Bad Beta, Good Beta, and Rare Events*

Francesco Bianchi

Princeton University

March 20, 2008

Abstract

This paper shows that rare events help explaining the cross section of asset returns, the reason being that they are important 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 depends 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. The ICAPM delivers excellent results when investors distinguish between an high- and a low-uncertainty regime.

1 Introduction

The macro-finance literature reveals a growing interest in the role that rare disasters can play in understanding asset returns. Barro (2006), following Rietz (2003), shows that rare

*I am grateful to Chris Sims for useful suggestions at the early stage of this work. I thank Rober Barro, Markus Brunnermeier, John Campbell, Emmanuel Farhi, Alisdair McKay, Motohiro Yogo, Adam Zawadowski, and seminar participants at Princeton University for helpful discussions and comments. Correspondence: 001 Fisher Hall, Department of Economics, Princeton University, Princeton, NJ, 08544, USA. E-mail: fbianchi@princeton.edu.

disasters are potentially important to explain several asset-pricing puzzles, like the high equity premium, low risk-free rate, and volatile stock returns. Gabaix (2007) extends Barro's results allowing for a time-varying intensity of rare disasters. Along the same lines, Farhi and Gabaix (2008) construct a model for exchange rates that yields a theory of the forward premium puzzle.

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 an high equity premium. Hence, if the probability of a disaster is not taken into account, the high risk premium will be ascribed to unreasonably high levels of risk aversion.

However, rare events can be important also for another reason: They are likely to have a large impact on the way agents form expectations. This could happen because they convey information about states of the world that do not occur frequently or because they stand out for having very specific charateristics. Indeed, they often imply very strong comovements in the variables of interest, with the result of being easier to interpret. Moreover, because of their effects, they are also very important to understand. Hence, it could well be that agents end up forming expectations relying heavily on what happened in extreme circumstances.

This paper provides support for this thesis 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 (2004). The model is based on the idea that unexpected (excess) returns can be decomposed into news about future cash flow and news about future discount rates. Accordingly, the usual CAPM beta can be decomposed into two betas, one for each of the two kinds of news. The economically motivated Intertemporal CAPM 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.

To describe the way agents form expectations a VAR is estimated over the sample 1928:12-2001:12. The news are obtained from the residuals according to a transformation of the coefficients. The sample is then divided in two parts: 1928:12-1963:6 and 1963:7-2001:12. The authors compute the cash-flow beta (bad beta) and the discount-rate beta (good beta) for the two sub-samples and they test their explanatory power on two sets of portfolios. Three models are considered: the static CAPM, the ICAPM and a factor model based on the two betas. It is well known that the CAPM does not perform very well after the '60s. On the other hand, the economically motivated model based on the two betas (ICAPM) does very well in both subsamples.

At this point it is important to notice that while the betas are computed over two distinct subsamples, the news are based on the VAR coefficients estimated over the entire sample, 1928:12-2001:12. The authors explain that a larger sample is likely to reduce the estimation bias that arises by having very persistent series. However, 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.

A first contribution of this paper is to show that the remarkable results attained by Campbell and Vuolteenaho depend on including the stock market crash that opened the Great Depression. Figure 1 reports the evolution of R^2 of the three models[1] for the post-60s subsample as the sample size shortens. Notice how excluding the first three years from the sample used in estimating the VAR has a substantial effect on the explanatory power of the ICAPM. When the entire sample 1928:12-2001:12 is used, $R^2 = 51.96\%$. However once the

[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.

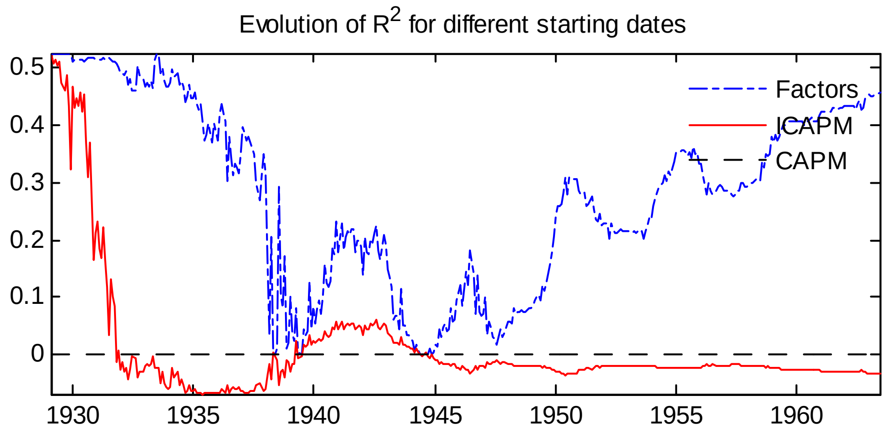

Figure 1: Evolution of R^2 for the three models as the sample used to compute the news shortens. For example: 1940:12 means that the news have been computed according to a VAR estimated over the sample 1940:12-2001:12. While the sample used to estimate the VAR is changing, the models are always tested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

first three years are excluded these results disappear and $R^2 \simeq 0\%$. There are several possible explanations for this failure,[2] but one of them stands out from others in being more thought provoking. As mentioned before, the sample spans the period 1928:12-2001:12. Therefore removing the first three years means to remove the dramatic market crash that marked the beginning of the Great Depression. After all, it is not surprising that these three years have such a large impact on the results of the authors. If investors form expectations giving a large weight to what happened in extreme circumstances, these early years are likely to be particularly important.

The second contribution of this paper is to formalize this intriguing explanation. Using a Markov-switching model with VAR coefficients and volatilities evolving according to two independent chains, I show that it is in fact possible to isolate a *'30s regime*. More importantly, when trying to maximize the explanatory power of the ICAPM it turns out that a large weight must be assigned to this regime, even if there is little evidence for it occurring

[2]Chen and Zhao (2007) and Bianchi (2003) would argue that the return decomposition approach is very sensitive to changes in the sample used to estimate the VAR.

again after the early years. The best version of the model delivers $R^2 = 52.01\%$ when the weight assigned to the *'30s regime* is equal to one. In a Markov-switching model this does not mean that agents expect the *'30s regime* to prevail forever. Instead, it implies that when forming expectations about the near future agents rely heavily on what happens during extraordinary events even if they are aware that in the long run both regimes will occur with a certain probability. This result supports the idea that events that occurred far in the past can still be extremely important in shaping the expectation mechanism, especially if they represent unique events like the Great Depression.

Finally, I consider an alternative version of the model where VAR coefficients and volatilities follow the same regime-switching process. The underlying assumption is that the dynamics of the state variables can be substantially different during periods characterized by high volatility, implying that agents should form expectations in a different way when they are surrounded by more uncertainty. Under the assumption that agents are fully aware of the structure of the model, the explanatory power of the ICAPM is extremely good: $R^2 = 58.26\%$, a value well above the one obtained with the fixed coefficient VAR ($R^2 = 51.96\%$). Considering that the early years of the sample were characterized by high volatility, I regard this as further evidence in favor of the idea pursued in this paper. Even in this case, the data suggest that the ICAPM works well when a large weight is placed on the statistical dynamics that prevail in exceptional or risky times. More importantly, in both cases the early years of the sample are crucial to the extent that they have a large impact on the way investors think about financial markets. Note that this does not imply that agents are not rational. In fact, the best performing model is one where agents form expectations in a very sophisticated and completely rational way.

The point is that many economic models require 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 modelled. In fact, a model could be rejected because the expectation mechanism is not captured adequately. This is what seems to happen when the first years of the sample are removed from the data used to compute the news. Extraordinary events are likely to have large effects on the performance of the stock market and, more in general, on the state of the economy. Moreover, rare events are usually characterized by statistical properties that are by definition very different from the usual ones. Finally, extreme events could be particularly informative about the structure of the economy.

The results of this paper can be linked to some recent contributions in the finance and macro literature. The importance assigned to the early years of the sample seems to echo the ideas of Barro (2006) and Rietz (2003) about the connection between rare events and the equity premium puzzle. Here rare events help to explain the cross section of returns through the expectation mechanism, having a large impact on the way people form expectations and interpret whatever happens on the stock markets. Gabaix (2007) extends Barro's results allowing for a time-varying intensity of rare disasters. He suggests that the value premium could be compensation for distress risk if value stocks did worse than growth stocks during disasters. To some extent this paper provides support for his intuition: when the '30s are removed from the estimates, the betas cannot explain the cross section of returns. However, Gabaix argues that *regular times* betas could be completely uninformative about the risk premia, given that they would not reflect the true risk associated with the different portfolios. His argument implies that to be informative the betas should be computed *during* rare disasters. In this paper *regular times* betas can still be informative as long as we take into account the importance of rare events in shaping the expectation mechanism. Bianchi *et al.* (2007) 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. Their results imply that large estimates of the term premium computed on the basis of fixed-coefficient models may reflect omitted parameter instability or, in other words, a failure in modeling the expectation mechanism.

