Rare Events, Financial Crises, and the Cross-Section of Asset Returns

Bianchi, Francesco

Duke University

1 January 2008

Rare Events, Financial Crises, and the Cross-Section of Asset Returns*

Francesco Bianchi
Duke University

January 2010

Abstract

This paper shows that rare events are important in explaining the cross section of asset returns because of their role in shaping agents' expectations. I reconsider the "bad beta, good beta" ICAPM proposed by Campbell and Vuolteenaho and I point out that the explanatory power of the model relies on including the stock market crash that opened the Great Depression. When using a Markov-switching VAR, a *'30s regime* is identified. This regime receives a large weight when forming expectations consistent with the ICAPM, suggesting that the way agents think about financial markets is shaped by what happens during extreme circumstances. From a technical point of view, the paper extends the present value decomposition of Campbell and Shiller to allow for Markov-switching dynamics in the law of motion of the state variables. This approach could shed new light on the sensitivity of the present value decomposition methodology to the sample choice.

1 Introduction

Financial market practitioners would agree that extreme financial crises are characterized by peculiar dynamics that set them apart from what happens on a daily basis. This paper

*I am grateful to Chris Sims for useful suggestions at the early stage of this work. I thank Robert Barro, Markus Brunnermeier, John Campbell, Jakub Jurek, Alisdair McKay, Motohiro Yogo, Adam Zawadowski, Barbara Rossi, and seminar participants at Princeton University, Duke University, the Board of Governors of the FRS, the EEA-ESEM conference, Bank of Italy, and the CEF 2008 conference for helpful discussions and comments. This paper was awarded the *Society for Computational Economics Student Prize.*
Correspondence: 213 Social Sciences Building, Duke University, Durham, NC, 27708-0097, USA. E-mail: francesco.bianchi@duke.edu.

shows that this is in fact the case and that taking these dynamics into account is important in understanding the cross section of asset returns. Particular emphasis will be put on the stock market crash that opened the Great Depression, providing support for the idea that rare events shape agents' expectations. In doing so, I will provide formulas to extend the present value decomposition of Campbell and Shiller (1988) to accommodate for the state variables evolving according to a Markov-switching process.

The macro-finance literature reveals a growing interest in the role that rare disasters can play in understanding asset returns. Barro (2006, 2009), following Rietz (1988), shows that rare events are potentially important in explaining several asset-pricing puzzles, such as the high equity premium, low risk-free rate, and volatile stock returns, and that the welfare costs of rare disasters can be very large for reasonable parameter values.

The idea that rare disasters should imply an increase in the risk premium sounds reasonable and conceptually it is not different from the argument underlying the Consumption CAPM: An asset is valuable if it pays during bad times, when the marginal utility of consumption is very high. If the stock market performs poorly during a disaster, it is fair to expect a high equity premium. Hence, if the probability of a rare event is taken into account, the high risk premium won't be mistakenly ascribed to unreasonably high levels of risk aversion. In fact, Barro et al. (2009), estimating an empirical model of consumption disaster with Bayesian methods, show that under Epstein-Zin-Weil preferences rare disasters can rationalize a sizeable equity premium for modest values of risk aversion.

Gabaix (2007) extends Barro's results to allow for a time-varying intensity of rare disasters. Among other things, he suggests that the small-value premium could be compensation for distress risk if value stocks did worse than growth stocks during disasters. Irrespective of how appealing this argument might sound, it seems at odds with the findings of Julliard and Ghosh (2008). They show that rare events cannot account for the Equity Premium Puzzle and the cross section of asset returns at the same time. On the one hand, the rare events explanation of the EPP requires an increase in the probability assigned to disasters. On the other hand, all stocks behave quite poorly during such events. Therefore, the rare-event explanation of the EPP significantly worsens the ability of the Consumption-CAPM to explain the cross-section of asset returns because it implies a reduction in the cross-sectional dispersion of consumption risk compared to the cross-sectional variation of average returns.

Proposing an alternative approach, this paper shows that rare events indeed matter for the cross section of asset returns because they have a large impact on the way agents think about financial markets. So far, the literature has focused on the behavior of assets during rare disasters, but it has neglected to address the role that rare events can play in shaping agents' expectations. However, several economic variables of interest exhibit strong comovements during extraordinary events, a feature that makes them relatively easy for agents to interpret. Moreover, rare events come with remarkable effects on the performance of the stock market and on the overall state of the economy. Hence, in an uncertain environment such as the one characterizing financial markets, it could be that agents form expectations relying heavily on what happens in extreme circumstances.

This argument has an important implication. If the only thing that matters is the way that assets behave during disasters, *regular times betas* could be completely uninformative about the risk premia, given that they would not reflect the risk associated with the different portfolios, and only *rare disasters betas* would reveal the information needed to price them. But if rare disasters did affect the expectation formation mechanism, then *regular times betas* would still be informative as long as expectations have been modeled adequately, i.e. as long as the role of rare events has been taken into account.

This paper provides support for such a thesis by showing that, exactly because of their impact on the expectation mechanism, rare events play an important role in explaining the cross section of asset returns. I reconsider the "bad beta, good beta" Intertemporal CAPM (ICAPM) proposed by Campbell and Vuolteenaho (2004a). 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 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. Campbell and Vuolteenaho show that the ICAPM is able to account for the small-value anomalies that arise starting from the early '60s.

The first contribution of this paper is to point out that this remarkable result relies on whether the dataset includes the stock market crash that opened the Great Depression. Campbell and Vuolteenaho implement the decomposition of the expected market return into cash-flow news and discount rate news using a VAR model estimated on the sample 1928:12-2001:12. They then split the sample into two parts: pre- and post-63. Figure 1 reports the evolution of R^2 of the ICAPM[1] for the post-63 subsample as the *full* sample used in estimating the initial VAR shortens. For example, 1940:12 means that cash-flow and discount-rate news have been computed according to a VAR estimated over the sample 1940:12-2001:12, with the ICAPM always tested over the same subsample (1963:07-2001:12).

[1] This 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. This is why R^2 can become negative.

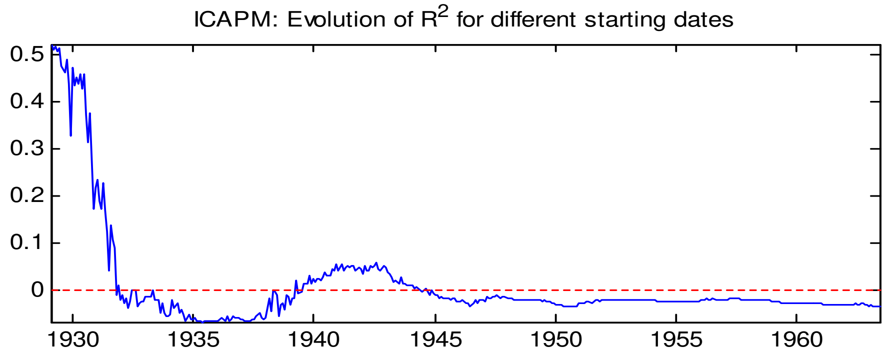

Figure 1: Evolution of R^2 for the ICAPM as the sample used to compute cash-flow and discount rate news shortens. For example: 1940:12 means that cash-flow and discount-rate news have been computed according to a VAR estimated over the sample 1940:12-2001:12. While the sample used to estimate the VAR is changing, the models are always tested over the same subsample (1963:07-2001:12).

Excluding the first three years has a substantial effect on the explanatory power of the ICAPM. When the entire sample is used, $R^2 = 51.96\%$. However, once the first three years are excluded these results disappear and $R^2 \simeq 0\%$.

Many economic models require researchers to take a stance on how agents form expectations. This implies that when testing these models two sets of assumptions are under the lens of the researcher. The first one is model specific, while the second refers to how expectations have been modeled. In fact, a model could be rejected because the expectation mechanism is not captured adequately. This is what seems to happen once the early '30s are excluded from the estimates. Removing these first years means getting rid of the dramatic market crash that marked the beginning of the Great Depression. If investors form expectations giving a large weight to what happens in extreme circumstances, it is not surprising that the events of the early '30s have such a large impact on the performance of the ICAPM. The Great Depression market crash is likely to convey information that is crucial in modeling the agents' expectation formation mechanism and this is why it is important to include those years in the estimates.

The second contribution of this paper is to formalize this appealing argument. Using a Markov-switching model with VAR coefficients and volatilities evolving according to two independent chains (MS-VAR), I show that it is in fact possible to isolate a *'30s regime*. An interesting feature of the *'30s regime* is that the value spread predicts low stock market returns, whereas this is not true for a large part of the remainder of the sample.[2] This

[2]The value spread is the difference in the log book-to-market ratios of small value and small growth stocks. Therefore, when the performance of growth stocks is particularly good, this variable increases.

property is quite relevant because it is implied by the ICAPM itself. If growth stocks deliver lower returns, and these returns are not explained by market betas, it must be the case that these assets provide hedging against future low market returns. Therefore, we would expect the value spread to enter with a negative coefficient in a regression that has future market returns as the dependent variable. Consistent with this finding, I show that in order to maximize the explanatory power of the ICAPM a large weight must be assigned to the *'30s regime*, even if there is little evidence of it occurring again after the early years. The best version of the model delivers $R^2 = 53.77\%$ when the *initial* weight assigned to the *'30s regime* is well above its historical occurrence, implying that when forming expectations agents rely heavily on what happens during extraordinary events, even if they are aware that different regimes might prevail in the *long run*.

The third contribution of this paper is methodological and consists of extending the present value decomposition methodology of Campbell and Shiller (1988) and Campbell (1991) to allow for Markov-switching dynamics in the law of motion of the financial variables. The formulas presented in the paper are specific for the model of Campbell and Vuolteenaho (2004a), but they can be easily modified to handle other models that make use of a present value decomposition. This approach, which formally isolates periods characterized by unusual dynamics, might prove useful in explaining why the present value decomposition methodology is often sensitive to the sample choice. The paper also provides a simple algorithm to estimate a Markov-switching VAR in reduced form with Bayesian methods. However, practitioners that are more familiar with frequentist econometrics can make use of the results presented here and use a method of their choice to estimate the MS-VAR.

The content of this paper can be summarized as follows. Section 2 presents a short review of the related literature. Section 3 explains how to obtain the two types of news starting from a fixed coefficient VAR and introduces the ICAPM. Readers that are familiar with Campbell and Vuolteenaho (2004a) may skip this section. Section 4 highlights the importance of the early years of the sample: Once these are removed the ICAPM delivers poor results. Section 5 shows that it is possible to isolate a *'30s regime* using a Markov-switching model. In section 6, I describe how to compute the news when the dynamics of the state variables are described by a Markov-switching process. Section 7 shows that a large weight must be assigned to the *'30s regime* in order to maximize the explanatory power of the ICAPM. Section 8 shows that if agents simply updated their beliefs according to a time-varying VAR, the ICAPM would not be able to account for the value-small anomalies, implying that the Great Depression had a long lasting impact on expectations. Section 9 concludes.

2 Related literature

The results of this paper can be linked to some recent contributions in the finance and macro literature. The role of the Great Depression is central in Cogley and Sargent (2007). They 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. They show that with sufficient initial pessimism, their model is able to generate substantial values for the market price of risk and equity premium and to predict high Sharpe ratios and forecastable excess stock returns. On the other hand, it could be that agents have a *limited capacity* when it comes to acquiring information. In a model of Rational Inattention (Sims (2003, 2006)) agents would probably find it optimal to devote more attention to extraordinary events. Such events are likely to be easier to interpret and more important to understand. As for the importance of allowing for parameter instability when modeling agents' expectations, this paper is close to the work of Bianchi et al. (2009). They use a Time-Varying Factor-Augmented VAR (FAVAR) to model the interaction between the yield curve and the real economy. When agents are assumed to form expectations according to the Time-Varying FAVAR, they find that deviations from the expectations hypothesis are rare.

The idea that rare events affect the 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)). As observed by Julliard and Ghosh (2008) this explanation would negatively affect the ability of the Consumption-CAPM to price the cross-section of asset returns, since such an expectation would reduce the cross-sectional dispersion of consumption risk across assets. Instead, the argument that I propose is based on the idea that rare events modify the way agents interpret *whatever* happens in stock markets. In other words, agents can form expectations and interpret events based on a limited number of episodes, without necessarily expecting a market crash to occur in the near future.

