}$$

where Z_t is a vector of state variables with the excess return ordered first. Assuming that agents form expectations using the above VAR, the two types of news can be obtained according to the following transformation of the residuals:

$$r_{t+1} - \mathbb{E}_t r_{t+1} = e_1' u_{t+1} \tag{5}$$

$$N_{CF,t+1} = (e_1' + e_1' \lambda) u_{t+1} \tag{6}$$

$$N_{DR,t+1} = e_1' \lambda u_{t+1} \tag{7}$$

where $\lambda = \rho A (I - \rho A)^{-1}$ and $e_1' = [1, 0, ..., 0]'$. The residuals of the different equations are given weights reflecting their persistence and their contribution in explaining the excess return. The first effect is captured by $(I - \rho A)^{-1}$, and the second by ρA.

Once the two series for the news have been obtained, the betas can be computed for a set of portfolios according to the following formulas:[9]

$$\widehat{\beta}_{i,CF} = \frac{\widehat{cov}(r_{i,t}, N_{CF,t})}{\widehat{var}(N_{CF,t} - N_{DR,t})} \tag{8}$$

$$\widehat{\beta}_{i,DR} = \frac{\widehat{cov}(r_{i,t}, -N_{DR,t})}{\widehat{var}(N_{CF,t} - N_{DR,t})} \tag{9}$$

where $r_{i,t}$ is the return of the *i-th* portfolio. Notice that the denominator is simply the sample variance of the unexpected excess returns, i.e., of the residuals of the VAR $r_{t+1} - \mathbb{E}_t r_{t+1}$. The market beta is obtained by summing the two betas.

Campbell (1993) derives an approximate discrete-time version of Merton's (1973) ICAPM. The pricing implications of the model are based on the first-order condition of an investor with Epstein and Zin (1989) preferences who holds a portfolio of tradable assets that contains all of her wealth. Campbell assumes that this portfolio is observable in order to derive testable asset-pricing implications from the first-order condition. Under appropriate assumptions about the parameters of the model, it can be shown that the price of risk for the discount-rate beta should equal the variance of the market return, while the price of risk for the cash-flow beta should be γ times greater, where γ is the investor's coefficient of relative risk aversion. Therefore, the model provides economically motivated restrictions and can be tested against the standard CAPM.

Three models can be examined: the static CAPM, the ICAPM, and an unrestricted factor model based on the two betas. Consider the cross-sectional regression

$$\overline{R}_i = g_0 + g_1 \widehat{\beta}_{i,CF} + g_2 \widehat{\beta}_{i,DR}$$

[9] Campbell and Vuolteenaho (2004) also allow for a lag in the formulas used to compute the betas to control for the possibility that not all stocks in the test-asset portfolios were traded frequently and synchronously. See page 1258 of their paper. In that case, the formulas for the betas become: $\widehat{\beta}_{i,CF} = \frac{\widehat{cov}(r_{i,t}, N_{CF,t})}{\widehat{var}(N_{CF,t} - N_{DR,t})} + \frac{\widehat{cov}(r_{i,t}, N_{CF,t-1})}{\widehat{var}(N_{CF,t} - N_{DR,t})}$ and $\widehat{\beta}_{i,DR} = \frac{\widehat{cov}(r_{i,t}, -N_{DR,t})}{\widehat{var}(N_{CF,t} - N_{DR,t})} + \frac{\widehat{cov}(r_{i,t}, -N_{DR,t-1})}{\widehat{var}(N_{CF,t} - N_{DR,t})}$. The results presented in this paper are very similar across the two specifications. I decided to present results based on a beta without a lag because these formulas are more common. The results for the alternative specification are available upon request.

where $\overline{R}_i$ is the time-series mean for the excess return of asset i. The CAPM model imposes the coefficient restriction $g_1 = g_2$, given that the single market beta is obtained summing the two betas: $\widehat{\beta}_{i,M} = \widehat{\beta}_{i,CF} + \widehat{\beta}_{i,DR}$. According to the ICAPM, the premia should be: $g_1 = \gamma\sigma_M^2$ and $g_2 = \sigma_M^2$, where σ_M^2 is the variance of the unexpected excess returns. Therefore, the ICAPM restricts the coefficient of the discount-rate beta, and it returns an estimate of the coefficient of relative risk aversion γ.[10] In the factor model the coefficients are not restricted. The model can be interpreted as a generalization of the ICAPM that allows the rational investor's portfolio to include Treasury bills as well as equities.

5.2 News in a Markov-switching framework

In this subsection, I report the formulas for the news when the state variables follow a MS-VAR as described by (1) and (2):

$$
\begin{aligned}
Z_{t+1} &= c_{\xi_t^\Phi} + A_{\xi_t^\Phi} Z_t + u_{t+1} && (10) \\
\Phi_{\xi_t^\Phi} &= \left[c_{\xi_t^\Phi}, A_{\xi_t^\Phi}\right] && (11)
\end{aligned}
$$

where $u_{t+1} = \Sigma^{1/2}_{\xi^\Sigma_{t+1}} \omega_{t+1}$ and $\omega_t \sim N(0, I)$. Suppose that agents know the parameters of the model and use them to decompose unexpected returns into news. In order to derive the news in this context, we need to be able to model the revision in expectations implied by the MS-VAR residuals taking into account the possibility of regime changes.

Define the conditional expectation $\mathbb{E}_0(Z_t) = \mathbb{E}(Z_t|\mathbb{I}_0)$ with $\mathbb{I}_0$ being the information set available at time 0. Notice that the expected value only depends on the realization of the Markov chain controlling the VAR coefficients up to time t, $\xi_1^\Phi ... \xi_t^\Phi$.[11] Define the $nm^\Phi \times 1$ column vector $q_t \equiv \left[q_t^{1\prime}, ..., q_t^{m^\Phi\prime}\right]'$ where $q_t^i = \mathbb{E}_0\left(Z_t 1_{\xi_t^\Phi = i}\right) = \mathbb{E}\left(Z_t 1_{\xi_t^\Phi = i}|\mathbb{I}_0\right)$ and $1_{\xi_t^\Phi = i}$ is an indicator variable that is one when regime i is in place. Note that:

$$q_t^i = \mathbb{E}_0\left(Z_t 1_{\xi_t = i}\right) = \mathbb{E}_0\left(Z_t|\xi_t = i\right)\pi_t^i$$

where $\pi_t^{\Phi,i} = P_0\left(\xi_t^\Phi = i\right) = P\left(\xi_t^\Phi = i|\mathbb{I}_0\right)$. Therefore we can obtain the conditional expecta-

[10] The asset pricing formulas of Campbell (1993) that represent the basis for Campbell and Vuolteenaho (2004) are derived assuming homoskedasticity. However, when modeling parameter instability it is important to allow for heteroskedasticity to avoid spurious changes in the VAR coefficients. This is why the MS-VAR was estimated allowing for heteroskedasticity. Given that the focus here is on the changes in dynamics implied by the Great Depression regime, I regard the idea of extending the analysis to price volatility in the spirit of Campbell, Giglio, Polk, and Turley (2014) as an interesting direction for future research, but beyond the scope of this paper. Furthermore, even in Campbell and Vuolteenaho (2004) there is not an immediate mapping between the volatility of the VAR innovations (based on estimates obtained over the entire sample) and the variance of the market returns used to price the assets (computed over two distinct subsamples).