The role of the Great Depression in shaping expectations is central in Cogley and Sargent (2007). Borrowing from Cecchetti *et al.* (2000) the idea that distorted beliefs may help to explain asset-price anomalies, 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 and initialize a new learning process. Following Friedman and Schwartz (1963), they argue that the Great Depression was one such event. They consider two alternative procedures to capture pessimism: (1) truncate a pre-1933 sample of consumption growth rates to oversample discouraging 1929-1933 observations; and (2) use a robustness calculation to twist an initial prior pessimistically. 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. In a different context Alonso (2006) proposes a model based on the idea of ambiguity: when investors are uncertain about the future, they systematically place a large weight on the least favorable outcome. To capture this idea, she uses the multiple-priors utility specification developed by Gilboa and Schmeidler (1989). Ambiguity aversion is modelled with a "maxmin" formulation: the consumer behaves as if he maximizes expected utility when choosing consumption and an asset portfolio under a worst-case belief that is chosen from a set of conditional probabilities. On the other hand, it could be that agents have a *limited capacity* when it comes to acquire information. In a model of Rational Inattention (Sims (2003, 2006)) agents would probably find optimal to devote more attention to extraordinary events. Such events are likely to be easier to interpret and more important to understand.

Finally, a *caveat* about the methodology used to derive the news. Chen and Zhao (2005) argue that it is potentially misleading to obtain cash-flow news and discount-rate news with the discount-rate news directly modeled but the cash-flow news calculated as the residual, given that discount-rate news cannot be accurately measured. They conduct several tests showing that the results of Campbell and Vuolteenaho (2004) are very sensitive to the set of variables that are included in the VAR. Bianchi (2003) highlights that the results are also very sensitive to the sample chosen to estimate the news. Bianchi (2003) represents the *pars destruens* of a more extended argumentation that finds its *pars costruens* in this paper. Here I offer an explanation for why the sample choice is so important.

The content of this paper can be summarized as follows. Section 2 explains how to obtain the news starting from a fixed coefficient VAR and it introduces the ICAPM. Readers that are familiar with Campbell and Vuolteenaho (2004) might want to skip this section. Section 3 makes clear how important the early years of the sample are: once the first years of the sample are removed the ICAPM delivers poor results. In section 4, using a time-varying VAR, I show that the parameters of the model were substantially different at the beginning of the sample but they have settled down since then. Section 5 shows that using a Markov-switching model it is possible to isolate a *'30s regime*. As a side product, I describe how to estimate a Markov-switching VAR in reduced form. In section 6 I point out that a large weight must be assigned to the *'30s regime* in order to maximize the explanatory power of the ICAPM and that the model delivers better results when agents form expectations assuming that variables evolve in a different way when uncertainty is high. Section 7 concludes.

2 The ICAPM

2.1 The Main Idea

The CAPM fails to describe average realized stock returns since the early 1960s, if 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 invoking 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 in two sub-betas: one reflecting the covariance with news about future cash flows, the other linked to news about discount rates. The previous argument suggests that given two assets with the same CAPM beta, the one with the highest cash-flow beta should have a larger return.

Using an intertemporal capital asset pricing model (ICAPM) along the lines of the one proposed by Robert Merton (1973), Campbell and Vuolteenaho (2004) show 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.

2.2 From the VAR representation to the betas

Using the loglinear approximation for returns introduced by Campbell and Shiller (1988) and Campbell (1991), unexpected excess returns can be approximated by:

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

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

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

Consider a VAR in companion form:

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

where Z_t is a vector containing four state variables: 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 news can be obtained according to the following transformation of the residuals:

$$\begin{aligned} r_{t+1} - E_t r_{t+1} &= e_1' u_{t+1} \\ N_{CF,t+1} &= \left(e_1' + e_1' \lambda\right) u_{t+1} \\ N_{DR,t+1} &= e_1' \lambda u_{t+1} \\ \lambda &= \rho\Gamma \left(I - \rho\Gamma\right)^{-1} \end{aligned}$$

The residuals of the four equations receive weights reflecting their explanatory power in the equation for the excess return and their persistence. The first effect is captured by $\rho\Gamma$, the second by $\left(I - \rho\Gamma\right)^{-1}$.

Finally the betas can be computed according to the following formulas:

$$\widehat{\beta}_{i,CF} = \frac{\widehat{cov}(r_{i,t}, N_{CF,t})}{\widehat{var}(N_{CF,t} - N_{DR,t})} + \frac{\widehat{cov}(r_{i,t-1}, N_{CF,t-1})}{\widehat{var}(N_{CF,t-1} - N_{DR,t-1})} \tag{2}$$

$$\widehat{\beta}_{i,DR} = \frac{\widehat{cov}(r_{i,t}, -N_{DR,t})}{\widehat{var}(N_{CF,t} - N_{DR,t})} + \frac{\widehat{cov}(r_{i,t-1}, -N_{DR,t-1})}{\widehat{var}(N_{CF,t-1} - N_{DR,t-1})} \tag{3}$$

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 usual market beta used in the CAPM can be obtained summing the two betas.

The final step consists in assessing if the two betas represent a substantial improvement over the single market beta. Following Campbell and Vuolteenaho (2004) I test three models: the standard CAPM, a factor model based on the two betas and the economically motivated Intertemporal CAPM.

Note that the methodology proposed by Campbell and Vuolteenaho (2004) and followed in this paper is based on a two-pass procedure. In the first pass the parameters of the VAR are estimated and then used to transform the residuals into news. Then, in the second pass, the betas are computed and used to fit the cross section of asset returns. An interesting alternative would be to use a one-pass procedure. This method would recognize that the fit of the model that is being tested provides some additional information about the parameters of the VAR. In a Bayesian framework, this approach could be implemented *via* a Gibbs sampling algorithm augmented with a Metropolis-Hastings step. The large uncertainty surrounding the VAR coefficients would be taken into account and more weight would be given to those draws of the parameters that improve the explanatory power of the model. This could also help the researcher in separating the news from the noise. On the other hand, it would require some substantial changes in the setup of the model and in the interpretation of the results. I believe it is beyond the scope of this paper to explore this alternative procedure and I regard it as an interesting option for future research.

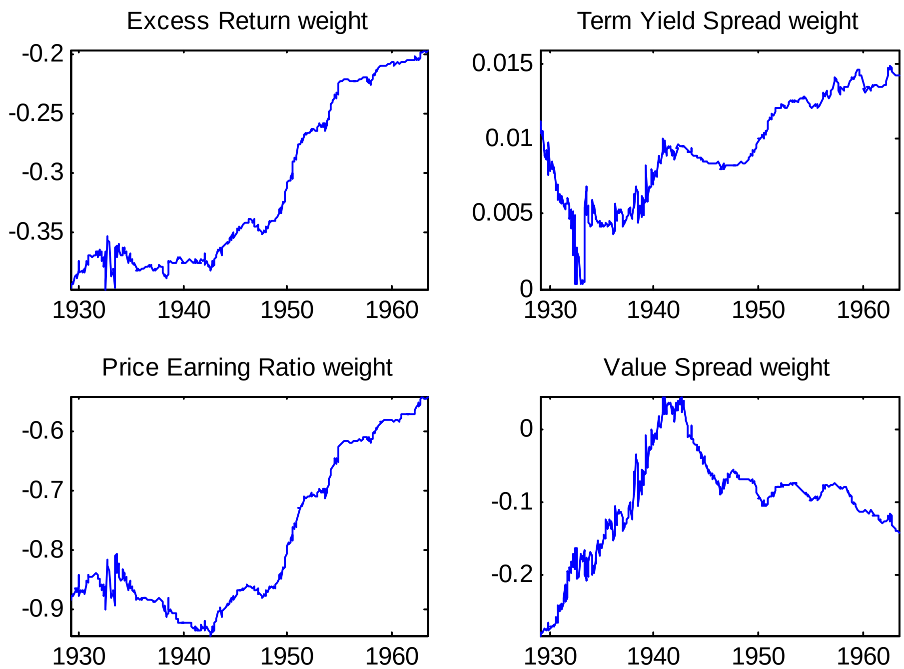

Figure 2: Evolution of the weights used to transform the residuals in discount-rate news as the sample size shortens. For example: 1940:12 means that the VAR has been estimated over the sample 1940:12-2001:12.

3 Excluding the early years

As mentioned before, Campbell and Vuolteenaho (2004) estimate the VAR on the entire sample 1928:12-2001:12. Then they split the analysis in two subsamples: 1928:12-1963:06 and 1963:07-2001:12. Here, 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 start assessing the performance of the different models by computing the news according to a VAR estimated on the entire sample and including the same variables used by Campbell and Vuolteenaho (2004): the excess log return on the CRSP value-weighted index, the term yield spread in percentage points, the log price earning ratio, and the small-stock value-spread.[3] This delivers the same results of Campbell and Vuolteenaho (2004). Then I shorten the sample by a single month at a time, while keeping the methodology described in section 2.2 unchanged. Therefore the sample size for the initial VAR is changing, while the subsample under investigation (1963:07-2001:12) is not.

3.1 The news

As a first step it is useful to consider how the vector used to transform the residuals in news varies while the sample shortens. Figure 2 describes the evolution of the weights used to construct the discount-rate news. The horizontal axis reports the starting date of the sample that is used to estimate the VAR coefficients, while the vertical axis shows the weight that the residual from the specified equation receives when computing $N_{DR,t+1}$. As the initial sample shortens the weights vary significantly. On the other hand, the residuals of the excess return equation are quite stable given that it is in general hard to predict stock returns. This implies that the results that follow are driven by the changes in the weights.