Lewellen et al. (2008) and Daniel and Titman (2006) have highlighted several drawbacks of the empirical methods used to test factor-model explanations of market anomalies. Among other things, they recommend testing models on a large set of portfolios and taking into account coefficient restrictions as implied by economic theory. The results of this paper are essentially robust to these critiques. The ICAPM is based on two factors and imposes economically motivated restrictions on the premia. Furthermore, good results are obtained even when imposing the *zero-beta-restriction*, i.e. that a portfolio with both betas equal to zero should deliver the risk free rate. Moreover, in testing the model, 20 risk-sorted portfolios are added to the 25 Fama-French portfolios.

Finally, there are 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. They conduct several tests showing that the results of Campbell and Vuolteenaho (2004a) are very sensitive to the set of variables that are included in the VAR. To overcome this drawback Campbell et al. (2007) use direct proxies for cash-flow and discount-rate news. Bianchi (2003) highlights that the results are also very sensitive to the sample chosen to estimate the news series. Thomas and Zhang (2007) and Wei and Joutz (2009) point out that the pre-50s data are important for the results of Campbell and Vuolteenaho (2004b), another important contribution from the same authors that makes use of a present value decomposition to test the inflation illusion hypothesis. Bianchi (2003) represents the *pars destruens* of a more extended argument that finds its *pars costruens* in this paper. Here I offer an explanation for why the sample choice is so relevant and provide new tools to handle this issue. However, it is important to keep in mind that cash-flow and discount-rate news are obtained through a highly non-linear transformation of the residuals. This means that even extremely small changes in the VAR coefficients can have important effects on the final results.

3 The ICAPM

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.

However, 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.

This means that 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. In fact, according to an Intertemporal capital asset pricing model along the lines of the one proposed by Merton (1973), 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.

The first step consists in obtaining estimates for the news. Using the loglinear approximation for returns introduced by Campbell and Shiller (1988), unexpected excess returns can be approximated by:

$$r_{t+1} - E_t r_{t+1} = \underbrace{(E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j}}_{N_{CF,t+1}} - \underbrace{(E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j}}_{N_{DR,t+1}} \tag{1}$$

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, 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 future market's cash flows and news about future market's discount returns, respectively.

The VAR methodology, introduced by Campbell (1991), provides an estimate for the terms $E_t r_{t+1}$ and $N_{DR,t+1} = (E_{t+1} - E_t) \sum\limits_{j=0}^{\infty} \rho^j \Delta r_{t+1+j}$. Then $N_{CF,t+1}$ is derived from (1).

Consider a VAR in companion form:

$$Z_{t+1} = a + \Gamma Z_t + u_{t+1} \tag{2}$$

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:

$$r_{t+1} - E_t r_{t+1} = e_1' u_{t+1} \tag{3}$$

$$N_{CF,t+1} = (e_1' + e_1' \lambda) \, u_{t+1} \tag{4}$$

$$N_{DR,t+1} = e_1' \lambda u_{t+1} \tag{5}$$

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

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

$$\widehat{\beta}_{i,CF} = \frac{\widehat{cov}(r_{i,t}, N_{CF,t})}{\widehat{var}(N_{CF,t} - N_{DR,t})} + \frac{\widehat{cov}(r_{i,t}, N_{CF,t-1})}{\widehat{var}(N_{CF,t} - N_{DR,t})} \tag{6}$$

$$\widehat{\beta}_{i,DR} = \frac{\widehat{cov}(r_{i,t}, -N_{DR,t})}{\widehat{var}(N_{CF,t} - N_{DR,t})} + \frac{\widehat{cov}(r_{i,t}, -N_{DR,t-1})}{\widehat{var}(N_{CF,t} - N_{DR,t})} \tag{7}$$

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 from the first equation (see (1)). The market beta is obtained by summing the two betas. This is different than the usual market beta because of the additional lag of the news terms. However, the poor performance of the CAPM does not depend on this assumption.

The final step consists of determining whether the two betas represent an improvement over the single market beta. 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

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

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 $\widehat{\beta}_{i,M} = \widehat{\beta}_{i,CF} + \widehat{\beta}_{i,DR}$. According to the ICAPM the premia should be: $g_1 = \gamma\sigma_M^2$ and $g_2 = \sigma_M^2$, where γ is the coefficient of relative risk aversion and 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. In the factor model the coefficients are not restricted. Each model can be tested with and without the constant g_0. Excluding the constant is equivalent to imposing that a portfolio with both betas equal to zero should deliver the risk-free return (zero-beta-restriction).

4 Excluding the early years

Campbell and Vuolteenaho (2004a) implement the steps described in section 3 estimating a fixed coefficient VAR(1) over the sample 1928:12-2001:12 and computing the news for the entire sample using formulas (4) and (5). Then they split the sample into two parts 1928:12-1963:6 and 1963:7-2001:12, compute the betas for the 25 Fama and French portfolios and 20 risk-sorted portfolios, and test the three models (the static CAPM, the "Bad Beta, Good Beta ICAPM", and the unrestricted factor model) over the two subsamples.

The ICAPM returns very good results over both sub-samples, whereas the CAPM performs very poorly over the second subsample (1963:7-2001:12). This important finding is explained by a change in the composition of the single market beta across the different portfolios. Specifically, over the second subsample, growth stocks turn out to have large market betas, but with a large good beta component, whereas value stocks have higher bad betas than growth stocks do. This justifies their high returns over the second subsample. On the other hand, the composition of the market beta was substantially homogenous before the '60s. This explains why the static CAPM, based on the single market beta, is able to explain the cross section of asset returns for the pre-'63 subsample.

At this point, it is important to point out that while the betas are computed over two distinct subsamples, cash-flow and discount-rate news are based on the VAR coefficients estimated over the entire sample, 1928:12-2001:12. Two important assumptions are implied by this choice. First, the dynamics of the variables included in the model must have been stable over time. Second, agents are somehow aware of these underlying parameters and they use them to form expectations.

I focus on the results for the second subsample. In particular, I investigate the importance of including the first years of the sample when estimating the VAR. I begin assessing the performance of the different models by computing the two types of news according to a VAR estimated on the entire sample. I then shorten the sample by a single month at a time, while keeping the methodology described in section 3 unchanged. I use the same state variables used by Campbell and Vuolteenaho (2004a): The excess log return on the CRSP value-weighted index, the term yield spread in percentage points, measured as the yield difference between ten-year constant-maturity taxable bonds and short-term taxable notes, the log price earning ratio, and the small-stock value-spread, the difference in the log book-to-market ratios of small value and small growth stocks. The models are tested on the 25 Fama and French portfolios and 20 risk-sorted portfolios.[3] Thus, the sample size for the initial VAR is the only thing that is changing, whereas the subsample under investigation (1963:07-2001:12) and the variables used to test the model are not.

4.1 Evolution of the explanatory power of the models

Figure 2 reports the evolution of R^2 for the three models.[4] Notice that excluding the first three years from the sample used in estimating the VAR has a dramatic effect on the ability of the ICAPM to explain the returns of the 44 portfolios (the smallest-growth portfolio is excluded from the estimates). When the entire sample 1928:12-2001:12 is used, $R^2 = 51.96\%$

[3] All the data used in this paper come from the dataset that Campbell and Vuolteenaho make available online (http://www.aeaweb.org/articles/).

[4] This 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.

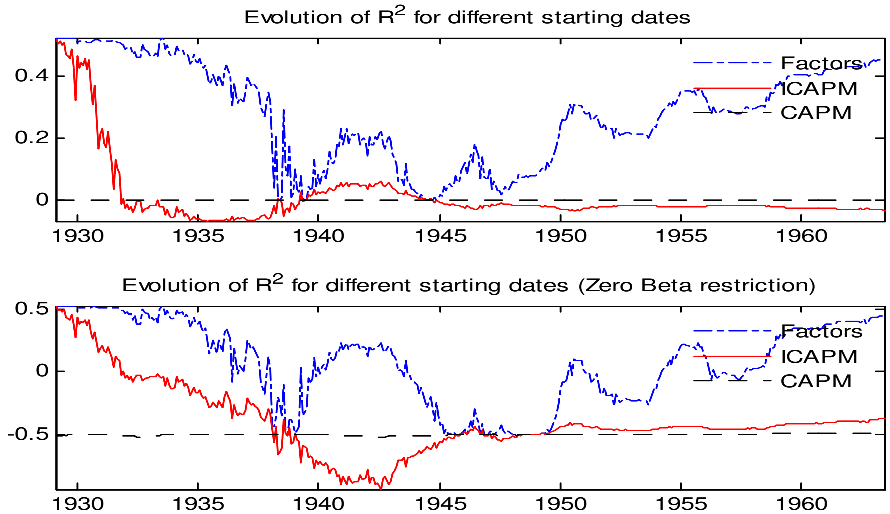

Figure 2: Evolution of R^2 for the three models as the sample used to compute the two types of news shortens. For example: 1940:12 means that cash-flow and discount-rate news have been computed according to a VAR estimated over the sample 1940:12-2001:12. While the sample used to estimate the VAR is changing, the models are always tested over the same subsample (1963:07-2001:12). The zero-beta-restriction implies that the return of a portfolio with both the betas equal to zero should equal the risk free rate.

for the unrestricted ICAPM and $R^2 = 51.57\%$ for the restricted ICAPM. However, once the first three years are excluded these remarkable results disappear.

A different argument applies to the factor model. Again the performance depends on the sample used to compute cash-flow and discount-rate news, but the factor model works well for different choices of the initial sample. In particular it has a good explanatory power when using only the second subsample to estimate the VAR and the two kinds of news.

The difference between the two models is given by the restrictions imposed by the ICAPM on the β_{DR} premium. When the entire sample is included, the factor model premia are 0.0526, for β_{CF}, and 0.015, for β_{DR}. The unrestricted premium of the discount rate beta is in this case very close to the value imposed by the ICAPM ($\widehat{\sigma}^2_M = 0.02$). However, this is not the case when the early years are removed. For example, when only the second subsample is used, the unrestricted premia turn out to be 0.0515, for β_{CF}, and -0.0436, for β_{DR}. The premium of the discount rate beta is now very different from the value implied by the ICAPM. These features reflect two crucial findings that will be described in the following two subsections: When the Great Depression is removed from the estimates: 1) The correlation between the two types of news moves from positive to negative, implying that the correlation between the betas moves from negative ($\simeq -0.38$) to strongly positive

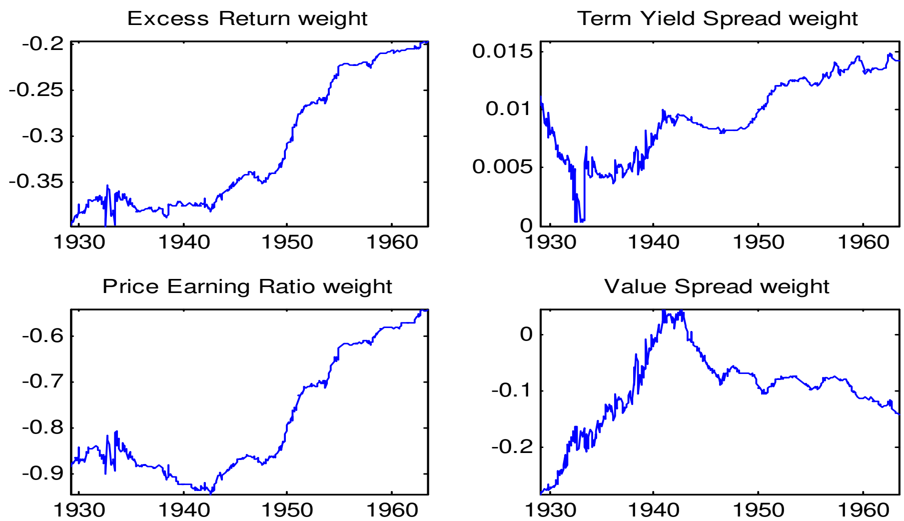

Figure 3: Evolution of the weights used to transform the residuals into discount-rate news as the sample size shortens. For example: 1940:12 means that the VAR has been estimated over the sample 1940:12-2001:12.

($\simeq 0.8$);[5] 2) The important feature that value stocks are characterized by larger cash-flow betas does not hold anymore.

Finally the graph shows that the CAPM presents an extremely poor performance independent of the sample used. This fact, together with the extreme sensitivity of the ICAPM to the sample choice, suggests that the coefficient estimates are important in decomposing the residuals into cash-flow news and discount rate news, but not for calculating unexpected excess returns. In other words, it is notoriously hard to predict stock market returns, and the only factor that really matters is how agents *interpret* what happens on the stock markets. The following two subsections will reinforce this intuitive argument.