[11] If we were interested in the evolution of the volatilities, we would also have to take into account the Markov chain controlling the evolution of the covariance matrix ξ_t^Σ. More details can be found in Bianchi (2014).

tion $\mathbb{E}_0(Z_t)$ as:

$$\mathbb{E}_0(Z_t) = \sum_{i=1}^{m} q_t^i = wq_t$$

where the matrix $w = [I_n, ..., I_n]$ is obtained placing side by side m^Φ n-dimensional identity matrices. This is a convenient result because while the law of motion of Z_t is not Markov, the law of motion of q_t is. Following Costa, Fragoso, and Marques (2004),[12] Bianchi (2014) shows that the law of motion of q_t is given by:

$$\begin{aligned} q_t &= C\pi_t^\Phi + \Omega q_{t-1} && (12) \\ \pi_t^\Phi &= H^\Phi \pi_{t-1}^\Phi && (13) \end{aligned}$$

with $\Omega = bdiag(A_1, ..., A_{m^\Phi})\left(H^\Phi \otimes I_n\right)$ and $C = bdiag(c_1, ..., c_{m^\Phi})$, where $\otimes$ represents the Kronecker product and *bdiag* is a matrix operator that takes a sequence of matrices and uses them to construct a block diagonal matrix. For the model considered in this paper, Ω and C are:

$$\Omega = \begin{bmatrix} A_1 h_{11}^\Phi & A_1 h_{12}^\Phi \\ A_2 h_{21}^\Phi & A_2 h_{22}^\Phi \end{bmatrix}, \; C = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}.$$

Under the assumption of mean square stability, the process for q_t converges to finite values. Then, given a sequence of probabilities $\pi^{\Phi,T}$ or a posterior draw for the regime sequence ξ^T, it is straightforward and computationally efficient to compute the sequences of discount-rate news and cash-flow news in one step (see Appendix E for a proof):

$$\begin{aligned} N_{DR}^T &= e_1' w \left[\lambda^q v^{q,T} + \lambda^\pi v^{\pi,T}\right] && (14) \\ N_{CF}^T &= e_1' w \left[(I_r + \lambda^q) v^{q,T} + \lambda^\pi v^{\pi,T}\right] && (15) \\ u^T &= e_1' w v^{q,T} && (16) \end{aligned}$$

where $\lambda^q = (I_{nm^\Phi} - \rho\Omega)^{-1}\rho\Omega$, $\lambda^\pi = (I_{nm^\Phi} - \rho\Omega)^{-1}\rho CH(I_r - \rho H)^{-1}$, $v_t^q = q_{t+1|t+1} - q_{t+1|t}$, and $v_{t+1}^\pi = \pi_{t+1|t+1}^\Phi - \pi_{t+1|t}^\Phi$, where $\pi_{t|t}^\Phi$ is a column vector whose i-th element coincides with $\pi_{t|t}^{\Phi,i} = P_t\left(\xi_t^\Phi = i\right)$, the probability of being in regime i at time t conditional on the information set available at time t. It is worth emphasizing that news now has two components. The first one is represented by the standard Gaussian innovation, while the second component derives from the revision in beliefs about the regime that is in place: $v_{t+1}^\pi = \pi_{t+1|t+1}^\Phi - \pi_{t+1|t}^\Phi$. For a given Gaussian innovation, the change in beliefs determines a change in the way the shocks are mapped into the future. When the two regimes coincide, formulas (14)-(16) collapse to (5)-(7). Therefore, the above formulas can be treated as a generalization of the ones used in Campbell and Vuolteenaho (2004).

[12] The results of Costa, Fragoso, and Marques (2004) are derived in the context of the engineering literature and cannot be directly applied to the MS-VARs generally used in economics because based on models in which regime changes are known one period in advance.

For the practical implementation of the formulas presented above, the vector of regime probabilities and parameters need to be replaced by their corresponding estimates. In the benchmark results presented below, I use the parameter estimates obtained using the entire sample and the corresponding filtered probabilities. Therefore, following Campbell and Vuolteenaho (2004) and many other contributions in the literature, I make the implicit assumption that agents in the model have more information than the econometrician. The econometrician tries to use the entire dataset to learn the parameters governing the economy and then uses the parameter estimates obtained over the entire sample to infer which regime agents thought was in place at each point in time. An alternative approach would be to assume that agents in the economy acts as econometricians and estimate the model recursively. In this second case, the agents' information set and the econometrician's information set are aligned (up to revision in the data). Results for this second approach are very similar and are described in Section 6.

It is worth pointing out that the approach described above can model situations in which not all m regimes are stable. This is because in order to be able to compute the news, we only need the discounted expectations to be stable. Mean square stability guarantees stability for first and second moments, i.e., covariance stationarity. Notice that this is in fact more than what is necessary for two reasons. First, the VAR implementation does not require the variance to be stable, but only that agents' expectations converge. Second, even if first moments are not stable, *discounted* first moments might be. However, it might be argued that imposing covariance stationarity is still desirable, given that it implies that agents' uncertainty converges to a finite value no matter the regime that is in place today. For the estimates considered in this paper, both regimes were determined to be stable.

In conclusion, MS models allow for temporary deviations from the stationarity assumption and are therefore useful when modeling agents' expectations formation mechanism. Instead, the assumption of stationarity is assumed to hold at each point in time when computing the news in a fixed coefficient framework. At the same time, the series for the news can still be conveniently derived using analytical expressions for the news. In other words, numerical integration is not necessary.

5.3 Evolution of the explanatory power of the models

Figure 5 reports the evolution of R^2 for the three models over rolling windows of 35 years.[13] For example, 1965 corresponds to the sample February 1930-January 1965. The size of the window is chosen in a way that the initial subsample roughly coincides with the first subsample used by Campbell and Vuolteenaho (2004). For each subsample, the betas are

[13] The R^2 is computed as $1 - RSS/RSM$ where RSS is the residual sum of squares and RSM is the residual sum of squares when only the constant is used as a regressor.

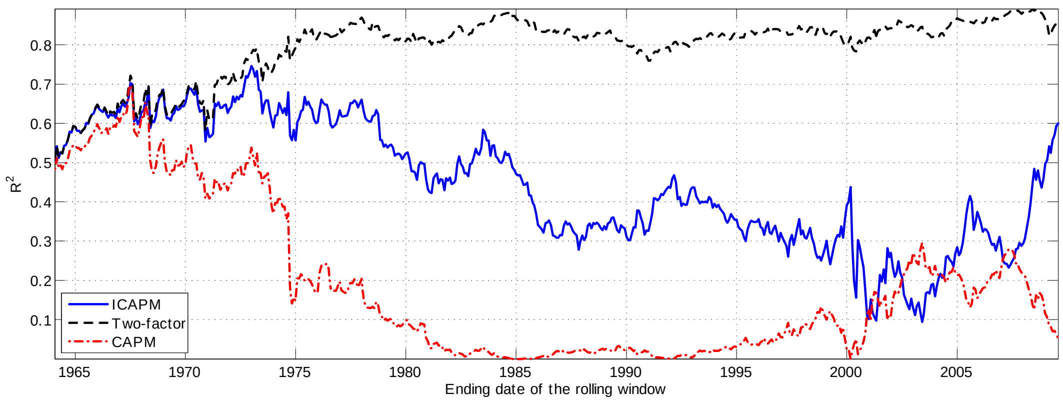

Figure 5: **Explanatory power of the ICAPM over moving windows.** The figure reports the explanatory power as measured by the R^2 of three models: The unrestricted two-factor model, the Intertemporal CAPM, and the traditional CAPM. The betas and average returns are computed over moving windows of 35 years. The horizontal axis reports the ending date of the rolling window. For example 1965 corresponds to the sample February 1930-January 1965. The dependent variables are the average returns of the 25 Fama-French portfolios.

computed according to the formulas (8) and (9). The dependent variables are the average returns of the 25 Fama-French portfolios over the same time period. I drop the extreme small-growth portfolio that is often found to be an outlier in asset-pricing models.

The explanatory power of all models is very high at the beginning of the sample, and initially it tends to increase as the window moves to the right. However, past the 1970s the performance of the CAPM starts to quickly deteriorate, with a very visible drop around 1975. On the contrary, the performance of the unrestricted two-factor model remains substantially high, with values often above 80%. However, this model does not impose economically motivated restrictions on the premia, so it is not surprising that it delivers a higher R^2. The ICAPM does very well until the mid-1980s, even if its performance starts following a downward trend. By the mid-1990s, the R^2 starts fluctuating around 30%, very far from the 60% attained during the first half of the sample. However, as the window approaches the most recent financial crisis, the explanatory power of the ICAPM increases steeply and the R^2 touches 60%. This is a remarkable improvement in fit given that the last time that the ICAPM explanatory power crossed the 60% threshold was toward the end of 1978, and it has not been larger than 50% since the first half of 1985. Finally, it is worth pointing out that the performance of the CAPM does not show any significant recovery.

These results have some suggestive implications. First of all, they highlight the role played by the Great Depression and the Great Recession. Once these events are included in the analysis, the ICAPM performance substantially improves. Furthermore, the fact that the performance of the ICAPM improves, while the explanatory power of the CAPM remains unsatisfactory, implies that distinguishing between the two sources of risk is crucial and that this distinction becomes particularly meaningful in the aftermath of exceptional events.