Figure 3 displays how the variance-covariance matrix of the news and the variance of unexpected returns vary as the initial sample shortens. The variance of unexpected returns is basically unaffected by changes in the sample used to compute the news, while the variance of the discount rate news and the covariance between the news move substantially. In particular, including the first years of the sample or not has a relevant impact on the covariance between the 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, if even more observations are excluded the covariance becomes negative. Remember that unexpected returns are equal to

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

the difference between cash-flow news and discount rate news (equation (1)). Therefore:

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

Hence, the sample used to model the expectation mechanism has a large impact on how return innovations are decomposed into news, but a minimal effect on the ability of the model in predicting stock market returns.

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. On the other hand, the ICAPM relies on the covariance between news and portfolio returns. Hence, how unexpected returns are decomposed into news is extremely important for this model. In other words, 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 intuitive argument will find support in what follows.

3.2 The betas and the models

The statistical properties of the news turned out to be extremely sensitive to the choice of the initial sample. Not surprisingly, the same happens with the betas. Cash-flow and discount-rate betas are computed for the 25 Fama-French ME and BE/ME-sorted portfolios and 20 portfolios sorted on past risk loadings with VAR state variables. Figure 4 reports the cash-flow betas and the discount rate betas. 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: the cash flow betas share an upward trend, while all the discount rate betas follow a common negative trend. Furthermore, these movements mimic the patterns of the variances of the news. Finally, the central finding of Campbell and Vuolteenaho, that value stocks have a larger cash-flow beta, disappears once the Great

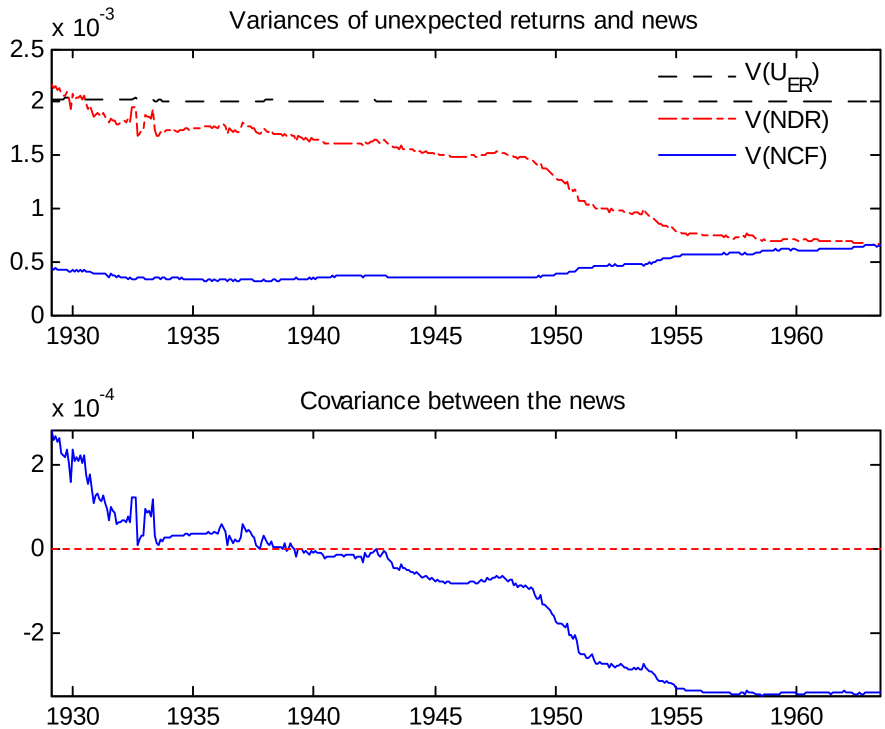

Figure 3: Evolution of the unexpected return variance and of the covariance matrix of the news over the subsample 1963:07-2001:12 as the sample size used to compute the news shortens. For example: 1940:12 means that the VAR has been estimated over the sample 1940:12-2001:12. Unexpected returns are the 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.

Depression has been excluded from the estimates (see solid lines in the left-top panel).

Consider the formulas to compute the betas ((2) and (3)). In the current exercise we are keeping the sample used to compute the betas fixed, hence the portfolio returns cannot be 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 sample choice: it is always very hard to predict stock market returns at monthly frequency. Putting things together we can conclude that these trends are driven by the changes in the estimates of the news that

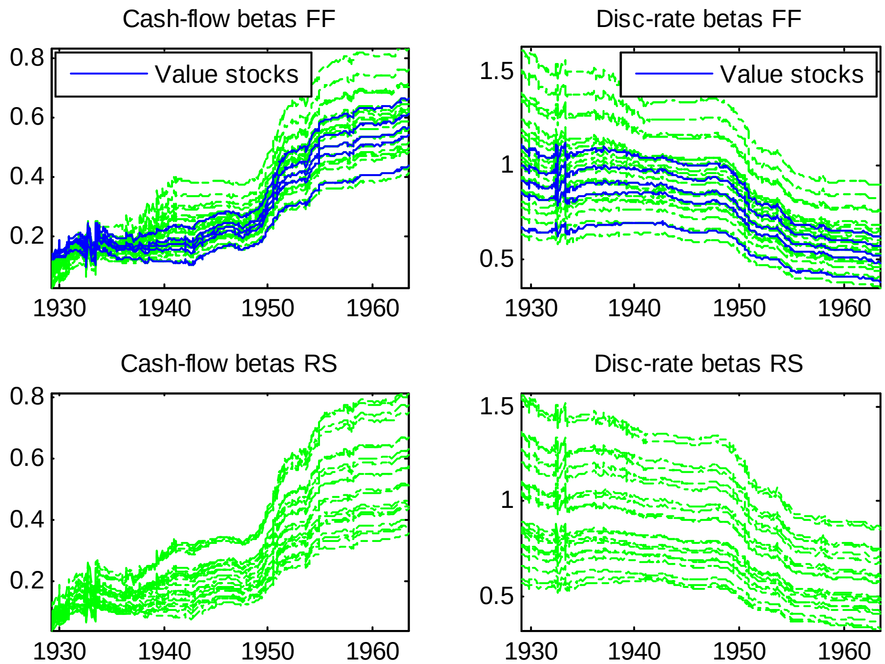

Figure 4: Time evolution of the betas over the subsample 1963:07-2001:12 for different starting dates of the sample used to compute the news. For example: 1940:12 means that the VAR has been estimated over the sample 1940:12-2001:12. The top panels refer to the Fama-French portfolios, the lower panels contain 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.

in turn depend on the VAR coefficient estimates.

Campbell and Vuolteenaho show that differences in the two betas can explain the high returns on value and small stocks relative to the predictions of the traditional CAPM. The idea is that the value of the single market beta is not enough to summarize the risk associated with a particular portfolio. On the other hand, a model based on the two betas is able to capture the sensitivity of the different assets to the two kinds of news that determine unexpected excess returns. Given a value for the market beta, the higher the cash-flow beta, the higher the risk of the portfolio.

Following Campbell and Vuolteenaho I test three models: the static CAPM, the ICAPM and a 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 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 is tested with and without the constant g_0. Excluding the constant is equivalent to impose that a portfolio with both betas equal to zero should deliver the same return of the risk-free asset.

Figure 5 reports the evolution of R^2 of the three models.[4] Notice how excluding the first three years from the sample used in estimating the VAR has a dramatic effect on the ability of the ICAPM of explaining 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, $R^2 = 51.96\%$ 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 the 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 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

[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.

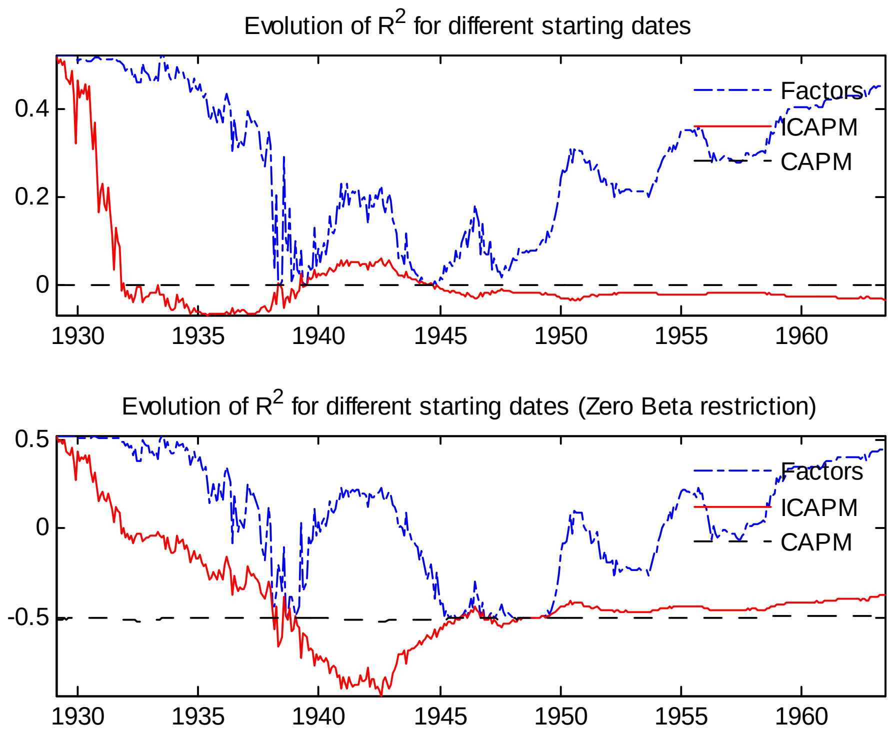

Figure 5: Evolution of R^2 for the three models as the sample used to compute the news shortens. For example: 1940:12 means that the news have been computed according to a VAR estimated over the sample 1940:12-2001:12. While the sample used to estimate the VAR is changing, the models are always tested 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.

in this case very close to the value imposed by the ICAPM ($\widehat{\sigma}_M^2 = 0.02$). 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 the fact that when the Great Depression is removed from the estimates: 1) the key finding of Campbell and Vuolteenaho, that value stocks have a larger cash-flow beta, does not hold anymore; 2) the correlation between the betas from negative

($\simeq -0.38$) becomes strongly positive ($\simeq 0.8$).[5]

Finally the graph shows how the CAPM presents an extremely poor performance independently by the sample used. This is a direct result of the fact that changing the initial sample does not affect the residuals in the excess returns equation.