4.2 Cash-flow and discount-rate news

As a first step, it is useful to consider how the vector used to transform the residuals into cash-flow and discount-rate news varies as the sample shortens. Figure 3 describes the evolution of the weights used to construct the discount-rate news. The horizontal axis reports the starting date of the sample over which the VAR has been estimated, while the vertical axis shows the weight assigned to the residual from the specified equation, when computing $N_{DR,t+1}$. As the initial sample shortens, the weights vary significantly. On the other hand, the residuals of

[5] The correlation between the betas becomes very close to 1 after WWII. Graphs of the premia and of the correlation between the betas are reported in appendix D.

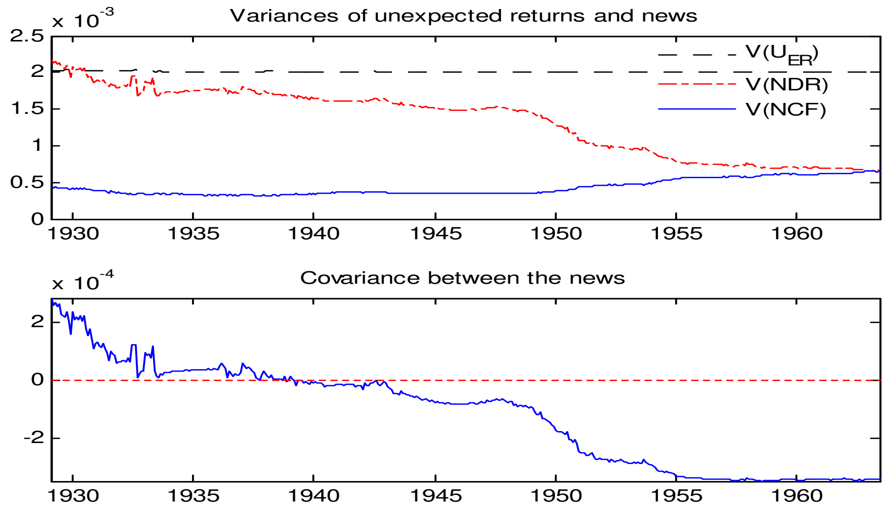

Figure 4: Evolution of the unexpected return variance and of the covariance matrix of the cash-flow and discount rate news over the subsample 1963:07-2001:12 as the sample size used to estimate the VAR shortens. For example: 1940:12 means that the VAR has been estimated over the sample 1940:12-2001:12. Unexpected returns are the residuals of the excess return equation and they are equal to the difference between cash-flow news (NCF) and discount rate news (NDR). Note that the sample used to estimate the VAR is changing, while the covariance matrix and the unexpected return variance refer to the same subsample.

the excess return equation are quite stable given that it is generally difficult to predict stock returns. This implies that the results that follow are driven by the changes in the weights.

Figure 4 displays how the variance-covariance matrix of the two types of news and the variance of unexpected returns vary as the initial sample shortens. The variance of unexpected returns is, for the most part, unaffected by changes in the sample used to compute cash-flow and discount rate news, while the variance of the discount rate news and the covariance between the two types of news move substantially. In particular, whether or not the first years of the sample are included has a relevant impact on the covariance between cash-flow and discount-rate news. If the early years are excluded, the covariance turns out to be very close to zero, while if they are included it is positive. Furthermore, when even more observations are excluded, the covariance becomes negative. Recall that unexpected returns are equal to the difference between cash-flow news and discount rate news (equation (1)). Therefore:

$$V\left(u_{ER}\right) = V\left(NCF_t - NDR_t\right) = V\left(NCF_t\right) + V\left(NDR_t\right) - 2COV\left(NDR_t, NCF_t\right)$$

Hence, the composition of the sample used to model the expectation mechanism has a large

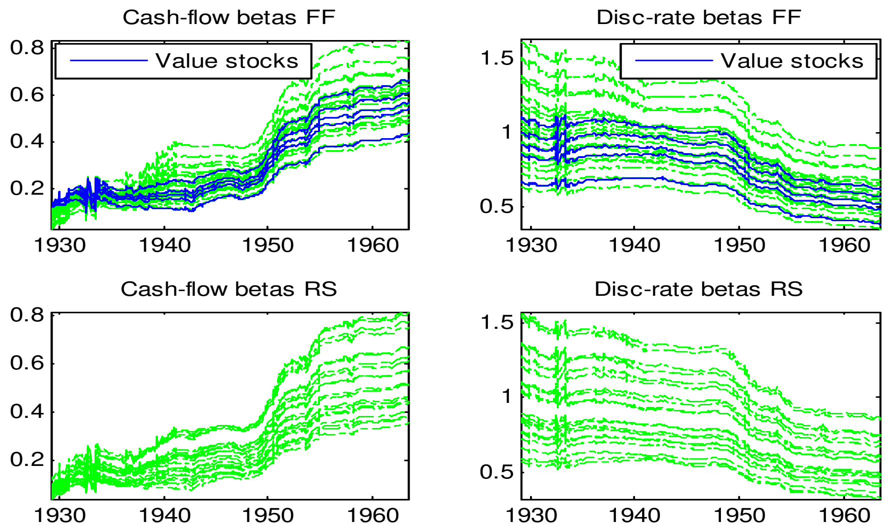

Figure 5: Time evolution of the betas over the subsample 1963:07-2001:12 for different starting dates of the sample used to compute cash-flow and discount rate news. For example: 1940:12 means that the VAR has been estimated over the sample 1940:12-2001:12. The top panels refer to the Fama-French portfolios, the lower panels present the results for the risk-sorted portfolios. Note that the sample used to estimate the VAR is changing, while the betas are always computed over the same subsample.

impact on how return innovations are decomposed into cash-flow and discount rate news, but a minimal effect on the model's capacity to predict stock market returns. This explains why $V(u_{ER})$ shows only imperceptible changes in response to the sample choice.

The CAPM is based on the covariance between portfolio returns and unexpected returns. Therefore, the sample choice should have a minimal effect on its performance. However, the ICAPM relies on the covariance between the two types of news and portfolio returns. Hence, the way that unexpected returns are decomposed into the two kinds of news is extremely important for this model. We would expect the sample choice to have a large impact on the explanatory power of the ICAPM, but minimal effect on the performance of the CAPM. This is exactly what we found in section 4.1.

4.3 Cash-flow and discount-rate betas

Figure 5 reports cash-flow and discount rate betas for the 25 Fama-French ME and BE/ME-sorted portfolios and the 20 portfolios sorted on past risk loadings with VAR state variables. The top panels refer to the Fama and French portfolios, while the lower panels contain the betas for the risk-sorted portfolios. The betas move in similar ways across the portfolios:

cash-flow betas exhibit upward trends, while all the discount rate betas follow a common negative trend. These movements mimic the patterns of the variances of cash-flow and discount-rate news. Finally, the central finding of Campbell and Vuolteenaho, that value stocks have a larger cash-flow beta, disappears once the Great Depression has been excluded from the estimates (see solid lines in the left-top panel).

Consider the formulas used to compute the betas ((6) and (7)). In the current exercise we keep the sample used to compute the betas fixed, hence we can rule out the possibility that portfolio returns are the source of this common trend. Moreover, the denominators are very stable because the residuals of the excess return equation are not very sensitive to the sample choice. We can conclude that these trends are driven by the changes in the estimates of the two news series, that in turn depend on the VAR coefficient estimates.

5 Was the Great Depression a rare event?

Was the Great Depression a rare event? The answer to this question seems so obvious, that it is hardly reasonable to spend an entire section to answer it: a simple "yes" would suffice. However, the goal here is not to state the obvious but to understand if a *'30s regime* can be formally detected and if this can help to clarify why the early years are so important.

Instead of being certain about the current state of the world, investors could have in mind a limited number of alternative scenarios. In this case, they would first obtain estimates of the parameters and then they would form expectations according to the probabilities assigned to different regimes. States that are not likely to occur in the near future could still receive a high weight when investors form expectations, especially if they were associated with extraordinary events. A Markov-switching model is the perfect tool to formalize this idea.

5.1 The model

As before, the state variables follow a VAR(1). However, in this model, both the VAR coefficients and the covariance matrix are allowed to switch across regimes:

$$Z_t = a(s_t^\Phi) + \Gamma(s_t^\Phi) Z_{t-1} + \Sigma(s_t^\Sigma)^{1/2}\omega_t \quad (8)$$

$$\Phi(s_t^\Phi) = \left[a(s_t^\Phi), \Gamma(s_t^\Phi)\right] \quad (9)$$

where $Z_t = [ER_t, TY_t, PE_t, VS_t]'$ denotes the data matrix, s_t^Σ and s_t^Φ are unobserved states and $\omega_t \sim N(0, I)$. Z_t contains the same state variables: Excess return (ER_t), term yield spread (TY_t), the (log) price earning ratio (PE_t), and the value spread (VS_t).

The model assumes that the VAR coefficients and the covariance matrix of the residuals follow two independent chains. This represents a convenient way to model heteroskedasticity and to allow for the possibility of changes in the dynamics of the state variables. A simplified model in which all the parameters follow a common chain can be obtained assuming $s_t^\Sigma = s_t^\Phi$.[6]

The unobserved states 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. Therefore, the probability of moving from one state to another is given by $P[s_t^\Phi = i | s_{t-1}^\Phi = j] = h_{ij}^\Phi$ and $P[s_t^\Sigma = i | s_{t-1}^\Sigma = j] = h_{ij}^\Sigma$.

Given $H^\Phi = [h_{ij}^\Phi]$ and $H^\Sigma = [h_{ij}^\Sigma]$ and a prior distribution for the initial state, we can obtain maximum likelihood estimates of the parameters of the model, conditional on the initial observation Z_1. In the process we obtain filtered estimates of the state, giving $P[s_t^\Phi = i | \{Z_s, s \leq t\}, \Phi(\cdot), \Sigma(\cdot), H^\Phi, H^\Sigma, s_t^\Sigma]$ and $P[s_t^\Sigma = i | \{Z_s, s \leq t\}, \Phi(\cdot), \Sigma(\cdot), H^\Phi, H^\Sigma, s_t^\Phi]$ for all i at each t. The filtered estimates of state probabilities can then be converted by a recursive algorithm to smoothed estimates. A similar recursive algorithm generates pseudo-random draws from the posterior distribution of the sequence of states $s^{\Phi,T}$ and $s^{\Sigma,T}$ conditional on $Z^T, \Phi(\cdot), \Sigma(\cdot), H^\Phi, H^\Sigma$ (where $x^T = \{x_t\}_{t=1}^T$).

5.2 Algorithm

The model is estimated with Bayesian methods and proper priors are put on all the parameters in the model. Given that the final goal is to establish whether it is possible to isolate a *'30s regime*, strong priors are imposed on the transition matrix for the VAR coefficients so as to obtain persistent regimes and avoid meaningless switches from one regime to another. In order to compute the news, we need to make sure that the model is stable. In principle, we only need convergence of the first moments, however I adopt a stronger concept of stability: Mean-square stability (MSS). A system is mean square stable if both the first and second moments converge. In the Gibbs sampling algorithm, draws of the model parameters that imply instability are rejected so that the priors are in fact truncated. Please refer to section 6 and appendix C for more details. To capture the idea that one regime might occur less frequently than the other, the persistence implied by the priors differs across the two VAR coefficients regimes, whereas the priors for all the other parameters are standard and

[6] Note that the regime switch is modeled for a VAR in its reduced form. Sims and Zha (2006) recommend working directly on the structural form of a 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.

symmetric across the two regimes.[7]

I report results for the posterior mode, based on the maximization of the log-posterior, and I employ a Gibbs sampling algorithm to draw from the posterior distribution. A detailed description of the prior distributions and the sampling method is given in appendix A. Here I summarize the Gibbs sampling algorithm which involves the following steps:

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

2. Sampling $\Phi(s_t^{\Phi})$ and $\Sigma(s_t^{\Sigma})$

   - If $s_t^{\Phi} = s_t^{\Sigma}$, i.e. if the VAR coefficients and the covariance matrix follow a common regime, standard results can be used: The VAR coefficients are sampled from a normal distribution and the covariance matrices are drawn from an inverted Wishart distribution (Uhlig (2005)).
   - If s_t^{Φ} and s_t^{Σ} are independent we need to proceed in two steps. Given $\Phi(\cdot)$ and s_t^{Φ} we can compute the residuals. Then, given s_t^{Σ}, $\Sigma(\cdot)$ can be drawn from an inverse Wishart distribution. When drawing the VAR coefficients we need to take into account the heteroskedasticity implied by the switches in $\Sigma(\cdot)$. This can be done using GLS or a Kalman filter. Appendix B describes the two methods.

3. Sampling H^{Φ} and H^{Σ}: Given the draws for the state variables s_t^{Φ} and s_t^{Σ}, the transition probabilities are independent of Y_t and the other parameters of the model. Therefore, they can be drawn using a Dirichlet distribution.