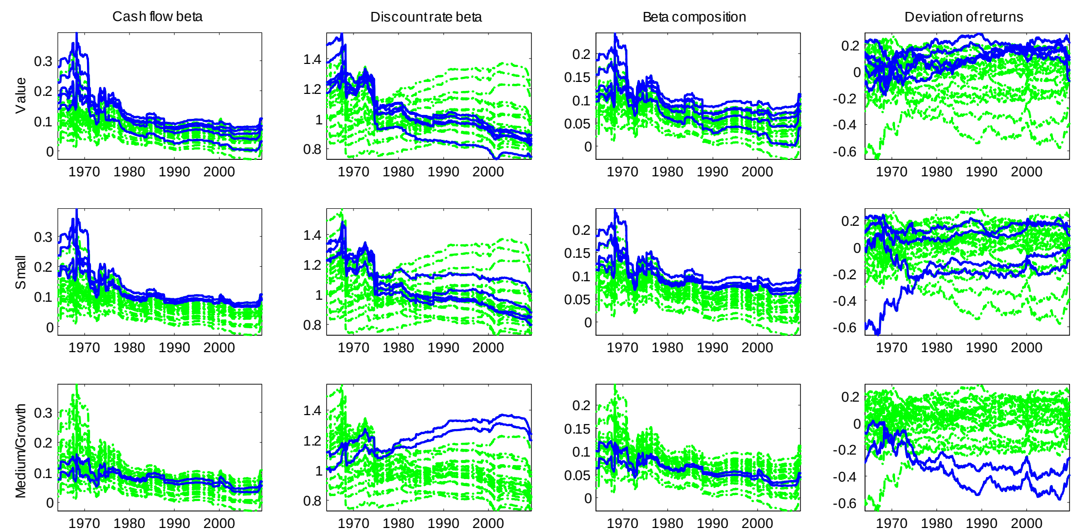

Figure 6: **Betas and predicted returns.** The first and second columns report cash flow betas and discount rate betas for the 25 Fama-French portfolios. The third column reports the composition of the market beta, and it is computed as the ratio between the cash-flow beta and the sum of cash-flow and discount-rate betas. Finally, the fourth column contains the deviations of portfolio returns from the values predicted by the Intertemporal CAPM. In each row, the solid blue lines refer to the portfolios indicated on the vertical axis of the first column.

To understand what drives the improvement in fit of the ICAPM, Figure 6 reports betas and deviations of portfolio returns from their predicted values for the 25 Fama-French portfolios. The first and second columns report cash-flow betas and discount-rate betas. The third column reports the composition of the market beta. This is computed as the ratio between the cash-flow beta and the sum of cash-flow and discount-rate betas. Recall that the market beta is obtained summing the two betas. Finally, the fourth column contains the deviations of portfolio returns with respect to the values predicted by the ICAPM. In each row, the solid blue lines refer to the portfolios indicated on the vertical axis of the figure.

A series of interesting patterns emerge. First, as in Campbell and Vuolteenaho (2004), over time, the value and small stocks experience a pronounced decline in the market beta with respect to the other portfolios. However, this decline is mostly driven by a fall in their discount rate betas, while their cash-flow betas remain on the upper side of the spectrum. This pattern implies a change in the composition of the market beta that in turn explains the success of the ICAPM over the CAPM. The ICAPM separates the different sources of risk associated with the two betas. Second, the most notable deviations of stock market returns from what is predicted by the ICAPM are caused by two medium/growth portfolios.[14] When analyzing the period antecedent the current crisis and excluding the Great Depression, these portfolios have stock market returns that are too low with respect to what is predicted by the ICAPM. While these stocks show a relatively large increase in their discount rate beta

[14] These two portfolios correspond to the ones with the second and third smallest market value among the five growth portfolios.

Sample	R^2	MPE	Risk Aversion
Pre-1964	50.54%	0.7454	4.8965
Post-1964	54.73%	0.5147	30.8862
Whole sample	60.58%	0.4593	11.3846

Table 2: Explanatory power of the ICAPM over different subsamples. The explanatory power of the ICAPM is assessed over three different samples: Pre-1964, post-1964, and the whole sample. The table reports R^2, mean pricing error, and the estimate for the coefficient of relative risk aversion.

and a stable composition for the beta, their average returns do not adequately reflect such an increase in risk. Finally, this anomaly is largely reduced toward the end of the sample, and at the same time, the returns of the small and value portfolios also move closer to their predicted returns.

From these results, we can infer that in order to adequately price the cross section of asset returns it is important to be able to observe the behavior of the assets during exceptional events such as the Great Recession. The relative performance of the different portfolios change substantially during these events. It is also important to emphasize that this is not the result of drastic changes in the betas. Even if we observe a partial increase in the cash-flow betas during the late years, the relative ranking of the portfolios with respect to the betas appears quite stable. It is therefore the change in the relative performance of the different portfolios during a time of distress that is largely responsible for the improvement in fit. In order to formalize this point, I regressed the average returns over the last window of time (June 1974-May 2009) on the betas computed using the window of time right before Bear Stearns received a loan from the Federal Reserve Bank of New York (April 1975-March 2008). The resulting R^2 is still high, 53.28%, even if lower than the value obtained aligning betas and average returns.

Summarizing, the analysis of the cross section of asset returns confirms the presence of similarities between the Great Depression and the Great Recession. The latter turned out to be, at least to date, a much less dramatic event. Nevertheless, it seems that both events are key to understanding the cross section of asset returns. To further corroborate this result, Table 2 breaks the sample into two parts, pre and post 1964. Notice that the R^2 is similar across the two subsamples and is quite high. Similarly, the R^2 computed over the whole sample is also very high. Therefore, the results suggest that as long as exceptional events are properly taken into account the ICAPM is able to correctly price the cross section of asset returns. Over the first subsample, the Great Depression plays a key role. Over the second subsample the Great Recession is enough to account for the behavior of the assets during rare events.

There are several possible explanations for why value stocks might perform worse during financial crises. Zhang (2005) argues that the value premium arises naturally in a neoclassical model because of costly reversibility and countercyclical price of risk. During bad times firms would find it optimal to disinvest, implying that assets in place are riskier than growth options. It seems reasonable that this distinction becomes particularly relevant during financial crises. Furthermore, almost by definition, value stocks include firms that markets believe might have less prospects of growth in the future. While this is not necessarily a problem during regular times, it can become a serious issue when credit availability is limited, real activity is low, and the price of risk is high. Similarly, Campbell and Vuolteenaho (2004) suggest that during the Great Depression and in its aftermath, value stocks might include a significant fraction of *fallen angels* that accumulated large amounts of debt during the crisis and were therefore inherently riskier.

5.4 Rare events and agents' expectations

In the previous subsection, I have analyzed in depth the role played by the Great Depression and the Great Recession in explaining the cross section of asset returns. A central result was that during the Great Recession, the fit of the ICAPM improved dramatically and went back to levels in line with the first half of the sample, suggesting that the presence of a rare event in the dataset is key to correctly pricing assets.

However, in the benchmark model, rare events also play another key role: they shape agents' expectations. This is because agents are assumed to be fully rational, and to take into account the possibility of regime changes when forming expectations. It is therefore interesting to ask what role this channel plays. In order to address this question, I reconsider the evolution of the explanatory power of the ICAPM under two different assumptions about the way agents form expectations. Under the benchmark model, agents take into account the possibility of regime changes. This corresponds to the case analyzed in the previous subsection. In the second case, agents form expectations according to the anticipated utility assumption. This implies that the probability assigned to the two regimes are not moving over time. Under this assumption, the series for the news are computed by replacing the estimated transition matrix H^{Φ} with the identity matrix in the formulas presented in Subsection 5.2. In other words, agents' beliefs are not evolving over time, but they are fixed to the filtered probabilities. Notice that a solution for the news still exists because the system turns to be mean square stable even under the anticipated utility assumption.