The results make clear that the coefficient estimates are important for 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 thing that really matters is how agents *interpret* what happens on the stock markets.

4 A Time-Varying approach

From the previous section it emerged that excluding the early years has devastating effects on the ability of the ICAPM in accounting for the value and size "anomalies".

If the statistical properties of the state variables have changed substantially over time it seems worth trying to capture these evolving dynamics in a formal way. Therefore in this section I make use of a Bayesian time-varying approach, aiming to capture these changes.

4.1 The model

As before, the state variables follow a VAR(1) but now both the VAR coefficients and the covariance matrix of the residuals are time-varying:

$$Z_t = a_t + \Gamma_t Z_{t-1} + v_t \tag{4}$$

[5] The correlation between the betas becomes really close to 1 after WWII. This results represent the other side of the coin of the change in the correlation between the news. Graphs of the premia and of the correlation between the betas are reported in appendix D.

where $Z_t = [ER_t, TY_t, PE_t, VS_t]'$ denotes the data matrix and $v_t = \Omega_t^{1/2}\omega_t$ with $\omega_t \sim N(0, I)$. Note that Z_t contains the same variables as before: excess return (ER_t), 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:

$$\Phi_t = \Phi_{t-1} + \eta_t \tag{5}$$

where $\Phi_t = vec\left[a_t, \Gamma_t\right]$.

The covariance matrix of the VAR innovations v_t is factored as

$$VAR\left(v_t\right) \equiv \Omega_t = A_t^{-1} H_t (A_t^{-1})' \tag{6}$$

The time-varying matrices H_t and A_t are defined as:

$$H_t \equiv \begin{bmatrix} h_{1,t} & 0 & 0 & 0 \\ 0 & h_{2,t} & 0 & 0 \\ 0 & 0 & h_{3,t} & 0 \\ 0 & 0 & 0 & h_{4,t} \end{bmatrix} \text{ and } A_t \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha_{21,t} & 1 & 0 & 0 \\ \alpha_{31,t} & \alpha_{32,t} & 1 & 0 \\ \alpha_{41,t} & \alpha_{42,t} & \alpha_{43,t} & 1 \end{bmatrix} \tag{7}$$

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-zero and non-one elements of the matrix A_t evolve as driftless random walks,

$$\alpha_t = \alpha_{t-1} + \varepsilon_t \tag{8}$$

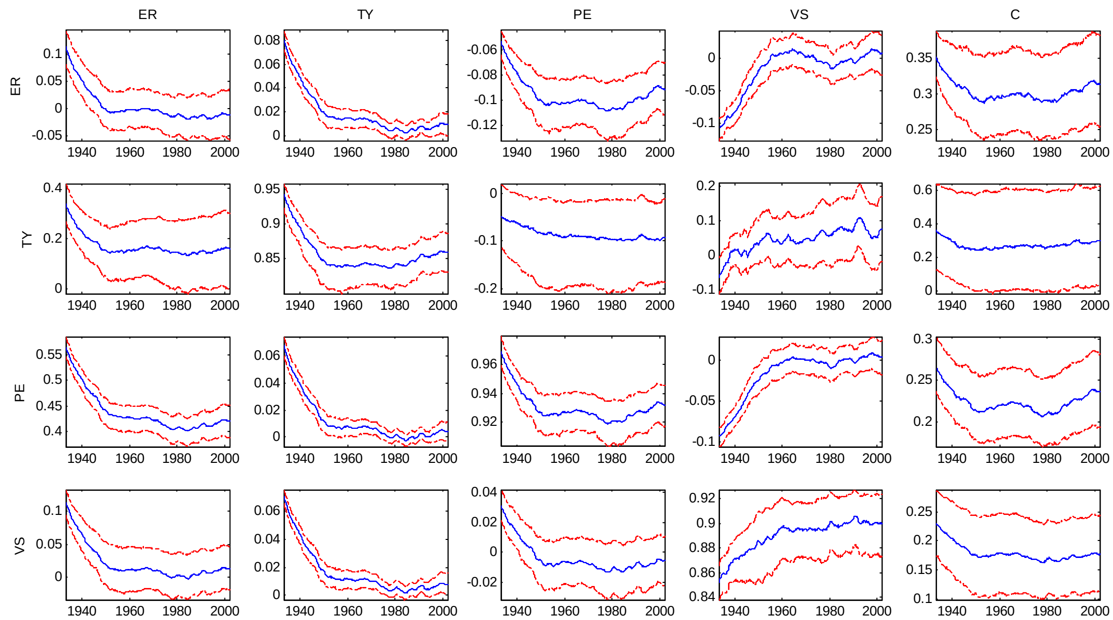

Figure 6: Time Varying VAR estimates: Evolution of the VAR coefficients over the sample 1928:12-2001:12. The graph reports the median and the 90% error bands based on Gibbs sampling draws from the posterior.

The vector $[v_t', \eta_t', \varepsilon_t', u_t']'$ is distributed as

$$[v_t', \eta_t', \varepsilon_t', u_t']' \sim N(0, V), \tag{9}$$

$$V = \begin{bmatrix} \Omega_t & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & G \end{bmatrix} \text{ and } G = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix} \tag{10}$$

The model is estimated using the Bayesian methods described by Kim and Nelson (1999). In particular, I employ a Gibbs sampling algorithm that approximates the posterior distribution (see appendix A for details).

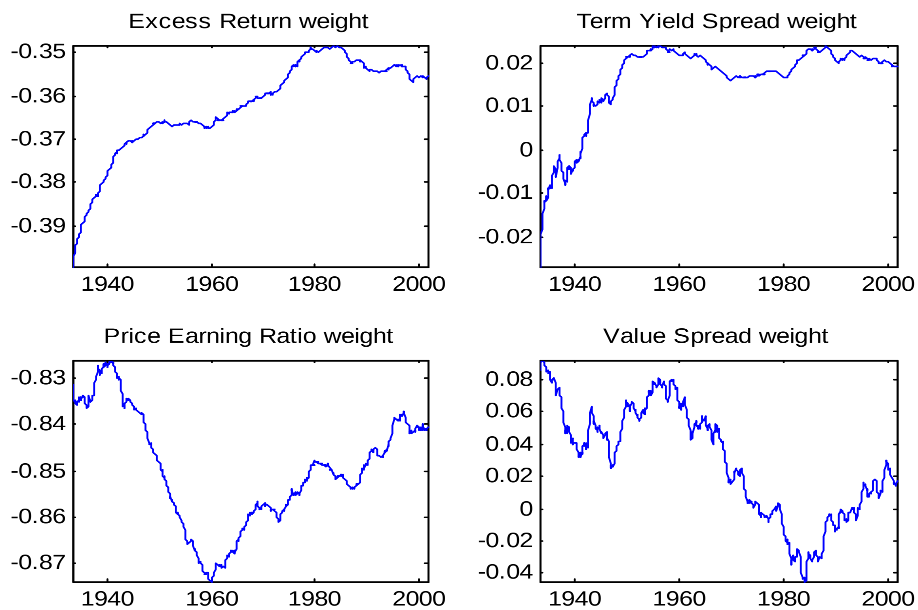

Figure 7: Time-Varying VAR estimates: Evolution of the weights used to transform the residuals in discount-rate news.

4.2 Results

Figures 6 and 7 show, respectively, the evolution of the VAR coefficients and of the weights used to compute the news. It is evident that the VAR coefficients have changed substantially since the early '30s while they have been remarkably stable throughout the remain of the sample. This supports the idea that the early years reflect a unique event with specific statistical properties. In section 3 it has been shown that these early years are extremely important to obtain the results of Campbell and Vuolteenaho. This could be because they simply contain additional information like other randomly selected subsamples. But figure 6 suggests that they are crucial because they convey information from an exceptional event. If investors formed expectations simply taking into account the current state of the world, we would expect them giving a low weight to these years. But the previous results and the ones that follow suggest exactly the opposite.

The results presented here are based on the smoothed estimates of the VAR coefficients. This makes them directly comparable with the ones obtained following the procedure of Campbell and Vuolteenaho.

Note that I am making the implicit assumption that investors are aware, at each point in time, of the parameters of the model. This is not necessarily a realistic assumption. On the other hand, it is also true that investors are likely to have a much larger information set than the one employed in the VAR. In any case, the results should be taken *cum grano salis*.