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

I use 500,000 Gibbs sampling iterations, discard the first 50,000 as burn-in and retain one every 50 of the remaining draws. The posterior moments vary little over the retained draws providing evidence of convergence.

[7] This implies that the features of the two regimes are not restricted, so differences will arise only because of the data. Results under a symmetric prior for H^{Φ} are available upon request and they are substantially equivalent to the ones presented here, with the only difference that it becomes harder to obtain MSS, even if convergence of first moments still holds.

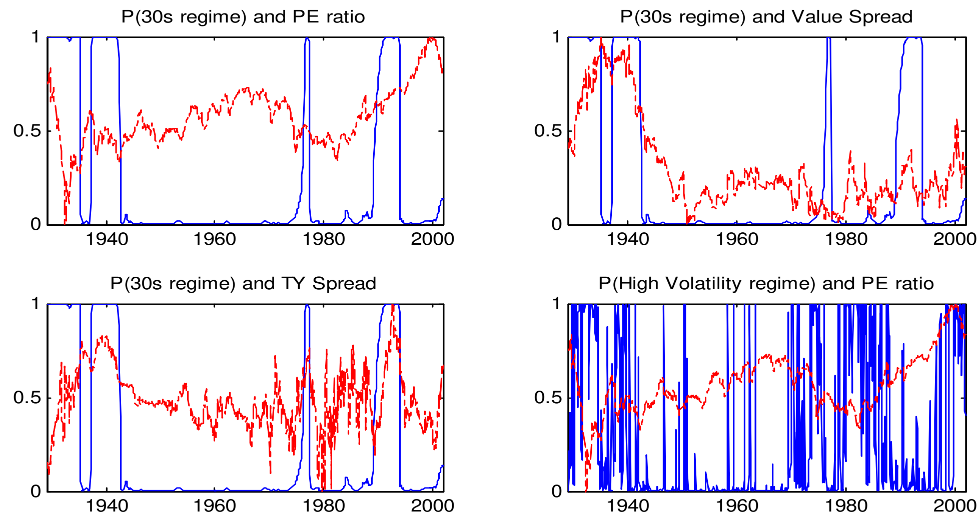

Figure 6: Markov-switching VAR with independent regimes. Starting from the lower-left panel and proceeding clockwise, the first three panels show the posterior mode probability of regime 1 for the VAR coefficients (the '30s regime) together with the term yield spread, the price-earning ratio, and the value spread. The last panel contains the probability of the high volatility regime with the price-earning ratio. The three state variables are normalized to fit into the graphs.

5.3 VAR Estimates

This section reports parameter estimates for the model described in section 5.1 for the case in which the number of regimes is equal to two for both chains, $m^{\Phi} = 2 = m^{\Sigma}$, for a total of four regimes. Figure 6 shows the smoothed probabilities of $s_t^{\Phi} = 1$ and $s_t^{\Sigma} = 1$ at the posterior mode. Table 1 reports posterior mode and 68% error bands for the parameters of the Markov-switching VAR.

Starting from the lower-left corner and proceeding clockwise, the first three panels of figure 6 report the smoothed posterior mode probability of regime 1 for the VAR coefficients together with the term yield spread, the price-earning ratio, and the value spread (the variables are normalized to fit in the graph). This regime clearly dominates the first decade, a period characterized by large market crashes and unusually high values for the term yield spread and the value spread. Then it occurs again in the mid-'70s and in the early 90s, when we observe very persistent increases in the term yield spread, and with some probability at the end of the sample, with the end of the IT bubble. The behavior of the value spread and the price earning ratio in the early '30s is worth noting. The largest stock market crash of US history came with a substantial increase in the value spread, which reached historic

$s_t^\Phi = 1$	ER_t	TY_t	PE_t	VS_t	$const$
ER_{t+1}	0.0295 (−0.0826,0.0713)	0.0008 (−0.0042,0.0036)	−0.0061 (−0.0122,0.0047)	−0.0112 (−0.0151,0.0005)	0.0375 (−0.0022,0.0575)
TY_{t+1}	−0.0424 (−0.2564,0.2419)	0.9827 (0.9555,0.9910)	−0.0008 (−0.0391,0.0374)	−0.0253 (−0.0147,0.0548)	0.0979 (−0.1509,0.1576)
PE_{t+1}	0.5165 (0.4259,0.5352)	−0.0021 (−0.0043,0.0008)	1.0009 (0.9958,1.0060)	−0.0047 (−0.0064,0.0051)	0.0058 (−0.0211,0.0185)
VS_{t+1}	−0.0254 (−0.0685,0.0256)	−0.0033 (−0.0060,0.0003)	−0.0044 (−0.0126,0.0007)	1.0063 (0.9968,1.0089)	0.0073 (−0.0102,0.0410)

$s_t^\Phi = 2$	ER_t	TY_t	PE_t	VS_t	$const$
ER_{t+1}	−0.0022 (−0.0401,0.0740)	0.0029 (−0.0000,0.0088)	−0.0142 (−0.0286,−0.0110)	0.0176 (0.0075,0.0234)	0.0236 (0.0135,0.0704)
TY_{t+1}	0.4538 (0.1958,0.7259)	0.8037 (0.7655,0.8184)	−0.0969 (−0.1440,−0.0652)	0.1346 (0.0793,0.1700)	0.1662 (0.0583,0.3595)
PE_{t+1}	0.4042 (0.3753,0.4565)	0.0010 (−0.0012,0.0048)	0.9947 (0.9861,0.9974)	0.0159 (0.0068,0.0188)	−0.0082 (−0.0145,0.0239)
VS_{t+1}	0.0064 (−0.0370,0.0512)	0.0018 (−0.0012,0.0060)	−0.0006 (−0.0067,0.0074)	0.9756 (0.9622,0.9809)	0.0351 (0.0144,0.0626)

$s_t^\Sigma = 1$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0074 (0.0074,0.0091)	0.0323 (−0.0298,0.1048)	0.7770 (0.7501,0.8039)	−0.0706 (−0.1465,−0.0120)
u_{TY}	0.0012 (−0.0012,0.0043)	0.1865 (0.1809,0.2221)	0.0341 (−0.0292,0.1049)	0.0056 (−0.0646,0.0674)
u_{PE}	0.0037 (0.0036,0.0046)	0.0008 (−0.0008,0.0027)	0.0341 (0.0030.0.0037)	−0.0926 (−0.1664,−0.0339)
u_{VS}	−0.0005 (−0.0011,−0.0001)	0.0002 (−0.0023,0.0024)	−0.0004 (−0.0008,−0.0002)	0.0058 (0.0057,0.0070)

$s_t^\Sigma = 2$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0012 (0.0012,0.0013)	−0.0087 (0.0619,0.0289)	0.7237 (0.0701,0.7437)	0.0654 (0.0226,0.1159)
u_{TY}	−0.0000 (−0.0003,0.0002)	0.0209 (0.0226,0.0266)	−0.0156 (−0.0625,0.0280)	−0.0149 (−0.0485,0.0422)
u_{PE}	0.0006 (0.0006,0.0007)	−0.0001 (−0.0003,0.0001)	0.0006 (0.0006,0.0007)	−0.0212 (−0.0617,0.0344)
u_{VS}	0.0001 (0.0000,0.0001)	−0.0001 (−0.0002,0.0002)	−0.0000 (−0.0000,0.0000)	0.0007 (0.0007,0.0009)

H^Φ	h_{11}^Φ	h_{22}^Φ
	0.9718 (0.9505,0.9767)	0.9940 (0.9730,0.9917)

H^Σ	h_{11}^Σ	h_{22}^Σ
	0.7706 (0.7191,0.8076)	0.9087 (0.9000,0.9309)

Table 1: The three sets of tables contain modes and 68% error bands for the parameters of the Markov-switching VAR. The first two tables report the estimates for the VAR coefficients. The second set of tables contains the elements of the covariance matrices (shown on and below the main diagonal) and the implied correlations (shown above the main diagonal). The last table contains the diagonal elements of the transition matrices.

heights. In other words, during the most severe recession that the US has ever experienced (so far), growth stocks were outperforming value stocks and this situation of disequilibrium lasted for a decade. A rational agent who is trying to hedge against risk is likely to find this pattern extremely informative. From now on I will refer to regime 1 as the *'30s regime*.

Overall, only some of the coefficients are tightly estimated. Nevertheless, the regimes are well identified and the VAR coefficients show some remarkable differences across the two regimes.[8] First of all, the coefficient of the value spread in the excess return equation is negative for the *'30s regime*, while it is positive for regime 2. This means that high returns on small growth stocks predict low stock market returns only under the *'30s regime*. The autoregressive component for the term yield spread is substantially closer to one when the *'30s regime* prevails. This explains the high probability of the *'30s regime* in the mid-70s and during the first half of the '90s. Furthermore, the excess return, the price earning ratio, and the value spread enter the term yield equation with coefficients that are substantially different across the two regimes. The coefficient of the excess return in the price earning equation is larger for the *'30s regime*, even if the bands partially overlap, and differences can be detected for the coefficients of the price earning ratio and term yield spread in the excess return equation. Finally, the autoregressive coefficients for these two variables are larger under the *'30s regime*.

As for the covariance matrix, regime 1 is associated with more uncertainty. Interestingly, value spread and excess return innovations are negatively correlated under the high volatility regime, whereas this is not the case under the low volatility regime. The panel in the low right corner of figure 6 reports the probability of the high volatility regime and the log price earning ratio. The probability of the high volatility regime is very high during the early years of the sample, but it increases again at different points in time. Note that sudden negative changes in the price-earning ratio are often associated with an increase in the probability of the high volatility regime.

6 News in a Markov-switching framework

Note that while regime 2 is stable, the *'30s regime* is not. Stationarity is necessary for the Campbell-Shiller approximation to hold. Therefore, it would be problematic to apply the VAR decomposition to the *'30s regime* alone. However, the system as a whole is stable, which is what we need to be able to decompose the residuals into news. In fact, the instability of the *'30s regime* represents an advantage of the approach used in this paper. The Markov-

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

switching model allows for temporary deviations from the stationarity assumption, provided that the model assigns a sufficiently high probability to the stable regime.

In what follows, I generalize the present value decomposition of Campbell and Shiller (1988) and Campbell (1991) for the case in which the law of motion of the state variables is described by a Markov-switching VAR. Here the focus is on discount rate and cash flow news, but the methods can be applied whenever we are looking for a present value decomposition in a Markov-switching framework. Define the column vectors q_t and ψ_t as:

$$q_t = \left[q_t^{1\prime}, ..., q_t^{m\prime}\right]', q_t^i = E\left(Z_t 1_{s_t=i}\right), \psi_t = \left[\psi_t^{1\prime}, ..., \psi_t^{m\prime}\right]',$$

where ψ_t^i is a column vector with all the elements equal to $p\left(s_t^\Phi = i\right)$ and $1_{s_t=i}$ is an indicator variable that is equal to one when regime i is in place and zero otherwise.

Note that:

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

Then, it can be shown that q_t evolves according to the following law of motion:

$$q_t = M\psi_t + \Omega q_{t-1} \tag{10}$$

$$\psi_t = G\psi_{t-1} \tag{11}$$

where $G = kron(H^\Phi, I_n)$ and, for the model considered in this paper, Ω and M are

$$\Omega = \begin{bmatrix} \Gamma\left(s_t = 1\right) h_{11}^\Phi & \Gamma\left(s_t = 1\right) h_{12}^\Phi \\ \Gamma\left(s_t = 2\right) h_{21}^\Phi & \Gamma\left(s_t = 2\right) h_{22}^\Phi \end{bmatrix}, \ M = \begin{bmatrix} diag\left(a\left(s_t = 1\right)\right) & 0 \\ 0 & diag\left(a\left(s_t = 2\right)\right) \end{bmatrix}$$

In order to compute the news, we need to make sure that the process described by (10) converges. As mentioned before, I assume mean-square stability, a concept of stability popular in the engineering literature. In fact, this is more than what is needed, given that it implies convergence of the second moments, whereas to compute the news we only need convergence of first moments. However, MSS is arguably a desirable feature because it implies that when trying to project the state vector into the future, the variance converges to a particular value that can be computed analytically. Costa et al. (2004) shows that MSS depends on the eigenvalues of the matrix governing the law of motion of the second moments of the homogeneous system. Appendix C describes these conditions and reports the general formulas for Ω and M.