Figure 7 presents the results. The blue solid line corresponds to the benchmark case, while the black dashed line reports the results for the case of anticipated utility. It is interesting to note that the two lines align very closely over the early subsamples, but around the mid-1970s, they depart, and since then, the benchmark model always outperforms the alternative specification. It is worth recalling that over the very same months the CAPM also had a drastic decline in the fit (see Figure 5). In other words, exactly when the distinction between the CAPM and the ICAPM becomes more meaningful, we observe a discrete drop in the fit

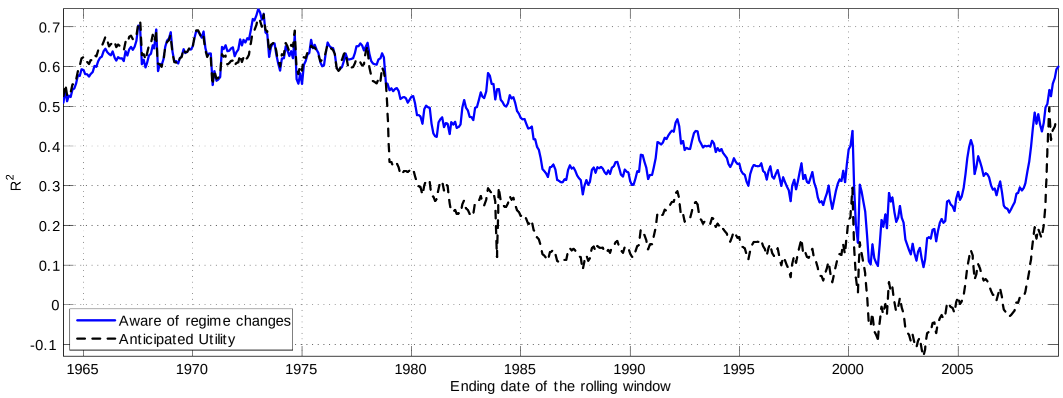

Figure 7: **The role of agents' beliefs.** The figure reports the explanatory power as measured by the R^2 for the ICAPM under two different assumptions about the way agents form expectations. The blue solid line corresponds to the benchmark case in which agents take into account the possibility of regime changes, while the black dashed line corresponds to the case in which agents form expectations according to the anticipated utility assumption. In this second case, the probability assigned to the two regimes are not moving over time. The betas and average returns are computed over moving windows of 35 years. The horizontal axis reports the ending date of the rolling window. For example, 1965 corresponds to the sample February 1930-January 1965. The dependent variables are the average returns of the 25 Fama-French portfolios.

of the ICAPM under the assumption of anticipated utility.[15] Finally, it is worth pointing out that the anticipated utility assumption becomes relatively more innocuous toward the end of the sample. This seems sensible given that this is the period of time during which the dynamics resembling the Great Depression present themselves. However, the gap in the explanatory power is still approximately 15%.

In summary, the Great Depression regime also plays a key role in accounting for the cross section of asset returns during regular times because it shapes the way agents form expectations. In fact, during regular times, it becomes particularly important to take into account the possibility of regime changes because no exceptional events are present over the sample. This is not enough to completely compensate for the lack of information, but it still determines an improvement in the fit of the ICAPM. In other words, it seems that the Great Depression played a key role in shaping agents' expectations about the behavior of financial variables during rare events.

6 Recursive Estimates

In the benchmark results presented above, I have used filtered probabilities to pin down agents' beliefs about entering the Great Depression regime and to compute the news. These

[15] Recall that the R^2 is computed as $1 - RSS/RSM$ where RSS is the residual sum of squares and RSM is the residual sum of squares when only the constant is used as a regressor. The ICAPM restricts the price of risk for the discount-rate beta to be equal to the variance of the market return. Therefore the R^2 can become negative. This is what happens over certain windows of time for the ICAPM when imposing the anticipated utility assumption.

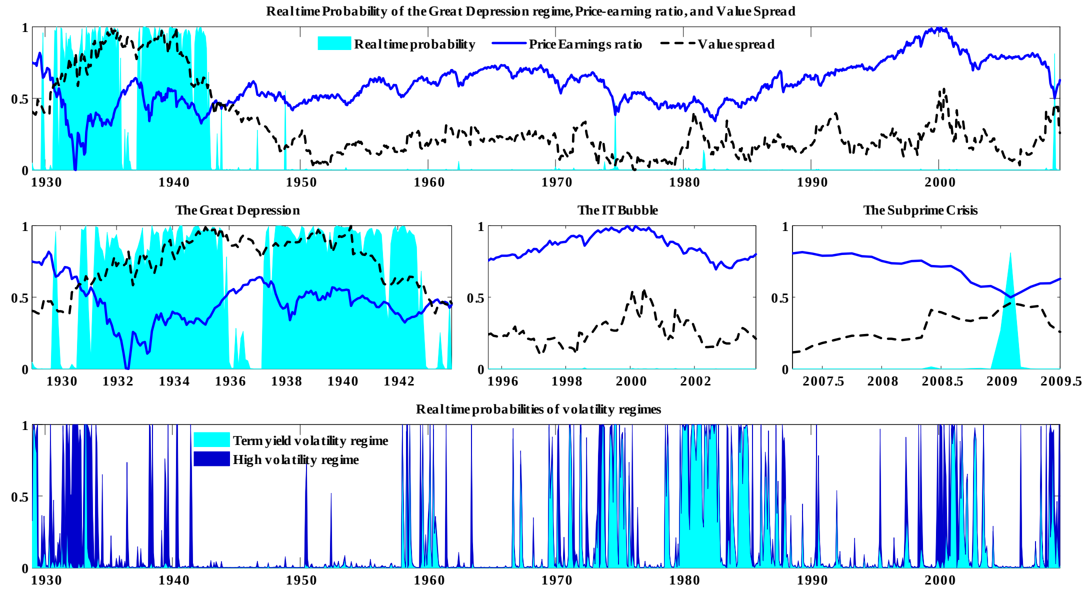

Figure 8: **Real time probabilities.** The figure reports the probabilities for the 1930 regime, the High volatility regime, and the Term yield volatility regime computed in real time starting with an initial sample spanning the period from December 1928 to January 1965. Then a month is added, the model is re-estimated, and the regime probabilities for that month are stored.

probabilities would represent the real time probabilities for an agent that has knowledge of all parameters of the model, but not of the regime in place. The choice of endowing the agent with the best possible estimates for the model parameters is consistent with the idea that in reality, agents have more information than the econometrician. Therefore, in using the whole sample, the econometrician is trying to obtain the most accurate estimates of what agents in fact know. It goes without saying that using the whole sample also improves the precision of the estimates.

An alternative approach would consist of assuming that the agents act as econometricians themselves, recursively estimating the MS-VAR as more data become available. I then conduct the following exercise. First, the MS-VAR is estimated over the sample December 1928-January 1965. The corresponding filtered probabilities and parameter estimates are used to compute the news over this initial subsample. Then a month at a time is added, with the result that the subsample keeps expanding until the whole sample is covered. For each of the expanded subsamples, the model is re-estimated, the *last* value for the regime probabilities and the corresponding parameter estimates are saved, and the news for this additional observation are computed and stored to reflect the updated estimates.[16]

Figure 8 reports the regime probabilities computed in real time for the Great Depression regime, the High volatility regime, and the Term Yield volatility regime. Even in this case, to facilitate the interpretation of the results, the periods corresponding to the Great Depression, the IT bubble, and the Great Recession are enlarged. Note that the results are similar to

[16] Accordingly, the priors are always set by only using the data available at each point in time.

Sample	R^2	MPE	Risk Aversion
Pre-1964	49.67%	0.7585	3.7665
Post-1964	45.24%	0.6226	14.1706
Whole sample	69.13%	0.3598	9.1992

Table 3: Explanatory power of the ICAPM over different subsamples based on news computed in real time. The explanatory power of the ICAPM is assessed over three different samples: Pre-1964, post-1964, and the whole sample. The table reports R^2, mean pricing error, and the estimate for the coefficient of relative risk aversion. The news are computed by using recursive estimates of the MS-VAR.

what is obtained when using the whole sample. In fact, the spike in the probability of the Great Depression regime at the beginning of 2009 is now even larger. In February 2009, the probability of the Great Depression regime computed in real time was 81.03%. This result reinforces the case in favor of the idea that agents might have feared a return to the Great Depression.

With respect to the filtered probabilities obtained using the whole sample, the only noticeable difference consists of an increase in the probability of the Great Depression regime in September 1974. When using the whole sample, the probability of the Great Depression regime during this month is 17.91%, while when using the recursive estimates this probability increases to reach 46.73%. Even in this case, the results can be rationalized in light of historical events. This is a period of time characterized by substantial uncertainty, induced by the end of the first oil shock, the resignation of President Nixon in August 1974 following the Watergate scandal, and the terrorist attack on the TWA Flight 841 from Tel Aviv to New York City after an intermediate stop in Athens.