Tables 1 and 2 report the betas for the 25 ME- and BE/ME-sorted portfolios and the 20-risk sorted portfolios. Even in this case, value stocks are not characterized by a larger cash-flow beta.

The last step consists in testing the three models that were introduced in the previous section. Table 3 reports the results. The ICAPM and CAPM turn out to have a very poor performance. The failure of the ICAPM seems in line with what obtained previously: when excluding the '30s the results change substantially. Here agents are supposed to form expectations according to time-varying coefficients. Therefore, even if the sample includes the '30s, the early years have a very low, if not null, impact on the expectations mechanism that prevails later in the sample.

On the other hand, the factor model still does a good job in explaining the returns of the 44 portfolios, with $R^2 = 64.18\%$.

5 Was the Great Depression a rare event?

This question sounds so trivial that it seems hardly reasonable to spend an entire section to answer it: a simple "yes" would be enough. 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 in clarifying why the early years are so important.

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 1: Cash Flow Beta, News computed according to a TV-VAR (1963:07-2001:12)

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 2: Discount Rate Beta, News computed according to a TV-VAR (1963:07-2001:12)

R^2	CAPM	ICAPM	Factor
pre '60s	62.27%	37.58%	71.34%
post '60s	2.18%	18.89%	64.18%

Table 3: Explanatory power of the three models when news are computed according to a Time-Varying VAR.

The time-varying approach that was used in the previous section suggested that the statistical properties of the state variables have evolved over time. However, if the news are computed according to the time-varying VAR coefficients the ICAPM turns out to perform poorly. A suggestive explanation relies on the idea that the way agents form expectations has not evolved along the same lines. It seems possible that even if agents were aware of the changes of the parameters, they would still place a positive probability on the resurgence of a regime that has characterized an important period of the American economic history. In a model with time-varying coefficients agents update the estimates for the *current* parameters and then they use them to form expectations. Therefore, the time-varying model seems an ideal tool to describe long-term dynamics, pointing out the time evolution of the parameters, but it implies that events that occurred far enough in the past do not have any impact on the way agents form expectations today.

Investors could have in mind a limited number of alternative scenarios. In this case, they would obtain estimates of the parameters under the alternative states of the world and then they would be forming expectations according to the probabilities that they put on each of them. Regimes that are not quite likely to occur in the near future could still be receiving a high weight when forming expectations if they were associated with extraordinary events. A Markov-switching model seems the perfect tool to formalize this idea. Moreover, it should be able to return a clear answer to our original question: Was the stock market crash associated with the Great Depression a rare event?

5.1 The model

As before, the state variables follow a VAR(1) but now both the VAR coefficients and the covariance matrix can switch across regimes:

$$Z_t = a(s_t^\phi) + \Gamma(s_t^\phi)Z_{t-1} + \Sigma(s_t^\Sigma)^{1/2}\omega_t \tag{11}$$

$$\Phi(s_t^\phi) = \left[a(s_t^\phi), \Gamma(s_t^\phi)\right] \tag{12}$$

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 $\omega_t \sim N(0, I)$. Note that Z_t contains the same variables as before: 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. A simplified model for the case in which all the parameters follow a common chain can be obtained assuming $s_t^\Sigma = s_t^\phi$. Note that the regime switch is modelled for a VAR in its reduced form. Sims and Zha (2006) recommend to work 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 like the ones that are included in the present model. Therefore, it seems more reasonable to proceed with this approach than trying to impose restrictions that are hard to justify.

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 a state to another is given by:

$$P[s_t^\phi = i | s_{t-1}^\phi = j] = h_{ij}^\phi \tag{13}$$

$$P[s_t^\Sigma = i | s_{t-1}^\Sigma = j] = h_{ij}^\Sigma \tag{14}$$

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 $y(1)$. In the process we obtain filtered estimates of the state, giving $P[s_t^\phi = i| \{y_s, s \leq t\}, \phi(\cdot), \Sigma(\cdot), H^\phi, H^\Sigma, s_t^\Sigma]$ and $P[s_t^\Sigma = i| \{y_s, s \leq t\}, \phi(\cdot), \Sigma(\cdot), H^\phi, H^\Sigma, s_t^\phi]$ 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 $\left\{s_t^\phi, t = 1, \ldots, T\right\}$ and $\left\{s_t^\Sigma, t = 1, \ldots, T\right\}$ conditional on $\{y_t, t = 1, \ldots, T\}, \phi(\cdot), \Sigma(\cdot), H^\phi, H^\Sigma$.

Under some restrictions on the form of H^ϕ and H^Σ, the model can be estimated using Gibbs sampling to draw from the posterior distribution of the parameters.

Proper priors are put on all the parameters in the model to allow for model comparison. Given that the final goal is to establish if it is possible to isolate a *'30s regime*, strong priors are imposed on the transition matrix for the VAR coefficients as a way to avoid meaningless switches from a regime to another.

5.2 Algorithm

The model in equations (11) to (13) is estimated using Bayesian methods. I report results for the posterior mode and I employ a Gibbs sampling algorithm to approximate the posterior distribution.

The mode estimates are based on the maximization of the log-posterior. 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_t^ϕ and s_t^Σ:

   - Following Kim and Nelson (1999) 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 and the VAR coefficients are sampled from a normal distribution and the covariance matrices are drawn 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 heteroschedasticity 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 and have a Dirichlet distribution.

4. Go to step 1.

I use 30000 Gibbs sampling replications and discard the first 20000 as burn-in. The posterior moments vary little over the retained draws providing evidence of convergence.

5.3 Results

5.3.1 Independent regimes

I start reporting the posterior mode estimates for a MS-VAR with the VAR coefficients and the covariance matrix evolving according to independent chains. The number of regimes is

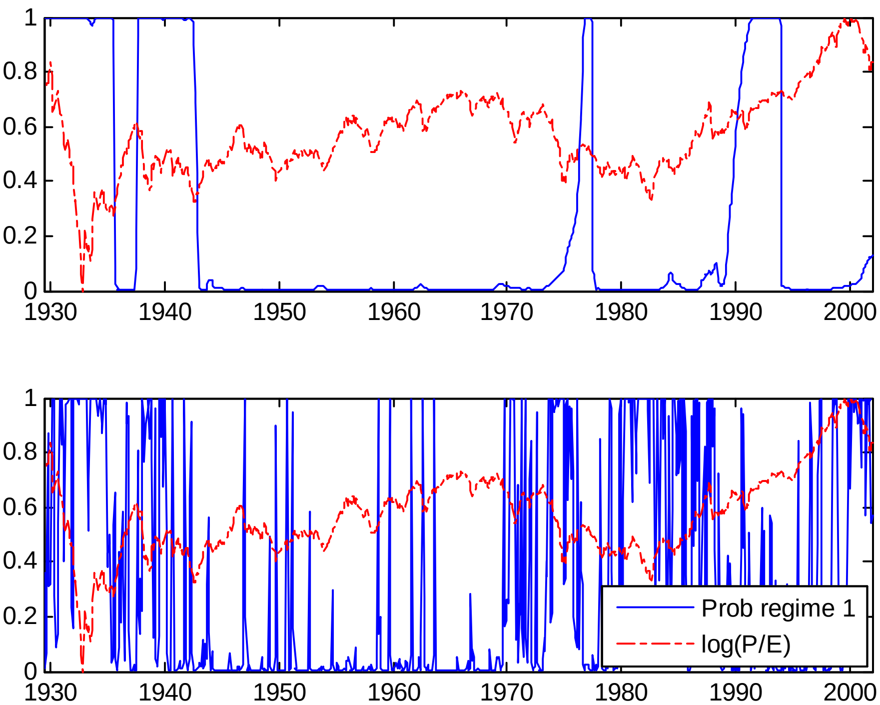

Figure 8: Markow Switching VAR with independent regimes. The top and lower panel report the posterior mode probabilities of regime 1 for the VAR coefficients and the covariance matrix respectively. The graph reports also the log price earning ratio.

equal to two for both chains: $m^{\phi} = 2 = m^{\Sigma}$. Figure 8 shows the smoothed probabilities of $s_t^{\phi} = 1$ and $s_t^{\Sigma} = 1$ in the top and lower panel respectively. Table 3 and 4 report the posterior mode estimates for the VAR coefficients and the covariance matrix.

From looking at table 3 it should be clear that the '30s are characterized by statistical dynamics that are quite different from the ones that prevail on the remaining of the sample. Regime 1 clearly dominates the first decade, even if it occurs again in the mid-'70s and in the early 90s. For this reason, from now on I will refer to regime 1 as the *'30s regime*. Across the two regimes the VAR coefficients show some remarkable differences.[6] The autoregressive

[6]However, the reader should keep in mind that we are talking about point estimates.