Define, $q_{t,t+j}^i = E_t\left(Z_{t+j} 1_{s_{t+j}=i}\right)$ and note that $q_{t,t}^i = E_t\left(Z_t 1_{s_t=i}\right) = Z_t * p\left(s_t^\Phi = i\right)$ and $Z_t = \sum_{i=1}^m q_{t,t}^i$. Appendix C shows that when MSS holds, we can obtain formulas for discount rate news and cash-flow news that resemble formulas (3)-(5):

$$\begin{aligned} N_{DR,t+1} &= e_1' w \left[\lambda^q v_{t+1}^q + \lambda^\psi v_{t+1}^\psi \right] \\ N_{CF,t+1} &= e_1' w \left[(I_r + \lambda^q) v_{t+1}^q + \lambda^\psi v_{t+1}^\psi \right] \\ u_{t+1} &= e_1' w v_{t+1}^q \end{aligned}$$

where $r = n * m^\Phi$ and

$$\begin{aligned} \lambda^q &= (I_r - \rho\Omega)^{-1} \rho\Omega \\ \lambda^\psi &= (I_r - \rho\Omega)^{-1} (I_r - \rho G)^{-1} \rho M G \\ v_{t+1}^q &= q_{t+1,t+1} - q_{t+1,t} \\ v_{t+1}^\psi &= \psi_{t+1,t+1} - \psi_{t+1,t} \end{aligned}$$

Note that given a sequence of probabilities (or a posterior draw $s^{\Phi,T}$) and a set of parameters, it is easy and computationally efficient to compute the entire sequences $v^{q,T}$, $v^{\psi,T}$, and u^T:

$$N_{DR}^T = e_1' w \left[\lambda^q v^{q,T} + \lambda^\psi v^{\psi,T} \right] \tag{12}$$

$$N_{CF}^T = e_1' w \left[(I_r + \lambda^q) v^{q,T} + \lambda^\psi v^{\psi,T} \right] \tag{13}$$

$$u^T = e_1' w v^{q,T} \tag{14}$$

When the two regimes coincide, formulas (12)-(14) collapse to (3)-(5). So the above formulas can be treated as a generalization of the ones used in Campbell and Vuolteenaho (2004a).

7 The importance of the Great Depression

Based on the results thus far, it seems plausible that the Great Depression represented an exceptional event not only for the real economy, but also for the statistical properties of the financial variables. It is, therefore, interesting to ask what role this rare event has in explaining the cross section anomalies. It could be that agents consider the *'30s regime* as a memory of the past, given that there is little evidence of it occurring again after the first decade of the sample. On the other hand, agents could put a certain probability on its future occurrence.

Suppose agents have in mind that the state variables can evolve according to two distinct

Model	$w_{\Gamma(1),t}$	$z.b.r.$	R^2	MPE
Naive	0.6420	no	52.91%	0.7596
Naive	0.5640	yes	49.09%	0.8211
Semi-sophisticated	0.9978	no	53.77%	0.7387
Semi-sophisticated	0.9723	yes	53.34%	0.7456
Perf. knowl.	$p_{\Gamma(1),t}$	no	17.29%	1.3215
Perf. knowl.	$p_{\Gamma(1),t}$	yes	15.04%	1.3575

Table 2: R2 and Mean Pricing Error for the ICAPM. Cash-flow and discount-rate news are computed based on a MS-VAR with the VAR coefficients and the covariance matrix evolving according to independent chains. Naive model: Agents assign time invariant weights to the two regimes. Semi-sophisticated: Agents assign initial arbitrary time-invariant probabilities to the two regimes, but they then update them using the transition matrix. Perfect knowledge: Weights assigned to the two regimes equal their probabilities and are updated using the transition matrix.

regimes. If they were fully aware of the model they would 1) come up with the probabilities to assign to each of the states of the world, 2) use the transition matrix H^{Φ} to update these probabilities, 3) form expectations about the future according to the updated probabilities, 4) revise their beliefs after having observed the state variables, and, finally, 5) compute cash-flow news and discount-rate news according to the revised beliefs. I refer to this specification as the *perfect knowledge* model. Under these assumptions the expectation error is the sum of two components: one depends on the revision in beliefs about the state of the world once new information is received, the other reflects the typical VAR shocks. On the opposite side of the spectrum is the *naive* model. In this case agents simply use a weighted average of the two sets of coefficients. They use these weights not only to form expectations, but also to decompose the expectation error. Finally, I consider an intermediate case: the *semi-sophisticated* model. In this setting agents are aware of the transition matrix H^{Φ} and they use it when computing cash-flow and discount-rate news. However, when forming expectations they come up with their own initial weights instead of using the estimated probabilities.[9] The results for the three different specifications are reported in table 2.

I shall begin by considering the *naive* model: Investors are aware of the two regimes, but instead of using the formulas described in section 6, they simply take a weighted average of the two sets of VAR coefficients. Note that this assumption is computationally convenient because the formulas of section 3 can be applied once Γ has been defined as $\Gamma = w_{\Gamma(1)}\Gamma(1) + \left(1 - w_{\Gamma(1)}\right)\Gamma(2)$, where $\Gamma(1)$ and $\Gamma(2)$ are the posterior mode estimates of the VAR coefficients. The posterior modes are preferred to the means because the difference

[9] In principle the initial weights could be time-varying. However, without further restrictions, the weights would adjust to a perfect fit. This is why I don't report results for this case.

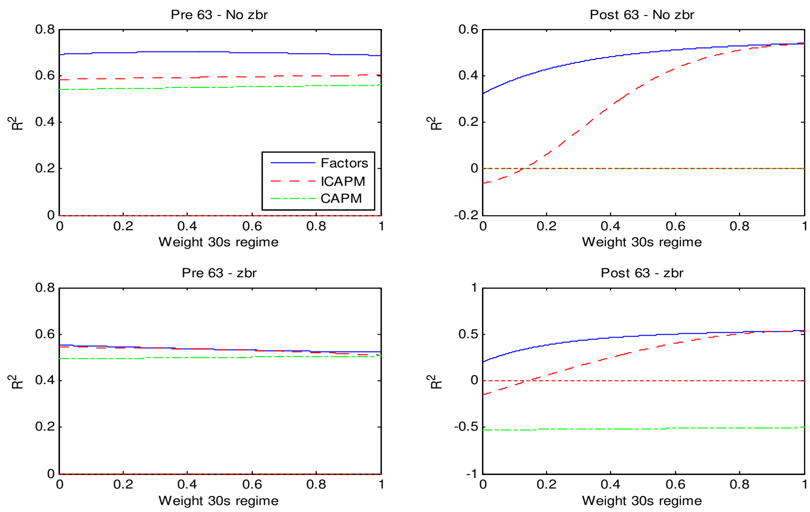

Figure 7: The figure shows how the explanatory powers of the unrestricted two-factor model, the ICAPM, and the CAPM vary in response to the *initial* weight assigned to the '30s regime under the semi-sophisticated specification. The left column refers to the pre-63 subsample, whereas the right column contains the results for the post-63 subsample. The first and the second row show results with and without the zero-beta-restriction, respectively.

in the posterior is particularly large (as explained by Sims et al. (2008) this happens quite frequently with Markov-switching models).

What are the weights that maximize the explanatory power (R_I^2) of the ICAPM? When imposing the zero-beta-restriction $R_I^2 = 49.09\%$ with $w_{\Gamma(1)} = 0.5640$, whereas $R_I^2 = 52.91\%$ with $w_{\Gamma(1)} = 0.6460$ when relaxing it. Note that the explanatory power of the model is large and it is obtained by assigning a weight to the *'30s regime* that is larger than its historical probability.

The results are even more suggestive under the *semi-sophisticated* specification. In this case, at each point in time agents assign arbitrary *initial* probabilities to the two regimes, but they are aware that in the *long run* both regimes will prevail. Therefore, when forming expectation they update the probabilities according to the transition matrix H^{Φ}. Figure 7 shows how the explanatory powers of the unrestricted two-factor model, the ICAPM, and the CAPM vary depending on the *initial* weight assigned to the *'30s regime*. The left column refers to the first subsample, whereas the right column contains the results for the second subsample. The first and the second row show results with and without the zero-beta-restriction, respectively. Note that the initial weight assigned to the *'30s regime* is substantially irrelevant for the first subsample: In this case the standard CAPM works well, so it is not important how the single market beta is decomposed into the two betas.

On the other hand, the weight of the *'30s regime* turns out to be very important when trying to explain the cross section of asset returns for the second subsample. The fit of the ICAPM model is maximized with $w_{\Gamma(1)} = 0.9978$, without the zero-beta-restriction, and $w_{\Gamma(1)} = 0.9723$, with the zero-beta-restriction. The correspondent R_I^2 is remarkably high, very similar across the two specifications (53.77% and 53.34%), and higher than the one obtained with the fixed coefficient VAR (51.96% and 51.57%). Furthermore, it is also very similar to the one obtained with the unrestricted two factor model that does not impose any economic restriction on the premia. This suggests that even if it is always possible to obtain a high R_I^2 by breaking the single market beta into two components, a sensible interpretation of the results might depend crucially on the way expectations have been modeled.

Finally, I consider the *perfect knowledge* model. Agents are fully aware of the MS model that drives the financial variables. They form expectations taking into account that $p_{t+1} = H^{\Phi}p_t$. Once the expectation error is revealed, they update their beliefs and they compute the two news series according to the MS process that drives the model. In this case the results for the second subsample are unsatisfactory ($R_I^2 = 17.29\%$ and $R_I^2 = 15.04\%$). Given the low weight that is assigned to the *'30s regime* over the second subsample (see figure 6), the poor performance associated with this specification is not surprising and is in line with the results of section 4.

These results support the original hypothesis: the weights given to rare events are not necessarily linked to the actual probability of their occurrence. The result for the *semi-sophisticated* specification is particularly interesting. Agents are aware of the two regimes and of the fact that in the long run both of them will occur with a certain probability. However, when forming expectations for the *immediate* future, they put a very large weight on the *'30s regime*. Obviously, this does not mean that they expect the Great Depression to occur with that exact probability. Rather, it suggests that the dynamic properties that prevailed during those years have a large impact on how agents think about financial markets.

8 A Time-Varying approach

A central result of this paper is that the statistical dynamics that characterize financial crises are important in explaining the value and stock anomalies. An alternative model is one in which agents are aware of parameter instabilities and update their beliefs disregarding what happened in extreme circumstances, as long as these events are far enough in the past. In this section I make use of a Bayesian Time-Varying VAR to model this kind of expectation formation mechanism.

8.1 The model

As before, the state variables follow a VAR(1), however, in this model, the VAR coefficients and the covariance matrix of the residuals are time-varying:

$$Z_t = a_t + \Gamma_t Z_{t-1} + v_t \tag{15}$$

where $Z_t = [ER_t, TY_t, PE_t, VS_t]'$ denotes the data matrix and $v_t = \Sigma_t^{1/2}\omega_t$ with $\omega_t \sim N(0, I)$. Once again, the state variable vector Z_t contains the excess return (ER_t), the term yield spread (TY_t), the (log) price earning ratio (PE_t), and the value spread (VS_t).

The VAR coefficients evolve according to a random walk: $\Phi_t = \Phi_{t-1} + \eta_t$, where $\Phi_t = vec\,[a_t, \Gamma_t]$. The covariance matrix of the VAR innovations v_t is factored as $VAR\,(v_t) \equiv \Sigma_t = A_t^{-1}H_t(A_t^{-1})'$. The time-varying matrices H_t and A_t are defined as:

$$H_t \equiv \begin{bmatrix} h_{1,t} & 0 & 0 & 0 \\ 0 & h_{2,t} & 0 & 0 \\ 0 & 0 & h_{3,t} & 0 \\ 0 & 0 & 0 & h_{4,t} \end{bmatrix} \text{ and } A_t \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha_{21,t} & 1 & 0 & 0 \\ \alpha_{31,t} & \alpha_{32,t} & 1 & 0 \\ \alpha_{41,t} & \alpha_{42,t} & \alpha_{43,t} & 1 \end{bmatrix} \tag{16}$$

with the $h_{i,t}$ elements evolving as geometric random walks, $\ln h_{i,t} = \ln h_{i,t-1} + u_t$, and the non-restricted elements of the matrix A_t evolve as driftless random walks, $\alpha_t = \alpha_{t-1} + \varepsilon_t$. The vector $[v_t', \eta_t', \varepsilon_t', u_t']'$ is distributed as $[v_t', \eta_t', \varepsilon_t', u_t']' \sim N(0, V)$, $V = blockdiag\,[\Sigma_t, Q, S, G]$ and $G = diag\,[\sigma_1^2, ..., \sigma_4^2]$.