Table 3 reports the results for the explanatory power of the ICAPM using the news computed in real time. Notice that the R^2 is still very large on both subsamples, even if somewhat lower than when computing the news using all the available information. However, the performance of the model over the whole sample is improved, with an R^2 close to 70%.

7 Conclusions

Using an MS-VAR, I have uncovered some key features that connect the behavior of the stock market during the Great Recession to the events that led to the Great Depression. I have identified a Great Depression regime and shown that its probability has been close to zero until the most recent recession. In February 2009, the probability of the Great Depression regime spiked to cross 50%, and it was larger than 80% when using real time estimates. During the early months of both the Great Depression and the Great Recession, the Value spread was increasing while the stock market was falling. However, during the Great Recession, this pattern eventually reverted, and the probability of the Great Depression regime experienced a sharp drop, arguably in response to robust government interventions, signaling that the

U.S. was in fact able to avoid a financial meltdown. Furthermore, a High volatility regime featuring a negative correlation between stock market returns and Value-spread innovations characterized both events.

Given that the Value spread measures the relative performance of small-growth stocks with respect to small-value stocks, the behavior of this variable during the Great Depression and the Great Recession suggest the idea that financial crises might have asymmetric effects on growth and value stocks. I formalized this idea by showing that the explanatory power of the Bad Beta, Good Beta ICAPM proposed by Campbell and Vuolteenaho (2004) improves significantly when the Great Recession is included in the analysis. This is because, conditional on the model, growth stocks performed relatively better *during* the Great Recession, justifying their lower returns *before* the Great Recession.

Finally, the explanatory power of the ICAPM is largely improved if agents form expectations taking into account the possibility of these events, even when the Great Recession and the Great Depression are not included in the cross-sectional regressions. Therefore, rare events also play a key role in shaping the expectations of fully rational agents. In light of their unusual statistical properties, rare events are indeed exceptional. Variables that tend to move in an apparently disconnected way suddenly reveal features that cannot be identified during regular times, when noise often dominates other sources of variation. This implies that while it is certainly important to study *rare-event betas*, *regular-time betas* can still be informative if rare events also have an impact on the way agents think about financial markets.

References

Ang, A., and G. Bekaert (2002): "Regime Switches in Interest Rates," *Journal of Business & Economic Statistics*, 20(2), 163–82.

Ang, A., and J. Chen (2007): "CAPM over the long run: 1926-2001," *Journal of Empirical Finance*, 14(1), 1–40.

Ang, A., and A. Timmermann (2012): "Regime Changes and Financial Markets," *Annual Review of Financial Economics*, 4(1), 313–337.

Backus, D., M. Chernov, and I. Martin (2011): "Disasters Implied by Equity Index Options," *Journal of Finance*, 66(6), 1969–2012.

Bai, H., K. Hou, H. Kung, and L. Zhang (2015): "The CAPM Strikes Back? An Investment Model with Disasters," LBS working paper.

Bansal, R., R. F. Dittmar, and C. T. Lundblad (2005): "Consumption, Dividends, and the Cross Section of Equity Returns," *Journal of Finance*, 60(4), 1639–1672.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of

Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

BARRO, R. (2009): "Rare Disasters, Asset Prices, and Welfare Costs," *The American Economic Review*, 99(1), 243–264.

BARRO, R. J. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121, 823–866.

BIANCHI, F. (2014): "Methods for Measuring Expectations and Uncertainty in Markov-switching Models," CEPR discussion paper 9705.

BIANCHI, F., M. LETTAU, AND S. C. LUDVIGSON (2013): "The Value of Forward Looking Communication," Forward Guidance: Perspectives from Central Bankers, Scholars and Market Participants, ed. W. den Haan, VoxEU.org Ebook, London.

BOLLERSLEV, T., AND V. TODOROV (2011): "Tails, Fears, and Risk Premia," *Journal of Finance*, 66(6), 2165–2211.

CAMPBELL, J. Y. (1991): "A Variance Decomposition for Stock Returns," *Economic Journal*, 101(405), 157–179.

——— (1993): "Intertemporal Asset Pricing Without Consumption Data," *American Economic Review*, 83, 487–512.

CAMPBELL, J. Y., S. GIGLIO, AND C. POLK (2013): "Hard Times," *Review of Asset Pricing Studies*, 3(1), 95–132.

CAMPBELL, J. Y., S. GIGLIO, C. POLK, AND R. TURLEY (2014): "An Intertemporal CAPM with Stochastic Volatility," Harvard working Paper.

CAMPBELL, J. Y., C. POLK, AND T. VUOLTEENAHO (2010): "Growth or Glamour? Fundamentals and Systematic Risk in Stock Returns," *Review of Financial Studies*, 23, 305–344.

CAMPBELL, J. Y., AND R. J. SHILLER (1988): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195–228.

CAMPBELL, J. Y., AND T. VUOLTEENAHO (2004): "Bad Beta, Good Beta," *American Economic Review*, 94, 1249–1275.

CHEN, H., S. JOSLIN, AND N.-K. TRAN (2012): "Rare Disasters and Risk Sharing with Heterogeneous Beliefs," *Review of Financial Studies*, 25(7), 2189–2224.

CHEN, L., AND X. ZHAO (2005): "Return decomposition," Michigan State and Kent State University, mimeo.

COGLEY, T., AND T. J. SARGENT (2006): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," *Review of Economic Dynamics*, 8, 262–302.

——— (2007): "The Market Price of Risk and the Equity Premium: A Legacy of the Great Depression?," *Journal of Monetary Economics*, 55(3), 454–476.

COSTA, O., M. FRAGOSO, AND R. MARQUES (2004): *Discrete-Time Markov Jump Linear Systems*. Springer, New York.

DANIEL, K., AND S. TITMAN (2006): "Testing Factor-Model Explanations of Market Anomalies," The Kellogg School at Northwestern University and NBER, mimeo.

Engsted, T., T. Q. Pedersen, and C. Tanggaard (2010): "Pitfalls in VAR Based Return Decompositions: A Clarification," working paper, University of Aarhus.

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57(4), 937–69.

Fama, E. F., and K. R. French (1992): " The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47(2), 427–65.

Fama, E. F., and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1), 3–56.

Gabaix, X. (2012): "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *Quarterly Journal of Economics*, 127(2), 645–700.

Geweke, J. (1992): "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," *Bayesian Statistics*, 4, 169–193.

Gourio, F. (2012): "Disaster Risk and Business Cycles," *American Economic Review*, 102(6), 2734–66.

Gulen, H., Y. Xing, and L. Zhang (2011): "Value versus Growth: TimeâĂŘVarying Expected Stock Returns," *Financial Management*, 40(2), 381–407.

Julliard, C., and A. Ghosh (2012): "Can Rare Events Explain the Equity Premium Puzzle?," *Review of Financial Studies*, 25(10), 3037–3076.

Kim, C.-J., and C. R. Nelson (1999a): "Has The U.S. Economy Become More Stable? A Bayesian Approach Based On A Markov-Switching Model Of The Business Cycle," *The Review of Economics and Statistics*, 81(4), 608–616.

Kim, C.-J., and C. R. Nelson (1999b): *State-Space Models with Regime Switching*. MIT Press, Cambridge, Massachusetts.

Koijen, R. S., H. Lustig, and S. V. Nieuwerburgh (2014): "The Cross-Section and Time-Series of Stock and Bond Returns," NBER Working Papers 15688, National Bureau of Economic Research, Inc.

Krugman, P. R. (2009): "Fighting Off Depression," New York Times OP-ED, January 4, 2009.

Lettau, M., and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56(3), 815–849.

Lettau, M., S. C. Ludvigson, and J. A. Wachter (2008): "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?," *Review of Financial Studies*, 21(4), 1653–1687.

Lewellen, J., S. Nagel, and J. Shanken (2010): "A Skeptical Appraisal of Asset-Pricing Tests," *Journal of Financial Economics*, 96(2), 175–194.

Martin, I. (2013a): "The Lucas Orchard," *Econometrica*, 81(1), 55–111.

Martin, I. W. (2013b): "Consumption-Based Asset Pricing with Higher Cumulants," *Re-*

view of Economic Studies, 80(2), 745–773.

Merton, R. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41(4), 867–887.