$s_t^\phi = 1$	ER_t	TY_t	PE_t	VS_t	$const$		w_{NDR}
ER_{t+1}	0.0355	0.0020	−0.0096	−0.0131	0.0493	u_{ER}	−0.5484
TY_{t+1}	−0.0289	0.9796	−0.0106	−0.0347	0.1489	u_{TY}	0.0698
PE_{t+1}	0.5228	−0.0016	0.9987	−0.0063	0.0141	u_{PE}	−1.0450
VS_{t+1}	−0.0345	−0.0016	−0.0013	1.0076	−0.0057	u_{VS}	0.5326

$s_t^\phi = 2$	ER_t	TY_t	PE_t	VS_t	$const$		w_{NDR}
ER_{t+1}	−0.0019	0.0027	−0.0130	0.0187	0.0186	u_{ER}	−0.3653
TY_{t+1}	0.4428	0.8076	−0.0894	0.1394	0.1336	u_{TY}	0.0057
PE_{t+1}	0.4022	0.0006	0.9946	0.0161	−0.0080	u_{PE}	−0.9148
VS_{t+1}	0.0030	0.0005	0.0013	0.9780	0.0264	u_{VS}	−0.0792

Table 4: Posterior mode estimates, VAR coefficients, MS-VAR with independent regimes

$s_t^\Sigma = 1$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0067	–	–	–
u_{TY}	0.0010	0.1697	–	–
u_{PE}	0.0034	0.0007	0.0028	–
u_{VS}	−0.0003	0.0000	−0.0003	0.0053

$s_t^\Sigma = 2$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0011	–	–	–
u_{TY}	−0.0000	0.0177	–	–
u_{PE}	0.0006	−0.0000	0.0006	–
u_{VS}	0.0000	−0.0000	−0.0000	0.0006

Table 5: Posterior mode estimates, Covariance matrix, MS-VAR with independent regimes

component for the excess return (ER) is relatively large and positive for the *'30s regime*, while it is small and negative for regime two. The autoregressive component for the term yield spread (TY) is substantially closer to one when the *'30s regime* prevails. The coefficient of the excess return in the price earning equation is also significantly larger for the *'30s regime*. Finally, the coefficient of the value spread (VS) in the excess return equation is negative for the *'30s regime*, while it is positive for regime 2. Table 4 shows also the weights used to transform the residuals into discount-rate news according to the two different regimes. The main differences occur for the weights given to the residuals of the term yield spread and of the value spread.

For what concerns the covariance matrix, table 5 shows that regime 1 is associated with more uncertainty. The probability of this regime is very high during the early years of the sample but it increases again at different points in time.

Figure 7 reports also the log price earning ratio.[7] Note how sudden negative changes in the price-earning ratio are often associated with an increase of the probability of the high volatility regime ($s_t^{\Sigma} = 1$). Furthermore, the stock market crashes of the '30s come with a high probability of the *'30s regime* ($s_t^{\Phi} = 1$).

So far I have shown results for the posterior mode. In what follows I present results based on draws from the posterior distribution.

Figure 9 is the analogous of figure 8. The graph shows the median and the 80% error bands for the probabilities of $s_t^{\phi} = 1$ and $s_t^{\Sigma} = 1$. For what concerns the VAR coefficients, regime 1 still dominates the early years of the sample. On the other hand, it is somehow harder to find evidence of resurgence of this regime after the '30s. The results for the covariance matrix are shown in the lower panel. They look more stable and they show the same pattern of the previous ones. The high volatility regime characterizes the '30s, but it shows up several other times after that.

[7] The variable is normalized to fit in the graph.

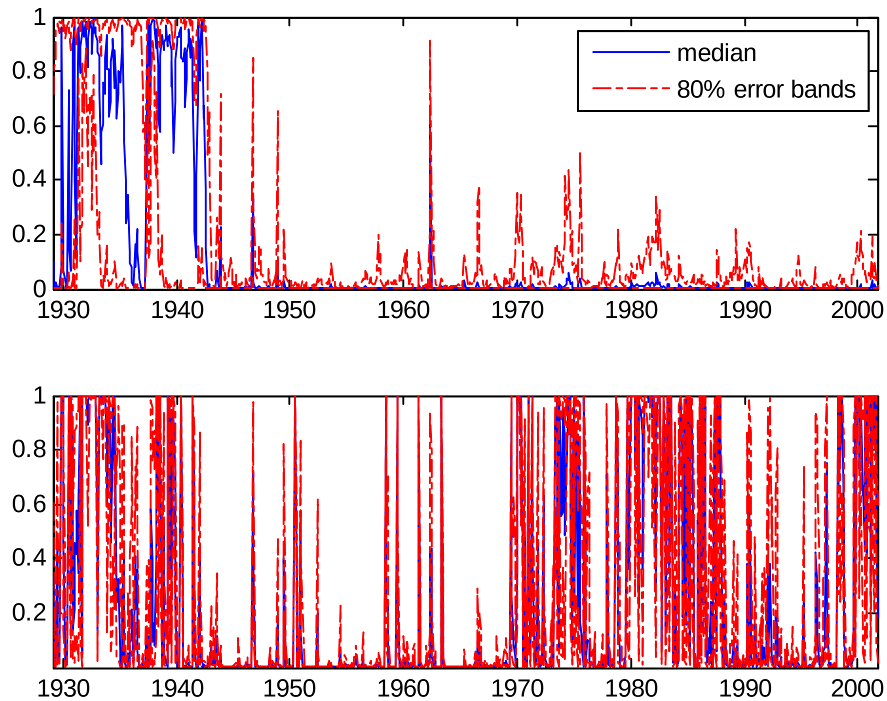

Figure 9: Markow Switching VAR with independent regimes. Gibbs sampling draws from the posterior of the probabilities of regime 1. The top panel refers to the VAR coefficients, while the lower panel contains the results for the covariance matrix.

Table 6 and 7 contain the means and standard errors for the VAR coefficients and covariance matrix according to the two regimes. The results are somehow different from the ones that have been shown before. However, the basic patterns can still be discerned. The autoregressive component of the excess return and the coefficient of this same variable in the price earning ratio are larger for the *'30s regime*. On the other hand, there is not any significant difference in the autoregressive coefficient of the term yield spread. Note that the standard errors are quite large and only some coefficients turn out to be tightly estimated.

$s_t^{\Phi}=1$	ER_t	TY_t	PE_t	VS_t	$const$		w_{NDR}
ER_{t+1}	0.0728 (0.0570)	0.0048 (0.0124)	−0.0298 (0.0196)	0.0293 (0.0166)	0.0029 (0.0136)	u_{ER}	−0.3257
TY_{t+1}	0.0488 (0.2131)	0.9266 (0.0260)	−0.0435 (0.0331)	0.0970 (0.0478)	−0.0170 (0.0627)	u_{TY}	0.0130
PE_{t+1}	0.5489 (0.0458)	0.0016 (0.0092)	0.9805 (0.0143)	0.0206 (0.0115)	−0.0025 (0.0096)	u_{PE}	−0.6503
VS_{t+1}	−0.0351 (0.0505)	−0.0021 (0.0074)	0.0092 (0.0106)	0.9908 (0.0105)	0.0035 (0.0116)	u_{VS}	0.5644

$s_t^{\Phi}=2$	ER_t	TY_t	PE_t	VS_t	$const$		w_{NDR}
ER_{t+1}	0.0212 (0.0551)	0.0018 (0.0074)	−0.0085 (0.0092)	0.0119 (0.0147)	0.0136 (0.0128)	u_{ER}	−0.3602
TY_{t+1}	0.2313 (0.2200)	0.9199 (0.0255)	−0.0203 (0.0264)	0.0500 (0.0513)	0.0265 (0.0644)	u_{TY}	0.0092
PE_{t+1}	0.4232 (0.0364)	0.0004 (0.0038)	0.9984 (0.0059)	0.0077 (0.0093)	−0.0072 (0.0097)	u_{PE}	−0.8925
VS_{t+1}	0.0060 (0.0516)	0.0001 (0.0069)	0.0070 (0.0164)	0.9750 (0.0206)	0.0173 (0.0118)	u_{VS}	0.0432

Table 6: MS-VAR coefficients, Independent regimes, MCMC estimates

$s_t^{\Sigma}=1$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0085 (0.0021)	–	–	–
u_{TY}	0.0013 (0.0030)	0.2278 (0.0555)	–	–
u_{PE}	0.0041 (0.0010)	0.0011 (0.0019)	0.0033 (0.0030)	–
u_{VS}	−0.0006 (0.0006)	−0.0007 (0.0027)	−0.0004 (0.0004)	0.0065 (0.0018)

$s_t^{\Sigma}=2$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0014 (0.0002)	–	–	–
u_{TY}	−0.0001 (0.0002)	0.0267 (0.0062)	–	–
u_{PE}	0.0007 (0.0001)	−0.0001 (0.0002)	0.0007 (0.0001)	–
u_{VS}	0.0000 (0.0001)	−0.0000 (0.0002)	−0.0000 (0.0000)	0.0009 (0.0002)

Table 7: MS-VAR covariance matrices, Independent regimes, MCMC estimates

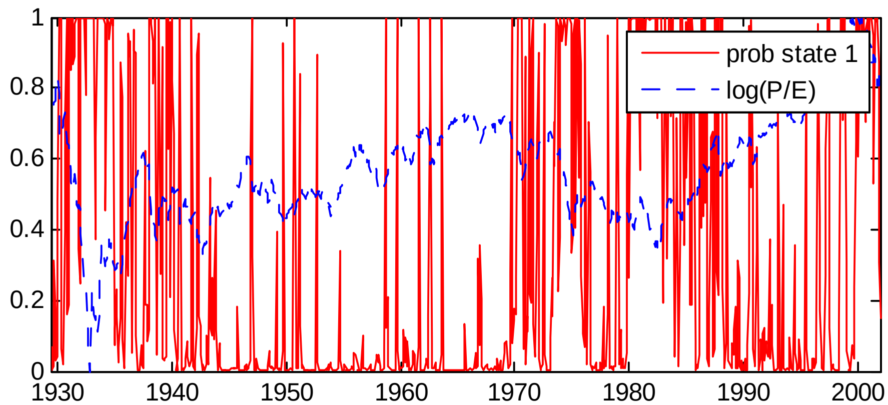

Figure 10: Markow Switching VAR with VAR coefficients and covariance matrix evolving according to a common chain. The graph reports the posterior mode probabilities of regime 1 and the log price earning ratio.