The model is estimated using the Bayesian methods described by Kim and Nelson (1999a). In particular, I employ a Gibbs sampling algorithm that approximates the posterior distribution (see appendix A for details). The priors and the starting values for the VAR coefficients are based on a fixed coefficient VAR estimated over the sample 1928:12-1932:12.

8.2 Results

Figure 8 shows the evolution of the VAR coefficients and of the weights used to compute the two types of news. It is evident that the VAR coefficients change substantially over the early '30s, while they are relatively stable throughout the remainder of the sample. Note that at the beginning of the sample excess returns were positively autocorrelated and the value spread was predicting low stock market returns; these features disappear over the reminder of the sample. These results reinforce the idea that the early years reflect an exceptional event with specific statistical properties.

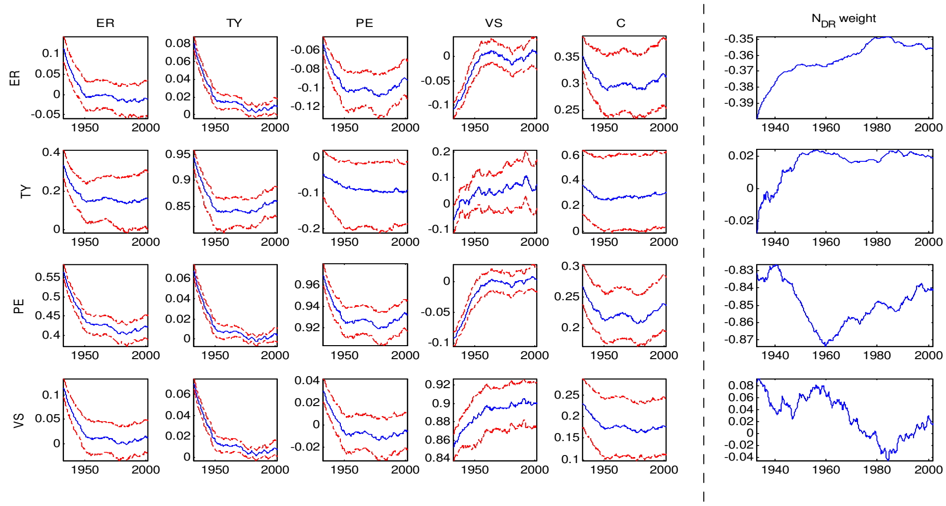

Figure 8: Time Varying VAR estimates: Evolution of the VAR coefficients and of the weights used to transform the residuals in discount-rate news (last column) over the sample 1933:01-2001:12. The priors and the starting values for the VAR coefficients are based on a fixed coefficient VAR estimated over the sample 1928:12-1932:12.

Tables 3 and 4 report the betas for the 25 ME- and BE/ME-sorted portfolios and the 20-risk sorted portfolios.[10] Note that value stocks are *not* characterized by larger cash-flow betas. Table 5 shows R^2 and Mean Pricing Error for the three models that were introduced in section 3. Both the ICAPM and CAPM turn out to have a very poor performance over the second subsample. On the other hand, the explanatory power of the factor model is extremely high on both sub-samples.

While the estimates point toward a substantial change in the law of motion of the state variables, the poor performance of the ICAPM confirms that the way agents form expectations might not have evolved along the same lines. It seems possible that even if agents were aware of the changes of the parameters, they would still extract a lot of information from a regime that has characterized an important period of American history. In a model with time-varying coefficients agents update the estimates for the *current* parameters and then use them to form expectations. Therefore, whereas the time-varying model might be an ideal tool to describe long-term dynamics, revealing the time evolution of the parameters, it could be unfit to capture the expectation mechanism because it implicitly assumes that events that occurred far enough in the past have no impact on the way agents form expectations today.

[10]The results presented here are based on the smoothed estimates of the VAR coefficients. This makes them directly comparable with the benchmark model with fixed coefficients.

Fama and French portfolios

	Growth	2	3	4	Value
Small	0.3181	0.2627	0.2159	0.1936	0.1846
2	0.3229	0.2516	0.2059	0.1961	0.2009
3	0.3147	0.2375	0.2059	0.1751	0.1944
4	0.3011	0.2404	0.2052	0.1902	0.2126
Large	0.2307	0.2174	0.1948	0.1721	0.1698

Risk-sorted portfolios

	Lo $\widehat{b}_{r_M}$	2	3	4	Hi $\widehat{b}_{r_M}$
Lo $\widehat{b}_{VS}$	0.1280	0.1873	0.2059	0.2600	0.3310
Hi $\widehat{b}_{VS}$	0.1558	0.1986	0.2623	0.2964	0.3595
Lo $\widehat{b}_{TY}$	0.1459	0.1999	0.2448	0.2898	0.3495
Hi $\widehat{b}_{TY}$	0.1445	0.1921	0.2295	0.2781	0.3358

Table 3: Cash Flow Betas for the subsample 1963:07-2001:12 when the news are computed according to a Time-Varying VAR.

Fama and French portfolios

	Growth	2	3	4	Value
Small	1.1761	1.0131	0.9104	0.8536	0.8742
2	1.1613	0.9600	0.8676	0.8121	0.8903
3	1.1027	0.9146	0.8072	0.7546	0.8298
4	1.0095	0.8791	0.8078	0.7468	0.8146
Large	0.8156	0.7701	0.7013	0.6361	0.6475

Risk-sorted portfolios

	Lo $\widehat{b}_{r_M}$	2	3	4	Hi $\widehat{b}_{r_M}$
Lo $\widehat{b}_{VS}$	0.5655	0.6731	0.7817	0.9408	1.1409
Hi $\widehat{b}_{VS}$	0.6021	0.7370	0.8635	1.0246	1.2406
Lo $\widehat{b}_{TY}$	0.6596	0.7355	0.8726	1.0151	1.2722
Hi $\widehat{b}_{TY}$	0.5865	0.6837	0.7720	0.9153	1.0916

Table 4: Discount Rate Betas for the subsample 1963:07-2001:12 when the news are computed according to a Time-Varying VAR.

R^2 (MPE)	CAPM	ICAPM	Factor
pre '60s	62.27% (1.2616)	37.58% (2.0873)	71.34% (0.9583)
post '60s	2.18% (1.5629)	18.89% (1.2959)	64.18% (0.5723)

Table 5: Explanatory power of the three models when cash-flow and discount rate news are computed according to a Time-Varying VAR.

9 Conclusions

In light of their unusual statistical properties, exceptional 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 gets the lion's share. It might be the case that these features play an important role in shaping the way agents form expectations. Until now, much of the attention has been devoted to *rare events betas*, betas computed *during* a rare event. However, *regular times betas* could still be informative if rare events have an impact on the way agents think about financial markets. This new channel could help reconcile the opposing views of Barro (2006) and Julliard and Ghosh (2008).

In support of this idea, I have pointed out that the ability of the "bad beta, good beta" ICAPM to account for the value and size anomalies depends on whether the data includes the early '30s. Given that the ICAPM is a model that imposes economically motivated restrictions, it is unlikely that its performance is uniquely determined by the sample choice. An alternative explanation is that some valuable information might be hidden in those early years, characterized by the market crash that opened the Great Depression

Using a Markov-switching VAR, I have shown that it is in fact possible to isolate a *'30s regime.* This regime is characterized by some interesting properties that set it apart from what happens over the reminder of the sample. Specifically, the value spread predicts low market returns and the state variables tend to be more autocorrelated. It turns out that in order to maximize the explanatory power of the ICAPM, a large weight must be assigned to the *'30s regime*, even if there is little evidence of it occurring again after the first decade. In deriving this result, I have introduced a generalization of the present value decomposition of Campbell and Shiller to allow for Markov-switching dynamics. This approach could be applied to other models that make use of the same methodology and it might provide insight into why the results are often sensitive to the sample choice.

When agents are assumed to form expectations using a Time-Varying VAR, taking into account parameter instabilities but implicitly disregarding what happened far in the past, the ICAPM returns very poor results. This reinforces the idea that rare events and financial crises have a very long lasting impact on the way agents form expectations. Therefore, economically motivated models that aim to explain stock market returns should address the fact that some events are more important than others in revealing the link between risk and return. Such events are likely to be particularly relevant when trying to model agents' expectations using statistical models.

The Great Depression, beginning with the devastating market crash that occurred on October 29, 1929, can certainly be considered a rare event in American history. It is hard 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. Financial markets and economic institutions have evolved significantly since then, but, as current events suggest, this does not imply that investors should disregard what happened in those years.[11] In fact, the '30s can be considered an extreme example of how a stock market crisis and a recession can negatively affect each other. Therefore, if economic theory is looking for a model that links returns to risk, it does not seem wise to leave out the Great Depression and the stock market crash that came with it, since agents are likely to devote a lot of attention to such events.

[11] Bianchi (2010) investigates the similarities between the Great Depression and the Great Recession and their implications for financial markets. Very preliminary results show that there are in fact some similarities.

References

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.

Barro, R., Nakamura, E., Steinsson, J. and Ursúa, J.: 2009, Crises and recoveries in a model of consumption disasters, Columbia University, working paper.

Bianchi, F.: 2003, The betas and the risk premium: An econometric investigation, Bocconi University undergraduate dissertation.

Bianchi, F.: 2010, The Great depression and the Great recession: A view from the stock market (working title), Duke University, working paper.

Bianchi, F., Mumtaz, H. and Surico, P.: 2009, The Great moderation of the term structure of U.K. interest rates, *Journal of Monetary Economics* **56**, 856–871.

Campbell, J. Y.: 1991, A variance decomposition for stock returns, *Economic Journal* **101**(405), 157–179.

Campbell, J. Y., Polk, C. and Vuolteenaho, T.: 2007, Growth or glamour? Fundamentals and systematic risk in stock returns, *Review of Financial Studies* **23**, 305–344.

Campbell, J. Y. and Shiller, R. J.: 1988, Interpreting cointegrated models, *Journal of Economic Dynamics and Control* **12**, 505–522.

Campbell, J. Y. and Vuolteenaho, T.: 2004a, Bad beta, good beta, *American Economic Review* **94**, 1249–1275.

Campbell, J. Y. and Vuolteenaho, T.: 2004b, Inflation illusion and stock prices, *American Economic Review* **94**, 19–23. (Proceedings issue).

Chen, L. and Zhao, X.: 2005, Return decomposition, Michigan State and Kent State University, mimeo.

Cogley, T. and Sargent, T. J.: 2006, Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S., *Review of Economic Dynamics* **8**, 262–302.

Cogley, T. and Sargent, T. J.: 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., Fragoso, M. and Marques, R.: 2004, *Discrete-Time Markov Jump Linear Systems*, Springer, New York.

Daniel, K. and Titman, S.: 2006, Testing factor-model explanations of market anomalies, The Kellogg School at Northwestern University and NBER, mimeo.

Farmer, R. E. A., Waggoner, D. and Zha, T.: 2009, Understanding markov-switching rational expectations models, *Journal of Economic Theory* **144**, 1849–1867.

Gabaix, X.: 2007, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, NBER and NYU working paper.

Jacquier, E., Polson, N. and Rossi, P.: 2004, Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics* **12**, 371–418.

Julliard, C. and Ghosh, A.: 2008, Can rare events explain the equity premium puzzle?, London School of Economics and Political Science, mimeo.

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

Lewellen, J., Nagel, S. and Shanken, J.: 2008, A skeptical appraisal of asset-pricing tests, *Journal of Financial Economics* . forthcoming.

Merton, R.: 1973, An intertemporal capital asset pricing model, *Econometrica* **41**(4), 867–887.

Primiceri, G.: 2005, Time varying structural vector autoregressions and monetary policy, *The Review of Economic Studies* **72**(3), 821–852.

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.

Sims, C. A.: 2003, Implications of rational inattention, *Journal of Monetary Economics* **50**(3), 975–1000.

Sims, C. A.: 2006, Rational inattention: Beyond the linear-quadratic case, *American Economic Review* **96**(2), 158–163. (Proceedings issue).

Sims, C. A., Waggoner, D. F. and Zha, T.: 2008, Methods for inference in large multiple-equation Markov-switching models, *Journal of Econometrics* **146**, 113–144.

Sims, C. A. and Zha, T.: 1998, Bayesian methods for dynamic multivariate models, *International Economic Review* **39**(4), 949–968.

Sims, C. A. and Zha, T.: 2006, Were there regime switches in US monetary policy?, *American Economic Review* **91**(1), 54–81.

Thomas, J. and Zhang, F.: 2007, Inflation illusion and stock prices: Comment, Yale University, working paper.

Uhlig, H.: 2005, What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* **52**, 381–419.

Veronesi, P.: 2004, The peso problem hypothesis and stock market returns, *Journal of*

Economic Dynamics and Control **28**(4), 707–725.