Nakamura, E., J. Steinsson, R. Barro, and J. Ursua (2013): "Crises and Recoveries in an Empirical Model of Consumption Disasters," *American Economic Journal: Macroeconomics*, 5(3), 35–74.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73, 1057–1084.

Primiceri, G. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852.

Raftery, A., and S. Lewis (1992): "How many iterations in the Gibbs sampler?," *Bayesian Statistics*, 4, 763–773.

Rietz, T. A. (1988): "The Equity Risk Premium: A Solution," *Journal of Monetary Economics*, 22, 117–131.

Sandroni, A. (1998): "Learning, Rare Events, and Recurrent Market Crashes in Frictionless Economies without Intrinsic Uncertainty," *Journal of Economic Theory*, 82(1), 1–18.

Shiller, R. J. (2000): *Irrational exuberance*. Princeton University Press.

Sims, C. A., and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 949–968.

——— (2006): "Were There Regime Switches in US Monetary Policy?," *American Economic Review*, 91(1), 54–81.

Solomon, D. (2009): "Market Pans Bank Rescue Plan," Wall Street Journal, 11 February 2009.

Veronesi, P. (2004): "The Peso Problem Hypothesis and Stock Market Returns," *Journal of Economic Dynamics and Control*, 28(4), 707–725.

Wachter, J. A. (2013): "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?," *Journal of Finance*, 68(3), 987–1035.

Zhang, L. (2005): "The Value Premium," *Journal of Finance*, 60(1), 67–103.

A Priors

Table 4 describes the priors used for the estimation of the MS-VAR. The priors are very loose and symmetric across regimes. Below, I describe more in detail how they have been obtained. The assumption of covariance stationarity implies a truncation of the priors as described in the table. The truncated prior is implemented by dropping the draws that imply non-stationarity. However, in the estimates the constraint implied by the truncation of the priors is rarely binding.

The priors for the VAR coefficients and the covariance matrix are symmetric across regimes and are obtained running univariate autoregressions for each endogenous variable:

$$y_{i,t} = a_i y_{i,t-1} + v_t \sigma_i$$

The prior for the VAR coefficients is:

$$B = vec\left(\Phi_{\xi_t^\Phi}\right) \sim norm\left(B_0, S_0 \otimes N_0^{-1})\right)$$

The autoregressive elements of B_0 are equal to the AR(1) coefficients, while all the other elements are set to zero. As in Sims and Zha (1998), the variance of the prior distribution is specified by a number of hyperparameters that pin down N_0. The choice of hyperparameters implies a fairly loose prior for the VAR coefficients. Let λ be a (5×1) vector containing the hyperparameters. The diagonal elements of N_0^{-1} corresponding to autoregressive coefficients are given as $\left(\frac{\lambda_0 \lambda_1}{\sigma_i l^{\lambda_3}}\right)^2$ where σ_j denotes the variance of the error from the AR regression for the jth variable and $l = 1...L$ denotes the lags in the VAR ($L = 1$ in the models considered in this paper). The intercept terms in N_0^{-1} are controlled by the term $(\lambda_0 \lambda_4)^2$. The choice for the hyperparameters are $\lambda_0 = 1$, $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 0.5$ and $\lambda_4 = 1$ and $S_0 = V_0 diag\left(\{\sigma_i^2\}_{i=1...n}\right)$, with $V_0 = 9$. The priors for the covariance matrices are symmetric across regimes and described by an inverse Wishart distribution with mean $S_0 = V_0 diag\left(\{\sigma_i^2\}_{i=1...n}\right)$, with $V_0 = 9$: $\Sigma_{\xi_t^\Sigma} \sim IW(S_0, V_0)$.

Each column of H^Φ and H^Σ is modeled according to a Dirichlet distribution whose properties are described in Table 4: $H^s(\cdot, i) \sim D(a_{ii}^s, a_{ij}^s)$, $s = \Phi, \Sigma$. I choose $a_{ii}^\Sigma = 10$, $a_{ij}^\Sigma = 2$, $a_{ii}^\Phi = 80$, $a_{ij}^\Phi = 2$. Note that the priors for the transition matrices are symmetric across regimes. I also estimated looser priors for the transition matrices and obtained very similar results.

$\xi_t^\Phi = 1,2$	ER_t	TY_t	PE_t	VS_t	$const$
ER_{t+1}	0.1236 (−0.0857,0.3316)	0 (−0.0216,0.0216)	0 (−0.1357,0.1357)	0 (−0.1552,0.1552)	0.0049 (−0.0331,0.0429)
TY_{t+1}	0 (−1.3427,1.3427)	0.9472 (0.8084,1.0866)	0 (−0.8765,0.8765)	0 (−1.0012,1.0012)	0.0823 (−0.1620,0.3265)
PE_{t+1}	0 (−0.2140,0.2140)	0 (−0.0221,0.0221)	0.9888 (0.8497,1.1282)	0 (−0.1593,0.1593)	0.0324 (−0.0066,0.0712)
VS_{t+1}	0 (−0.1868,0.1868)	0 (−0.0195,0.0195)	0 (−0.1218,0.1218)	0.9909 (0.8514,1.1303)	0.0147 (−0.0193,0.0488)

$\xi_t^\Sigma = 1,2,3$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0546 (0.0797,0.1981)	0 (−0.0610,0.0610)	0 (−0.0097,0.0097)	0 (−0.0085,0.0085)
u_{TY}	0 (−0.5660,0.5660)	0.3515 (0.5124,1.2733)	0 (−0.0624,0.0624)	0 (−0.0545,0.0545)
u_{PE}	0 (−0.5660,0.5660)	0 (−0.5660,0.5600)	0.0559 (0.0814,0.2024)	0 (−0.0086,0.0086)
u_{VS}	0 (−0.5660,0.5660)	0 (−0.5660,0.5660)	0 (−0.5660,0.5660)	0.0489 (0.0714,0.1773)

H^Φ	$\xi_t^\Phi = 1$	$\xi_t^\Phi = 2$
$\xi_{t+1}^\Phi = 1$	0.9875 (0.9599,0.9912)	0.0125 (0.0088,0.0401)
$\xi_{t+1}^\Phi = 2$	0.0125 (0.0088,0.0401)	0.9875 (0.9599,0.9912)

H^Σ	$\xi_t^\Sigma = 1$	$\xi_t^\Sigma = 2$	$\xi_t^\Sigma = 3$
$\xi_{t+1}^\Sigma = 1$	0.8182 (0.5954,0.8327)	0.0909 (0.0553,0.2315)	0.0909 (0.0553,0.2315)
$\xi_{t+1}^\Sigma = 2$	0.0909 (0.0553,0.2315)	0.8182 (0.5952,0.8324)	0.0909 (0.0553,0.2315)
$\xi_{t+1}^\Sigma = 3$	0.0909 (0.0553,0.2315)	0.0909 (0.0553,0.2315)	0.8182 (0.5955,0.8324)

Table 4: Priors for the parameters. The three sets of tables contain modes and 68% error bands for the priors of the parameters of the Markov-switching VAR. The priors are obtained running univariate autoregressions for each of the variables in the model, and they are symmetric across regimes.

B Likelihood and regime probabilities

Define the combined regime $\xi_t \equiv \left(\xi_t^\Phi, \xi_t^\Sigma\right)$, the associated transition matrix $H \equiv H^\Phi \otimes H^\Sigma$, and vector $\theta_{\xi_t} \equiv \left(\Phi_{\xi_t^\Phi}, \Sigma_{\xi_t^\Sigma}\right)$ with the corresponding set of parameters. For each draw of the parameters θ_{ξ_t} and H, we can then compute the filtered probabilities $\pi_{t|t}$, or smoothed probabilities $\pi_{t|T}$, of the regimes conditional on the model parameters. The filtered probabilities reflect the probability of a regime conditional on the data up to time t, $\pi_{t|t} = p(\xi_t|Y^t; H, \theta_{\xi_t})$, for $t = 1, ..., T$, and are part of the output obtained computing the likelihood function associated with the parameter draw H, θ_{ξ_t}. The filtered probabilities can be obtained using the following recursive algorithm:

$$\pi_{t|t} = \frac{\pi_{t|t-1} \odot \eta_t}{\mathbf{1}' \left(\pi_{t|t-1} \odot \eta_t\right)} \quad (17)$$

$$\pi_{t+1|t} = H \pi_{t|t} \quad (18)$$

$$p(Z_t|Z^{t-1}) = \mathbf{1}' \left(\pi_{t|t-1} \odot \eta_t\right) \quad (19)$$

where η_t is a vector whose jth element contains the conditional density $p(Z_t|\xi_t = i, Z^{t-1}; H, \theta_{\xi_t})$, the symbol $\odot$ denotes element by element multiplication, and $\mathbf{1}$ is a vector with all elements equal to 1. To initialize the recursive calculation, we need an assumption on the distribution of ξ_0. We assume that the six regimes have equal probabilities $p(\xi_0 = i) = 1/6$ for $i = 1...m$.