5.3.2 Common regimes

I consider now a MS model with VAR coefficients and covariance matrix evolving according to the same chain ($s_t^{\Sigma} = s_t^{\phi}$). Table 8 and 9 report the posterior mode estimates for the VAR coefficients and the covariance matrix, while figure 10 shows the probability of regime 1.

Regime 1 is characterized by high volatility and it occurs with high probability during the '30s and the '80s, but it shows up again in the '70s and at the end of the sample. Note how the posterior mode estimates of the covariance matrix are very close to the ones obtained under the assumption of independent regimes. Once again the VAR coefficients show some remarkable differences across the two regimes.[8] The autoregressive component for the excess return (ER) is large and positive for regime one, while it is large and negative for regime two. The coefficient of the excess return in the price earning equation is significantly larger in regime 1. The coefficient of the value spread (VS) in the excess return equation is negative for regime 1, while it is positive for regime 2. These differences resemble the results for the

[8]However, the reader should keep in mind that we are talking about point estimates. The results based on the MCMC simulation are available upon request.

case with independent regimes. However, the autoregressive component for the term yield spread (TY) is now substantially closer to one under regime 2.

5.3.3 Model comparison

The previous results make clear that when allowing for regime changes, a *'30s regime* emerges from the data. This result is interesting by itself because it gives an insight into why the early years are so relevant for the performance of the ICAPM. However, it is still an open question if a MS-VAR has to be preferred to the fixed parameter VAR to describe the Data Generating Process. Comparing the marginal data density of the models is a way to answer this question.

Let θ be a $(k \times 1)$ vector containing all the parameters of the model. Moreover denote the likelihood function and the prior density by $p(Y_T|\theta)$ and $p(\theta)$ respectively. The marginal data density is given by:

$$p(Y_T) = \int p(Y_T|\theta)p(\theta)d\theta \tag{15}$$

The modified harmonic mean (MHM) method of Gelfand and Dey (1994) can be used to approximate (15) numerically. This method is based on the following result:

$$p(Y_T)^{-1} = \int_{\Theta} \frac{h(\theta)}{p(Y_T|\theta)p(\theta)} p(\theta|Y_T)d\theta \tag{16}$$

where Θ is the support of the posterior probability density. The weighting function $h(\theta)$ is a probability density whose support is contained in Θ. A numerical approximation of the integral on the right hand side of (16) can be obtained by montecarlo integration:

$$\begin{aligned} \widehat{p}(Y_T)^{-1} &= \frac{1}{N}\sum_{i=1}^{N} m\left(\theta^i\right) \\ m\left(\theta^i\right) &= \frac{h(\theta^i)}{p(Y_T|\theta^i)p(\theta^i)} \end{aligned}$$

$s_t^\phi = 1$	ER_t	TY_t	PE_t	VS_t	$const$		w_{NDR}
ER_{t+1}	0.1621	0.0144	−0.0211	−0.0206	0.0776	u_{ER}	−0.3961
TY_{t+1}	0.0541	0.7511	−0.0508	0.1634	0.0065	u_{TY}	0.0069
PE_{t+1}	0.5976	0.0073	0.9875	−0.0102	0.0421	u_{PE}	−0.8247
VS_{t+1}	0.0330	0.0097	0.0076	0.9747	0.0223	u_{VS}	−0.0973

$s_t^\phi = 2$	ER_t	TY_t	PE_t	VS_t	$const$		w_{NDR}
ER_{t+1}	−0.1032	−0.0033	−0.0077	0.0079	0.0236	u_{ER}	−0.3621
TY_{t+1}	0.1927	0.9567	−0.0219	0.0299	0.0412	u_{TY}	0.0471
PE_{t+1}	0.3422	−0.0031	1.0003	0.0083	−0.0102	u_{PE}	−0.9825
VS_{t+1}	−0.0573	−0.0026	−0.0040	1.0009	0.0098	u_{VS}	−0.5125

Table 8: MS-VAR coefficients, Common regimes, posterior mode

$s_t^\Sigma = 1$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0062	–	–	–
u_{TY}	0.0020	0.1673	–	–
u_{PE}	0.0031	0.0013	0.0026	–
u_{VS}	−0.0005	0.0001	−0.0004	0.0054

$s_t^\Sigma = 2$	u_{ER}	u_{TY}	u_{PE}	u_{VS}
u_{ER}	0.0012	–	–	–
u_{TY}	−0.0001	0.0204	–	–
u_{PE}	0.0006	−0.0001	0.0006	–
u_{VS}	0.0000	0.0001	−0.0000	0.0006

Table 9: MS-VAR Covariance matrices, Common regimes, posterior mode

where θ^i is the ith draw from the posterior distribution of $p(\theta|Y_T)$. As long as $m(\theta)$ is bounded above the montecarlo approximation converges at a practical rate.

Gertler (1999) suggests an implementation based on the posterior simulator. The weighting function $h(\theta)$ is a truncated multivariate Gaussian density.[9] The mean $\overline{\theta}$ and the covariance $\overline{\Omega}$ are obtained from the posterior simulator. To ensure the boundness condition, choose $p \in (0,1)$ and take

$$h(\theta) = p^{-1} N\left(\theta; \overline{\theta}, \overline{\Omega}\right) I_{\widehat{\Theta}_M}$$

$$\widehat{\Theta}_M = \left\{\theta : \left(\theta - \overline{\theta}\right)' \overline{\Omega}^{-1} \left(\theta - \overline{\theta}\right) \leq \chi^2_{1-p}(k)\right\}$$

where $I_{\widehat{\Theta}_M}$ is an indicator function that is equal to one when $\theta \in \widehat{\Theta}_M$. If $\widehat{\Theta}_M \subsetneq \Theta$, the domain of integration needs to be redefined as $\widehat{\Theta}_M \cap \Theta$.

Table 10 reports the log marginal data density for diffetent values of p. A smaller value of p implies a better behavior of $m(\theta)$ over the domain $\widehat{\Theta}_M$, but also a greater simulation error due to a smaller number of draws $\theta^i \in \widehat{\Theta}_M$. The first model does not allow for switches in any of the parameters: it is the standard fixed coefficients VAR. The second and the third model allow for changes in the VAR coefficients and the covariance matrix, the difference between the two being that the former assumes that s_t^{ϕ} and s_t^{Σ} are independent, while the latter assumes a common regime. Finally the fourth model allows for switches in the covariance matrix only. The model with independent regimes clearly dominates both the fixed coefficient model and the Markov-switching model with common regimes. On the other, it is dominated by a model with a Markov-switching covariance matrix but fixed coefficients. However, it is important to take into account that strong priors were imposed on the transition matrix of

[9] Sims *et al.* (2006) point out that while the approach proposed by Geweke works generally well when dealing with fixed coefficients models, problems can arise when it is applied to Markov-switching models. Their argument relies on the fact that in this kind of models the posterior tends to be Non-Gaussian. Therefore, they suggest to replace the Gaussian distribution with elliptical distributions centered at the posterior mode. However, the two methods return similar results in the present case, probably because of the low dimensionality of the Markov-switching model.

Model	$p = 0.1$	$p = 0.3$	$p = 0.5$	$p = 0.7$
Fixed parameters	$1,397$	$1,396$	$1,396$	$1,396$
MS Common regimes	$1,743$	$1,744$	$1,715$	$1,716$
MS Independent regimes	$1,797$	$1,793$	$1,794$	$1,775$
MS Covariance only	$1,853$	$1,846$	$1,821$	$1,821$

Table 10: Marginal data density (log)

the VAR coefficients to avoid uninteresting switches across regimes. This choice has a large impact on the log-posterior of the MS-VAR with independent regimes.

The results that have been shown so far support the idea that the '30s convey valuable information about an exceptional event. When a fixed coefficient VAR is estimated over the entire sample the evidence referring to this rare event is incorporated in the estimates, even if not explicitly recognized. Note that this is not necessarily something we should try to avoid because in this contest the final goal is not to come out with an estimate of the data generating process, but to represent the way agents form expectations. The results make clear that the *'30s regime* does not occur frequently after the early years of the sample, but this does not mean that it does not have an impact on the way investors form expectations. Therefore, as long as some weight is put on the occurrence of the *'30s regime* (directly or indirectly), the ICAPM returns good results. The following section formalizes this idea.

6 The importance of the Great Depression

From what has been shown it seems plausible that the Great Depression represented an exceptional event not only for the real economy, but also for what concerns the statistical properties of the financial variables. It is, therefore, interesting to ask what role this rare event has in the formation of expectations. It could be that agents consider the *'30s regime* as a memory of the past, given that there is little evidence for it occurring again after the first decade of the sample. On the other hand, they could put a certain probability on its occurrence. The fact that the ICAPM works only when the news are computed according to a fixed coefficients VAR that includes the '30s suggests that these years are important in shaping agents' expectations.