Wei, C. and Joutz, F.: 2009, Inflation illusion or no illusion: What did pre- and post-war data say?, George Washington University, working paper.

A Bayesian algorithms

A.1 Markov-Switching VAR

I consider the most general case where both the VAR coefficients and the covariance matrix can switch and the regimes are assumed to be independent. The algorithms for the other models are simplified versions of the one reported here.

1. Sampling s_t^{Φ} and s_t^{Σ}:

   - Following Kim and Nelson (1999b) I use a Multi-Move Gibbs sampling to draw s_t^{Φ} from $f(s_t^{\Phi}|Z_T, \Phi(\cdot), \Sigma(\cdot), H^{\Phi}, H^{\Sigma}, s_t^{\Sigma})$ and s_t^{Σ} from $f(s_t^{\Sigma}|Z_T, \Phi(\cdot), \Sigma(\cdot), H^{\Phi}, H^{\Sigma}, s_t^{\Phi})$.

2. Sampling $\Phi(s_t^{\Phi})$ and $\Sigma(s_t^{\Sigma})$

   - If $s_t^{\Phi} = s_t^{\Sigma}$, i.e. if the VAR coefficients and the covariance matrix follow a common regime, we can use standard results and sample the VAR coefficients from a normal distribution and draw the covariance matrices from an inverted Wishart distribution.
   - If s_t^{Φ} and s_t^{Σ} are independent we need to proceed in two steps. Given $\Phi(\cdot)$ and s_t^{Φ} we can compute the residuals. Then, given s_t^{Σ}, $\Sigma(\cdot)$ can be drawn from an inverse Wishart distribution. When drawing the VAR coefficients we need to take into account the heteroskedasticity implied by the switches in $\Sigma(\cdot)$. This can be done using GLS or using a Kalman filter. Appendix B describes the two methods.
   - The priors are the same across regimes and are obtained running univariate autoregressions for each endogenous variable:

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

     The prior for the VAR coefficients is:

     $$B = vec\left(\Phi(\cdot)\right) \sim norm\left(B_0, \Sigma(\cdot) \otimes inv(N_0)\right)$$

     The constants and the autoregressive elements of B_0 are based on the univariate regressions, 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 $inv(N_0)$ 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 $inv(N_0)$ are controlled by the term $(\lambda_0\lambda_4)^2$. The choice for the hyperparameters are $\lambda_0 = .5$, $\lambda_1 = [.3, .2, .2, .2]$, $\lambda_2 = 1$, $\lambda_3 = 1$ and $\lambda_4 = 1$.

- The prior for $\Sigma(\cdot)$ is described by an inverse Wishart distribution with mean $S_0 = V_0 diag\left(\{\sigma_i^2\}_{i=1...n}\right)$:

$$\Sigma(\cdot) \sim IW(S_0, V_0)$$

with $V_0 = 11$.

3. Sampling H^Φ and H^Σ:

- Each column of H^Φ and H^Σ is modeled according to a Dirichlet distribution:

$$\begin{aligned} H^\Phi(\cdot, i) &\sim D(a_{ii}^\Phi, a_{ij}^\Phi) \\ H^\Sigma(\cdot, i) &\sim D(a_{ii}^\Sigma, a_{ij}^\Sigma) \end{aligned}$$

where $a_{ii}^\Sigma = 10$, $a_{ij}^\Sigma = 1$, $a_{22}^\Phi = 120$, $a_{21}^\Phi = 1$, $a_{11}^\Phi = 120$, and $a_{12}^\Phi = 5$. Note that the prior on H^Φ implies that regime 1 is less persistent than regime 2. However, no restrictions are imposed on when regime 1 occurred or on its characteristics.

- Given the draws for the state variables s_t^Φ and s_t^Σ, 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:

$$\begin{aligned} H^\Phi(\; : \; , i) \sim D(a_{ii}^\Phi + \eta_{ii}^\Phi, a_{ij}^\Phi + \eta_{ij}^\Phi) \\ H^\Sigma(\; : \; , i) \sim D(a_{ii}^\Sigma + \eta_{ii}^\Sigma, a_{ij}^\Sigma + \eta_{ij}^\Sigma) \end{aligned}$$

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.

A.2 Time-Varying VAR

A.2.1 Priors

VAR coefficients

The prior for the VAR coefficients is obtained *via* a fixed coefficients VAR model estimated over the sample 1928:12 to 1932:12. Φ_0 is therefore set equal to

$$\Phi_0 \sim N(\Phi^{OLS}, V^{OLS})$$

Elements of H_t

Let $\hat{v}^{ols}$ denote the OLS estimate of the VAR covariance matrix estimated on the pre-sample data described above. The prior for the diagonal elements of the VAR covariance matrix (16) is as follows:

$$\ln h_0 \sim N(\ln \mu_0, I_4)$$

where μ_0 are the diagonal elements of $\hat{v}^{ols}$.

Elements of A_t

The prior for the off diagonal elements A_t is

$$A_0 \sim N\left(\widehat{\alpha}^{ols}, V\left(\widehat{\alpha}^{ols}\right)\right)$$

where $\widehat{\alpha}^{ols}$ are the off diagonal elements of $\hat{v}^{ols}$, with each row scaled by the corresponding element on the diagonal. $V\left(\widehat{\alpha}^{ols}\right)$ is assumed to be diagonal with the diagonal elements set equal to 10 times the absolute value of the corresponding element of $\widehat{\alpha}^{ols}$.

Hyperparameters

The prior on Q is assumed to be inverse Wishart

$$Q_0 \sim IW\left(\bar{Q}_0, T_0\right)$$

where $\bar{Q}_0$ is assumed to be $var(\Phi^{OLS}) \times 10^{-4}$ and T_0 is the length of the sample used for calibration.

The prior distribution for the blocks of S is inverse Wishart:

$$S_{i,0} \sim IW(\bar{S}_i, K_i)$$

where $i = 1...4$ indexes the blocks of S. $\bar{S}_i$ is calibrated using $\hat{a}^{ols}$. Specifically, $\bar{S}_i$ is a diagonal matrix with the relevant elements of $\hat{a}^{ols}$ multiplied by 10^{-3}.

Following Cogley and Sargent (2006), I postulate an inverse-Gamma distribution for the elements of G,

$$\sigma_i^2 \sim IG\left(\frac{10^{-4}}{2}, \frac{1}{2}\right)$$

A.2.2 Simulating the Posterior Distributions

The model is a VAR with drifting coefficients and covariances. This model has become fairly standard in the literature and details on the posterior distributions can be found in a number of papers including Cogley and Sargent (2006) and Primiceri (2005). Here, I describe the algorithm briefly.

VAR coefficients Φ_t

The Time-Varying VAR coefficients are drawn using the methods described by Kim and Nelson (1999b).

Elements of H_t

Following Cogley and Sargent (2006), the diagonal elements of the VAR covariance matrix are sampled using the methods described by Jacquier et al. (2004).

Element of A_t

Given a draw for Φ_t the VAR model can be written as

$$A_t'\left(\tilde{Z}_t\right) = u_t$$

where $\tilde{Z}_t = Z_t - a_t - \sum_{p=1}^{P} \Gamma_{t,p} Z_{t-p} = v_t$ and $VAR\left(u_t\right) = H_t$. This is a system of equations with time-varying coefficients and given a block diagonal form for $Var(\varepsilon_t)$ the standard methods for state space models described by Kim and Nelson (1999b) can be applied.

VAR hyperparameters

Conditional on Z_t, $\Phi_{l,t}$, H_t, and A_t, the innovations to $\Phi_{l,t}$, H_t, and A_t are observable, which allows us to draw the hyperparameters—the elements of Q, S, and the σ_i^2—from their respective distributions.

B Drawing the MS-VAR coefficients

Suppose we have draws for $\left\{s_t^{\Phi}\right\}_{t=1\ldots T}$, $\left\{s_t^{\Sigma}\right\}_{t=1\ldots T}$, and $\Sigma(\cdot)$. Define $X_t = [1, Z_{t-1}]'$ and $\Phi(s_t^{\Phi}) = \left[a(s_t^{\Phi}), \Gamma(s_t^{\Phi})\right]$. Rewrite 8 as:

$$Z_t = \Phi(s_t^{\Phi}) X_t + \Sigma(s_t^{\Sigma})^{1/2} \omega_t$$

Given $\left\{s_t^{\Phi}\right\}_{t=1\ldots T}$ we can collect all the draws for $s_t^{\Phi} = j^{\Phi}$, for $j^{\Phi} = 1, \ldots, m^{\Phi}$. We obtain a fixed coefficient VAR with a Markov-switching covariance matrix. Notice that this is a convenient way to model heteroskedasticity.

$$\begin{aligned} Z_{t'} &= \Phi_{j^\Phi} X_{t'} + \Sigma(s^\Sigma_{t'})^{1/2}\omega_{t'} \\ t' &\in \left\{t : s^\Phi_t = j^\Phi\right\} \end{aligned}$$

Our goal is to draw from the posterior distribution of Φ_{j^Φ}. Note that if $\Sigma(\cdot)$ were a diagonal matrix, we could simply divide all the observations equation by equation by the standard deviation of the residuals, i.e. by the square roots of the diagonal elements of $\Sigma(\cdot)$. If we are a not willing to make this assumption we need to proceed in a different way. Here I propose two possible ways to deal with heteroskedasticity: the first one is based on GLS, the second one on the Kalman filter.

B.1 Generalized least squares

The first method is a generalization of the generalized least squares. This method is very straightforward to apply when an univariate regression is involved. Consider the following example:

$$\underset{T\times 1}{Y} = \underset{(T\times k)(k\times 1)}{X\beta} + \underset{(T\times T)(T\times 1)}{C^{-1}\epsilon} \tag{17}$$

where $C^{-1}(C^{-1})' = \Omega = E(\epsilon\epsilon')$ and $\epsilon \sim N(0, I_T)$. Note that Ω_T the covariance matrix of the residuals *across time* and we assume that is not singular. Then we can rewrite 17 as:

$$CY = CX\beta + \epsilon$$

The transformed model allows to apply OLS:

$$\begin{aligned} \beta_{GLS} &= \left(X'C'CX\right)^{-1}\left(X'C'CY\right) \\ &= \left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}Y\right) \end{aligned}$$

We can then combine this estimate with the priors on β following standard results. Note that when there is no correlation between observations the matrix Ω becomes diagonal and each observation is weighted according to the reciprocal of the square root of the variance.

Now consider a VAR and assume that the covariance matrix of the residuals can assume m^Σ possible values. In the present case we will make the assumption that the residuals are not correlated across time, but the results can be generalized.

At each point in time we have

$$\underset{(1\times n)}{y_t} = \underset{(1\times k)(k\times n)}{x_t\beta} + \underset{(1\times n)}{\varepsilon_t} \underset{(n\times n)}{\left(C\left(s_t^{\Sigma}\right)\right)^{-1}}$$

with $\left(C\left(s_t^{\Sigma}\right)^{-1}\right)\left(C\left(s_t^{\Sigma}\right)^{-1}\right)' = \Sigma\left(s_t^{\Sigma}\right)$.

Let y_t^i be the observation at time t for variable i. We can rewrite the model as

$$\underset{(1\times nT)}{\widetilde{Y}} = \underset{(1\times nk)(nk\times nT)}{\widetilde{B}\ \ \widetilde{X}} + \underset{(1\times nT)(nT\times nT)}{\widetilde{\varepsilon}\ \ \widetilde{C}^{-1}}$$

where $\widetilde{Y}$ and $\widetilde{\varepsilon}$ are $(1\times nT)$ vectors obtained ordering all the observation for y_t and ε_t in two row vectors, $\widetilde{B} = vec\left(\beta\right)'$, $\widetilde{X} = \left[bdiag\left(x_1'\right), ..., bdiag\left(x_T'\right)\right]$ and $\left(\widetilde{C}^{-1}\right)\left(\widetilde{C}^{-1}\right)' = \Omega$. Note that Ω is the covariance matrix of $\widetilde{\varepsilon}$. If we assume that there is no cross correlation between residuals over time the matrix Ω is block diagonal and the same thing is true for the matrix $\widetilde{C}^{-1}$. In a model with MS covariance matrix $\Omega = bdiag\left(\left\{\Sigma\left(s_t^{\Sigma}\right)\right\}_{t=1...T}\right)$ and $\widetilde{C}^{-1} = bdiag\left(\left\{\left(C\left(s_t^{\Sigma}\right)\right)^{-1}\right\}_{t=1...T}\right)$.