The likelihood for the entire data sequence Z^T is obtained multiplying the one-step-ahead conditional likelihoods $p(Z_t|Z^{t-1})$:

$$p\left(Z^T|\theta\right) = \prod_{t=1}^{T} p\left(Z_t|Z^{t-1}\right)$$

The smoothed probabilities reflect all the information that can be extracted from the whole data sample, $\pi_{t|T} = p(\xi_t|Z^T; H, \theta_{\xi_t})$. The final term $\pi_{T|T}$ is returned with the final step of the filtering algorithm. Then a recursive algorithm can be implemented to derive the other probabilities:

$$\pi_{t|T} = \pi_{t|t} \odot \left[H' \left(\pi_{t+1|T} \left(\div\right) \pi_{t+1|t}\right)\right]$$

where $(\div)$ denotes element by element division.

Finally, it is possible to obtain the filtered and smoothed probabilities for each of the two independent chains by integrating out the other chain. For example, if we are interested in $\pi^{\Phi}_{t|t} = p(\xi^{\Phi}_t|Y^t; H, \theta_{\xi_t})$ we have:

$$\pi^{\Phi,i}_{t|t} = p(\xi^{\Phi}_t = i|Y^t; H, \theta_{\xi_t}) = \sum_{j=1}^{m} p(\xi_t = \{i, j\}|Y^t; H, \theta_{\xi_t})$$

Similarly, the smoothed probabilities are obtained as:

$$\pi^{\Phi,i}_{t|T} = p(\xi^{\Phi}_t = i|Y^T; H, \theta_{\xi_t}) = \sum_{j=1}^{m} p(\xi_t = \{i, j\}|Y^T; H, \theta_{\xi_t}).$$

C Gibbs sampling algorithm

Both the VAR coefficients and the covariance matrix can switch and the regimes are assumed to be independent. Draws for the parameters of the model can be made following the following Gibbs sampling algorithm:

1. Sampling ξ^{Φ}_t and ξ^{Σ}_t given $\Phi_{\xi^{\Phi}_t}, \Sigma_{\xi^{\Sigma}_t}, H^{\Phi}, H^{\Sigma}$: Following Kim and Nelson (1999b) I use a Multi-Move Gibbs sampling to draw ξ^{Φ}_t from $f(\xi^{\Phi}_t|Z^T, \Phi_{\xi^{\Phi}_t}, \Sigma_{\xi^{\Sigma}_t}, H^{\Phi}, H^{\Sigma}, \xi^{\Sigma}_t)$ and ξ^{Σ}_t from $f(\xi^{\Sigma}_t|Z^T, \Phi_{\xi^{\Phi}_t}, \Sigma_{\xi^{\Sigma}_t}, H^{\Phi}, H^{\Sigma}, \xi^{\Phi}_t)$.

2. Sampling $\Sigma_{\xi^{\Sigma}_t}$ given $\Phi_{\xi^{\Phi}_t}, \xi^{\Phi}_t, \xi^{\Sigma}_t$: Given $\Phi_{\xi^{\Phi}_t}$ and $\xi^{\Phi,T}$, we can compute the residuals of the MS-VAR at each point in time. Then, given ξ^{Σ}_t, we can group all the residuals that pertain to a particular regime. Therefore, $\Sigma_{\xi^{\Sigma}_t}$ can be drawn from an inverse Wishart distribution for $\xi^{\Sigma}_t = 1...m^{\Sigma}$.

3. Sampling $\Phi_{\xi^{\Phi}_t}$ given $\Sigma_{\xi^{\Sigma}_t}, \xi^{\Phi}_t, \xi^{\Sigma}_t$: When drawing the VAR coefficients, we need to take into account the heteroskedasticity implied by the switches in $\Sigma_{\xi^{\Sigma}_t}$. This can be done following the following steps for each $i = 1...m^{\Phi}$:

(a) Based on $\xi^{\Phi,T}$, collect all the observation such that $\xi_t^{\Phi} = i$.

(b) Divide the data that refer to $\xi_t^{\Phi} = i$ based on $\xi^{\Sigma,T}$. We now have a series of subsamples for which VAR coefficients are covariance matrices are fixed: $\left(\xi_t^{\Phi} = i,\ \xi_t^{\Sigma} = 1\right)$, ..., $\left(\xi_t^{\Phi} = i,\ \xi_t^{\Sigma} = m^{\Sigma}\right)$. Denote these subsamples with $\left(y_{i,\xi_t^{\Sigma}},\ x_{i,\xi_t^{\Sigma}}\right)$ where the $y_{i,\xi_t^{\Sigma}}$ and $x_{i,\xi_t^{\Sigma}}$ denote left-hand-side and right-hand-side variables in the MS-VAR. Notice that some of these subsamples might be empty.

(c) Apply recursively the formulas for the posterior of VAR coefficients conditional on a known covariance matrix. Therefore, for $j = 1...m^{\Sigma}$ the following formulas need to be applied recursively:

$$\begin{aligned} P_T^{-1} &= P_L^{-1} + \Sigma_{\xi_t^{\Sigma}}^{-1} \otimes (x'_{i,\xi_t^{\Sigma}} x_{i,\xi_t^{\Sigma}}) \\ B_T &= B_L + (\Sigma_{\xi_t^{\Sigma}}^{-1} \otimes x'_{i,\xi_t^{\Sigma}}) vec(y_{i,\xi_t^{\Sigma}}) \\ P_L^{-1} &= P_T^{-1}, B_L = B_T \end{aligned}$$

where the algorithm is initialized using the priors for the VAR coefficients $B_L = B_0$ and $P_L^{-1} = P_0^{-1} = \left(S_0 \otimes N_0^{-1}\right)^{-1}$. Notice that this implies that if there are not any observations for a particular regime, then the posterior will coincide with the priors. With proper priors, this is not a problem.

(d) Make a draw for the VAR coefficients $vec\left(\Phi_{\xi_t^{\Phi}}\right) \sim N\left(P_T B_T, P_T\right)$ with $\xi_t^{\Phi} = i$.

4. Sampling H^{Φ} and H^{Σ}: Given the draws for the state variables $\xi^{\Phi,T}$ and $\xi^{\Sigma,T}$, the transition probabilities are independent of Y_t and the other parameters of the model and have a Dirichlet distribution. For each column of H^{Φ} and H^{Σ}, the posterior distribution is given by:

$$H^s(:,i) \sim D(a_{ii}^s + \eta_{ii}^s, a_{ij}^s + \eta_{ij}^s),\ \ s = \Phi, \Sigma$$

where η_{ij}^{Φ} and η_{ij}^{Σ} denote respectively the numbers of transitions from state i^{Φ} to state j^{Φ} and from state i^{Σ} to state j^{Σ}.

D Properties of the regimes

Figure 9 reports the distribution for the difference between the parameter of the VAR coefficients. Figure 10 reports the distribution for the difference between the elements of the covariance matrix under the Term Yield volatility regime and the Low volatility regime. Finally, Figure 11 reports the difference between the elements of the covariance matrix under the Term Yield volatility regime and the High volatility regime. These are computed taking

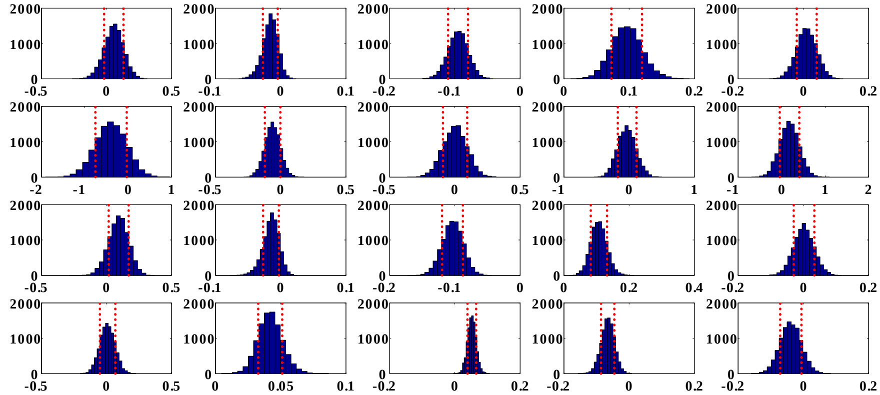

Figure 9: The figure contains histograms and 68% error bands for the pairwise differences of the VAR coefficients across the two regimes. This can be regarded as a "test" for the null hypothesis that the two parameters are the same across the two regimes.