Once the model has been rewritten in this way we can proceed with the transformation that will make possible to apply OLS:

$$\underset{(1\times nT)(nT\times nT)}{\widetilde{Y}\ \ \widetilde{C}} = \underset{(1\times nk)(nk\times nT)(nT\times nT)}{\widetilde{B}\ \ \widetilde{X}\ \ \widetilde{C}} + \underset{(1\times nT)}{\widetilde{\varepsilon}}$$

From here on standard methods apply.

A simple example will help in understanding the notation. Suppose there are only two endogenous variables $(n = 2)$ and three regressors $(k = 3)$. The sample size is $T = 3$.

We get

$$\begin{bmatrix} y_1^1 \\ y_1^2 \\ y_2^1 \\ y_2^2 \\ y_3^1 \\ y_3^2 \end{bmatrix}' = \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \beta_{12} \\ \beta_{22} \\ \beta_{32} \end{bmatrix}' \begin{bmatrix} x_1' & & x_2' & & x_3' & \\ & x_1' & & x_2' & & x_3' \end{bmatrix}$$

$$+ \begin{bmatrix} \varepsilon_1^1 \\ \varepsilon_1^2 \\ \varepsilon_2^1 \\ \varepsilon_2^2 \\ \varepsilon_3^1 \\ \varepsilon_3^2 \end{bmatrix}' \begin{bmatrix} C\left(s_1^\Sigma\right)^{-1} & & \\ & C\left(s_2^\Sigma\right)^{-1} & \\ & & C\left(s_3^\Sigma\right)^{-1} \end{bmatrix}$$

The block diagonal structure is particularly convenient because $\widetilde{C}$ can be computed taking the inverse of the matrices that lay on the main diagonal. Moreover the variables can be transformed separately for each t.

B.2 The Kalman filter

In this paper, I make use of an alternative method that is based on the Kalman filter algorithm (in fact, it is simply Bayesian updating). The observation equation is given by

$$\begin{aligned} Z_{t'} &= \Phi_{j^\Phi} X_{t'} + \omega_{t'} \Sigma(s_{t'}^\Sigma)^{1/2} \\ t' &\in \Upsilon = \left\{t : s_t^\Phi = j^\Phi\right\} \end{aligned}$$

while the transition equation is

$$\begin{aligned} \Phi_{j^\Phi,i} &= \Phi_{j^\Phi,i-1} \\ i &= 1...\#\Upsilon \end{aligned}$$

The estimate of the VAR coefficients is updated depending on the covariance matrix $\Sigma(s_{t'}^\Sigma)$. When we are in a low volatility regime the observation errors receive a large weight, while the opposite occurs when a high volatility regime is in place. For a detailed description of the Kalman filter see Kim and Nelson (1999a).

C Computing cash-flow and discount rate news in the perfect knowledge case

Consider a MS-VAR:

$$Z_t = a(s_t) + \Gamma(s_t) Z_{t-1} + \Sigma(s_t)^{1/2} \varpi_t$$

where Z_t is a column vector containing n variables and $s_t = 1, ..., m$, with m the number of regimes, evolves following the transition matrix H.

Define the column vectors q_t and ψ_t as:

$$q_t = \left[q_t^{1\prime}, ..., q_t^{m\prime}\right]', q_t^i = E\left(Z_t 1_{s_t=i}\right), \psi_t = \left[\psi_t^{1\prime}, ..., \psi_t^{m\prime}\right]',$$

where ψ_t^i is a column vector with all the elements equal to $p\left(s_t^{\Phi} = i\right)$ and $1_{s_t=i}$ is an indicator variable that is equal to one when regime i is in place and zero otherwise.

Note that:

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

We can define the law of motion of q_t as:

$$\begin{aligned} q_t &= M\psi_t + \Omega q_{t-1} & (18) \\ \psi_t &= G\psi_{t-1} & (19) \end{aligned}$$

where:

$$\begin{aligned} G &= kron(H, I_n) \\ \Omega &= blockdiag\left(\Gamma(s_t = 1), ..., \Gamma(s_t = m)\right) * G \\ M &= diag\left(a(s_t = 1)', ..., a(s_t = m)'\right) \end{aligned}$$

In order to compute the news, we need to make sure that the process described by (18) converges. I adopt the concept of mean-square stability (MSS) of Costa et al. (2004).[12] A system is mean square stable if both the first and second moments converge. Costa et al. (2004) provide conditions for stability of homogenous systems and they show that these conditions can be immediately extended to non-homogeneous systems as long as s_t is an ergodic chain and the shocks are bounded.

[12] This concept of stability has been recently used by Farmer et al. (2009) to provide a solution algorithm to a class of Markov-switching Dynamic Stochastic General Equilibrium models. The formulas presented here slightly differ from the ones of Costa et al. (2004) because of the different time indexation of the regimes.

MSS depends on the eigenvalues of the matrix governing the law of motion of the second moments of the homogeneous system:

$$Z_t = \Gamma(s_t) Z_{t-1}$$

$$Q_t = \Xi Q_{t-1} \tag{20}$$

$$Q_t = \begin{bmatrix} Q_t^1 \\ \vdots \\ Q_t^m \end{bmatrix}, Q_t^i = E\left(Z_t Z_t' 1_{s_t=i}\right)$$

$$\Xi = blockdiag\left(\Gamma(s_t=1) \otimes \Gamma(s_t=1), ..., \Gamma(s_t=m) \otimes \Gamma(s_t=m)\right) * G$$

Theorem 3.9 of Costa et al. (2004) states the system is MSS if and only if all eigenvalues of Ξ are inside the unit circle.

To compute the news, define:

$$\begin{aligned} q^i_{t+j,t} &= E_t\left(Z_{t+j} 1_{s_{t+j}=i}\right) = E\left(Z_{t+j} 1_{s_{t+j}=i} | \iota_t\right) \\ e_1' &= [1,0,0,0]',\ r = m*n \end{aligned}$$

where ι_t contains all the information that agents have at time time, including the probability of being in one of the m regimes. Note that $q^i_{t,t} = Z_t p(s_t = i)$.

Now, consider the formula for the discount rate news:

$$N_{DR,t+1} = (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j}$$

The first term is:

$$\begin{aligned} E_{t+1} \sum_{j=1}^{\infty} \rho^j r_{t+1+j} &= \sum_{j=1}^{\infty} \rho^j e_1' w q_{t+1+j,t+1} \\ &= e_1' w \left[\rho q_{t+2,t+1} + \rho^2 q_{t+3,t+1} + \rho^3 q_{t+4,t+1} + ...\right] \\ &= e_1' w \left(I_r - \rho\Omega\right)^{-1} \left[\rho\Omega q_{t+1,t+1} + \left(I_r - \rho G\right)^{-1} \rho M G \psi_{t+1,t+1}\right] \end{aligned}$$

The second term is:

$$\begin{aligned} E_t \sum_{j=1}^{\infty} \rho^j r_{t+1+j} &= \sum_{j=1}^{\infty} \rho^j e_1' w q_{t+1+j,t} \\ &= e_1' w \left(I_r - \rho\Omega\right)^{-1} \left[\rho\Omega q_{t+1,t} + \left(I_r - \rho G\right)^{-1} \rho M G \psi_{t+1,t}\right] \end{aligned}$$

Therefore:

$$\begin{aligned}
N_{DR,t+1} &= e_1' w \left(I_r - \rho\Omega\right)^{-1} \left[\rho\Omega \left(q_{t+1,t+1} - q_{t+1,t}\right) + \left(I_r - \rho G\right)^{-1} \rho M G \left(\psi_{t+1,t+1} - \psi_{t+1,t}\right)\right] \\
&= e_1' w \left(I_r - \rho\Omega\right)^{-1} \left[\rho\Omega \left(q_{t+1,t+1} - \Omega q_t - M G \psi_t\right) + \left(I_r - \rho G\right)^{-1} \rho M G \left(\psi_{t+1,t+1} - G\psi_t\right)\right] \\
&= e_1' w \left(I_r - \rho\Omega\right)^{-1} \left[\rho\Omega v_{t+1}^{q} + \left(I_r - \rho G\right)^{-1} \rho M G v_{t+1}^{\psi}\right] \\
&= e_1' w \left[\lambda^q v_{t+1}^{q} + \lambda^{\psi} v_{t+1}^{\psi}\right]
\end{aligned}$$

where

$$\begin{aligned}
\lambda^q &= \left(I_r - \rho\Omega\right)^{-1} \rho\Omega \\
\lambda^{\psi} &= \left(I_r - \rho\Omega\right)^{-1} \left(I_r - \rho G\right)^{-1} \rho M G
\end{aligned}$$

Then, we can easily compute the residuals:

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

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 MS states) and a set of parameters, it is easy and computationally efficient to compute the entire sequences $v^{q,T}$, $v^{\psi,T}$, and u^T:

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

D Additional graphs

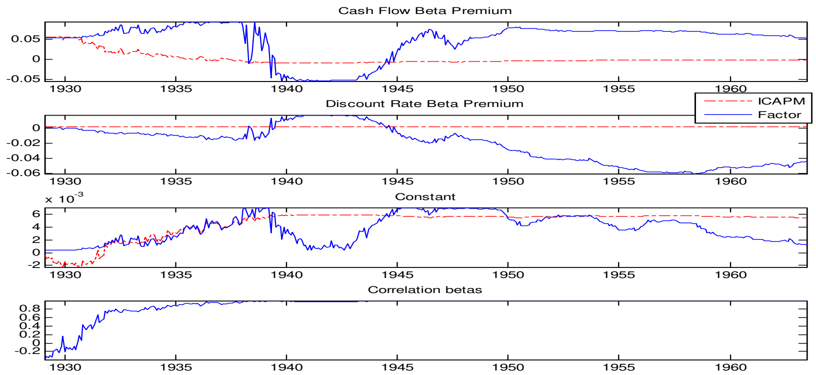

Figure 9: Fixed coeefficient VAR. Evolution of the premia as the sample size shortens. For example: 1940:12 means that the VAR has been estimated over the sample 1940:12-2001:12. The last plot describes the evolution of the correlation between the betas.

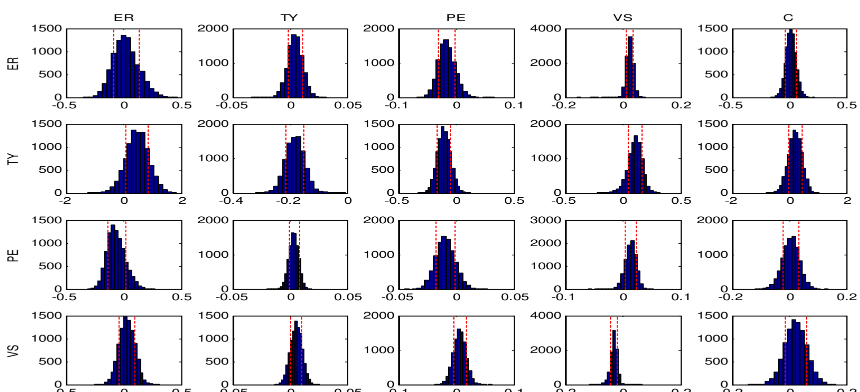

Figure 10: The figure contains histograms and 68% error bands for the pairwise differences of the VAR coefficients across the two regimes. This can be regarded as a "test" for the null the two coefficients are the same across the two regimes.

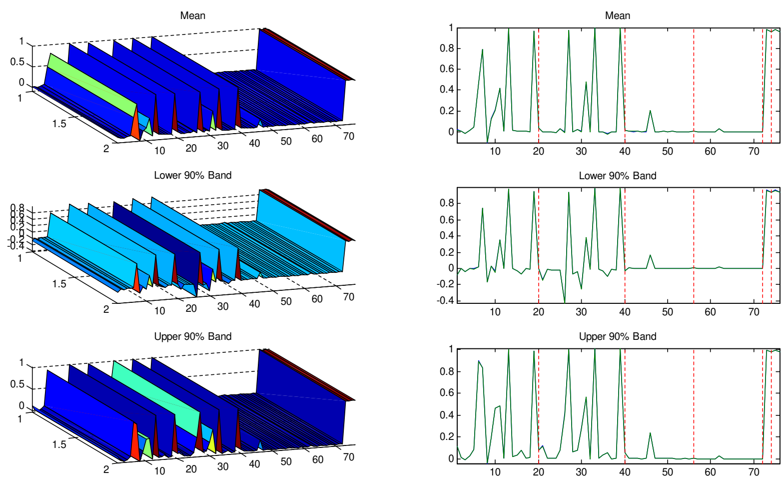

Figure 11: The figure shows means and 90% error bands computed on two not overlapping windows of 225,000 Gibbs sampling draws obtained dividing in two parts the 450,000 draws used in the paper (the retainment rate is 2%).