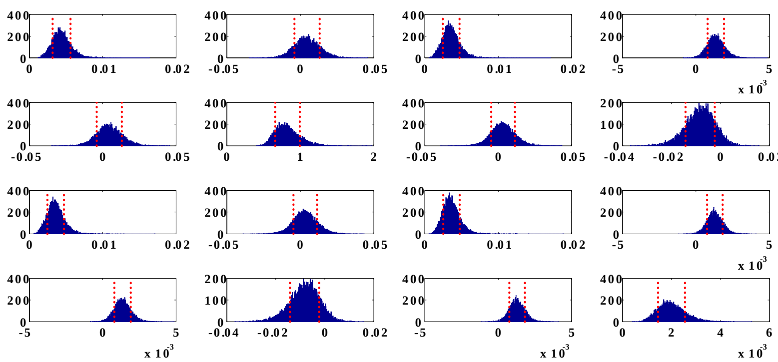

Figure 10: The figure contains histograms and 68% error bands for the pairwise differences of the covariance matrix under the Term Yield Volatility regime and the Low volatility regime. This can be regarded as a "test" for the null hypothesis that the two parameters are the same across the two regimes.

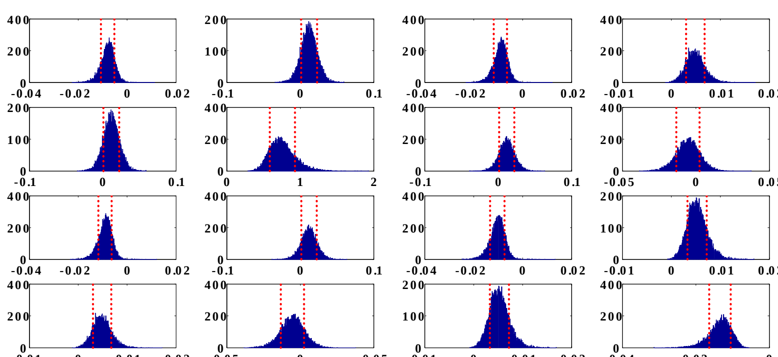

Figure 11: The figure contains histograms and 68% error bands for the pairwise differences of the covariance matrix under the Term Yield Volatility regime and the High volatility regime. This can be regarded as a "test" for the null hypothesis that the two parameters are the same across the two regimes.

the difference between the corresponding parameters for each draw from the Gibbs sampling algorithm.

E Cash-flow and discount rate news with regime changes

Consider an MS-VAR:

$$Z_t = c_{\xi_t} + A_{\xi_t} Z_{t-1} + R_{\xi_t} \Sigma_{\xi_t} \varepsilon_t$$

where Z_t is a column vector containing n variables observable at time t and $\xi_t = 1, ..., m$, with m the number of regimes, evolves following the transition matrix H.

Define the column vectors q_t and π_t:

$$q_t = \left[q_t^{1\prime}, ..., q_t^{m\prime}\right]', q_t^i = \mathbb{E}_0\left(Z_t 1_{\xi_t = i}\right), \pi_t = \left[\pi_t^1, ..., \pi_t^m\right]',$$

where $\pi_t^i = P_0(\xi_t = i)$ and $1_{\xi_t=i}$ is an indicator variable that is equal to 1 when regime i is in place and zero otherwise. The law of motion for $\widetilde{q}_t = [q_t', \pi_t']'$ is then given by

$$\underbrace{\begin{bmatrix} q_t \\ \pi_t \end{bmatrix}}_{\widetilde{q}_t} = \underbrace{\begin{bmatrix} \Omega & CH \\ & H \end{bmatrix}}_{\widetilde{\Omega}} \begin{bmatrix} q_{t-1} \\ \pi_{t-1} \end{bmatrix} \tag{20}$$

where $\pi_t = [\pi_{1,t}, ..., \pi_{m,t}]'$, $\Omega = bdiag(A_1, ..., A_m) H$, and $C = bdiag(c_1, ..., c_m)$. Recall that:

$$\mathbb{E}_0(Z_t) = \sum_{i=1}^{m} q_t^i = w q_t,\ w = \left[\underbrace{I_n, ..., I_n}_{m}\right]$$

To compute the news, define:

$$\begin{aligned} q_{t+s|t}^i &= \mathbb{E}_t\left(Z_{t+s} 1_{\xi_{t+s}=i}\right) = \mathbb{E}\left(Z_{t+s} 1_{\xi_{t+s}=i} | \mathbb{I}_t\right) \\ e_1' &= [1, 0, 0, 0]',\ mn = m * n \end{aligned}$$

where $\mathbb{I}_t$ contains all the information that agents have at time t, including the probability of being in one of the m regimes. Note that $q_{t|t}^i = Z_t \pi_t^i$.

Now consider the formula for the discount rate news:

$$N_{DR,t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t) \textstyle\sum_{j=1}^{\infty} \rho^j r_{t+1+j}$$

The first term is:

$$\begin{aligned}
\mathbb{E}_{t+1}\sum_{j=1}^{\infty}\rho^{j}r_{t+1+j} &= \sum_{j=1}^{\infty}\rho^{j}e_1'wq_{t+1+j|t+1} \\
&= e_1'w\left[\rho q_{t+2|t+1}+\rho^{2}q_{t+3|t+1}+\rho^{3}q_{t+4|t+1}+....\right] \\
&= e_1'w\left(I_r-\rho\Omega\right)^{-1}\left[\rho\Omega q_{t+1|t+1}+\rho CH\left(I_r-\rho H\right)^{-1}\pi_{t+1|t+1}\right]
\end{aligned}$$

The second term is:

$$\begin{aligned}
\mathbb{E}_{t}\sum_{j=1}^{\infty}\rho^{j}r_{t+1+j} &= \sum_{j=1}^{\infty}\rho^{j}e_1'wq_{t+1+j|t} \\
&= e_1'w\left(I_r-\rho\Omega\right)^{-1}\left[\rho\Omega q_{t+1|t}+\rho CH\left(I_r-\rho H\right)^{-1}\pi_{t+1|t}\right]
\end{aligned}$$

Therefore:

$$\begin{aligned}
N_{DR,t+1} &= e_1'w\left[\lambda^{q}v_{t+1}^{q}+\lambda^{\pi}v_{t+1}^{\pi}\right] \\
\lambda^{q} &= \left(I_r-\rho\Omega\right)^{-1}\rho\Omega \\
\lambda^{\pi} &= \left(I_r-\rho\Omega\right)^{-1}\rho CH\left(I_r-\rho H\right)^{-1}
\end{aligned}$$

Then, we can easily compute the residuals:

$$\begin{aligned}
u_{t+1} &= Z_{t+1}-\mathbb{E}_{t}Z_{t+1} \\
e_1'u_{t+1} &= r_{t+1}-\mathbb{E}_{t}\left(r_{t+1}\right)
\end{aligned}$$

and the news about future cash flows can be obtained as:

$$N_{CF,t+1}=e_1'u_{t+1}+N_{DR,t+1}$$

Note that given a sequence of probabilities or a draw for the MS states and a set of parameters, it is easy and computationally efficient to compute the entire sequences $v^{q,T}$, $v^{\pi,T}$, and u^{T}:

$$\begin{aligned}
N_{DR}^{T} &= e_1'w\left[\lambda^{q}v^{q,T}+\lambda^{\pi}v^{\pi,T}\right] \\
N_{CF}^{T} &= e_1'w\left[\left(I_r+\lambda^{q}\right)v^{q,T}+\lambda^{\pi}v^{\pi,T}\right] \\
u^{T} &= e_1'wv^{q,T}
\end{aligned}$$