In what follows I test this idea considering three alternative specifications. Suppose agents have clear in mind that the state variables can evolve according to two distinct 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 unexpected returns and 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 incoming news. On the opposite side of the spectrum there is the *naive* model. In this case agents simply choose the probabilities to assign to the two states without updating them. They use these weights not only to form expectations, but also to decompose the expectation error. Finally, I consider an intermediate case: the *sophisticated* model. In this setting agents are aware of the transition matrix H^{ϕ} and they use it when computing the news. However, when forming expectations they come up with their own initial weights instead of using the estimated probabilities.[10]

In what follows, I first consider the MS-VAR with independent regimes to show the importance of the *'30s regime*. Then I examine the MS-VAR with common regimes where the focus is on the role played by periods characterized by high uncertainty. The two sets of results are reported in tables 11 and 12 respectively.

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

6.1 Independent regimes

Let's start considering the *naive* model: Agents are aware of the two regimes and they assign time-invariant probabilities to them without any update. This means that they form expectations and they compute news assuming that these probabilities are not going to change in the future. Note that this assumption is computationally convenient because the formulas of section 2.2 can be applied once Γ has been defined as:

$$\Gamma = w_{\Gamma_1}\Gamma_1 + (1 - w_{\Gamma_1})\,\Gamma_2$$

where Γ_1 and Γ_2 are the posterior mode estimates of the VAR coefficients. I use the posterior modes instead of the means because the difference in the posterior is particularly large (as explained by Sims, Wagoner, and Zha (2007) this happens quite frequently with Markov-switching models). According to the posterior mode estimates the *'30s regime* is not relegated to the first years of the sample given that it shows up again in the mid-'70s and in the early '90s. The definition "*'30s regime*" refers to the fact that this regime dominated the years of the Great Depression.

What are the weights that maximize the explanatory power (R_I^2) of the ICAPM? If the model is tested over the first subsample (1928:12-2001:12), the ICAPM achieves $R_I^2 = 61.28\%$ with $w_{\Gamma_1} = 0.7030$. Note how the explanatory power of the model is large and it is obtained assigning a weight to the *'30s regime* that is larger than its historical probability. However, it is more interesting to look at the subsample 1963:07-2001:12. In this case the explanatory power of the ICAPM is maximized when $w_{\Gamma_1} = 0.5531$ and the maximum is $R_I^2 = 49.89\%$. This value is large and very close to the one that was obtained with the fixed coefficient VAR ($R^2 = 51.96\%$). Moreover, once again, the weight assigned to the *'30s regime* is remarkably high, especially if compared with the probability of occurrence of this regime during the second half of the sample.

The results are even more suggestive under the *sophisticated* specification. At each point in time agents assign the same probabilities to the two regimes (i.e. $w_{\Gamma_1,t,t} = w_{\Gamma_1}\ \forall t$), but they are aware that in the long run both regimes will prevail. Therefore, when forming expectation they update the probabilities of the two regimes according to the transition matrix H^{Φ} (i.e. $w_{\Gamma_1,t+i,t} = \left(H^{\Phi}\right)^i w_{\Gamma_1}$). Under these hypothesis, the fit of the ICAPM model is now maximized when $w_{\Gamma_1} = 1$ and R_I^2 is remarkably high: 52.01% for the second subsample, 62.25% for the first subsample. Note how the result for the second subsample outperforms what it was obtained with the fixed coefficient VAR ($R^2 = 51.96\%$).

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 $p_{t+1} = H^{\Phi}p_t$. Then, once the expectation error is revealed, they update their beliefs and they compute the news according to the MS process that drives the model. A detailed explanation of how to compute the news in the *perfect knowledge* case can be found in Appendix C. In this case the results for the second subsample are unsatisfactory ($R_I^2 = -6.42\%$), while they still are good over the first subsample $R_I^2 = 61.39\%$. Given the low weight that is assigned to the *'30s regime* over the second subsample (see figure 8), the poor performance associated with this specification is not surprising and in line with the results obtained with the Time-Varying model.

These results strongly support the original hypothesis: the weights given to rare events are not necessarily linked to the actual probability of their occurrence. Agents could give a larger weight to states of the world that would affect their welfare extensively if they occurred. Alternatively, it could be that some events are simply easier to interpret and more important to understand. Therefore agents would end up forming expectations relying heavily on few relevant observations. The result for the *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 it comes to form expectations for the *immediate* future they end up putting all the weight on the *'30s regime*. Obviously this does not mean that they expect the Great Depression to occur with probability one. Rather, it suggests that the dynamic properties that prevailed during those years have a large impact on how agents think about financial markets.

When a fixed coefficient VAR is estimated over the entire sample, evidence from different states of the world is mixed together. If the data were really characterized by a unique regime, the performance of the model should not be so sensitive to small variations in the sample used to compute the news. Obviously it could be that the data that are removed contain much more information than the remaining ones. However, this explanation does not seem very appealing once it has been taken into account that over a sample of more than 800 observations, it is enough to remove few of them to cause the ICAPM to fail. But independently from which one of the two hypothesis is more realistic, they both reflect the same fundamental idea: when trying to reproduce with statistical models the way investors form expectations, some events are more important than others. This could happen because these events convey information about states of the world that are not frequently observed or because they highlight the characteristics of the current (and unique) data generating process.

6.2 Common regimes

Let's now assume that agents believe that uncertainty in the financial markets shows up with specific dynamic features. In other words, investors have in mind a Markov-switching model in which covariance matrices and VAR coefficients share a common chain. Here the goal is to highlight the role of periods characterized by more uncertainty. In practice, these periods turn out to be often associated with a decrease in the Price Earning ratio, for this reason I distinguish between a *risky times regime* (regime 1, high volatility) and a *regular*

Model	w_{Γ_1}	Sample	ICAPM	Factor
Naive	0.7030	pre 60s	61.28%	–
Naive	0.5531	post 60s	49.89%	–
Naive	0.7480	pre 60s	–	69.89%
Naive	0.7439	post 60s	–	54.74%
Sophisticated	1	pre 60s	62.25%	–
Sophisticated	1	post 60s	52.01%	–
Sophisticated	0.7100	pre 60s	–	69.32%
Sophisticated	1	post 60s	–	52.07%
Perf. knowl.	$p_{\Gamma_1,t}$	pre 60s	61.39%	70.12%
Perf. knowl.	$p_{\Gamma_1,t}$	post 60s	−6.42%	17.82%

Table 11: The news are computed according to a MS-VAR with the VAR coefficients and the covariance matrix evolving according to independent chains.

times regime (regime 2, low volatility).

Let's start asking what are the weights over the two regimes that maximize R_I^2 in the context of the *naive model*. Over the second subsample the maximum is 46.73% and it is achieved with $w_{\Gamma_1} = 0.3854$, while if the first subsample is considered the maximum is attained with $w_{\Gamma_1} \simeq 1$ ($R_I^2 = 59.30\%$).

Moving to the *sophisticated model*, it turns out that the explanatory power of the ICAPM is maximized when a large weight is given to the *risky times regime*. Over the subsample 1928:12-1963:6 the maximum is 62.83% with $w_{\Gamma_1} = 1$, while over the subsample 1963:7-2001:12 the maximum is 33.16% with $w_{\Gamma_1} = 0.7350$.

However the most interesting result is obtained when considering the *perfect knowledge* model. Under this assumption the ICAPM achieves $R_I^2 = 63.20\%$ and $R_I^2 = 58.26\%$ respectively for the first and the second subsample. Note that the result for the second subsample represents a clear improvement over the fixed coefficient model. On the other hand, the ICAPM under the *perfect knowledge* specification had a very poor performance when using the MS model with independent regimes. To understand why the results are so different, it is enough to remember that the probability of the *'30s regime* is basically zero for large part of the sample, while the probability of the *risky times regime* in the MS with common regimes becomes high at different points in time.

Considering that the probability of the *risky times regime* is particularly high during the early years of sample, it is clear that the two sets of results go in the same direction. In both cases the data suggest that the ICAPM works well when a large weight is placed on the statistical dynamics that prevail in exceptional or risky times. More importantly, in both cases the early years of the sample are crucial to the extent that they have a large impact on the way investors think about financial markets.

Many economic models require to take a stance on how agents form expectations. This implies that when testing them 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 modelled. In fact, a model could be rejected because the expectation mechanism is not captured adequately. This is what seems to happen when the first years of the sample are removed from the data used to compute the news. Extraordinary events are likely to have large effects on the performance of the stock market and, more in general, on the state of the economy. Moreover, rare events are usually characterized by statistical properties that are by definition very different from the usual ones. More in general, it could be that there is more information in few extreme events than in a large number of usual ones. The Great Depression can surely be considered a unique event in American history and it started with the devastating market crash that occurred on October 29, 1929, the Black Tuesday. Even if it is not clear if the market crash was the cause or a consequence of the Great Depression, there is no doubt that the two events are intimately related. Therefore, if economists are looking for a model that links returns to risk, it does not seem to be a good idea to leave out an event like the Great Depression or the stock market crash that came with it.

Model	w_{Γ_1}	Sample	ICAPM	Factor
Naive	1	pre 60s	59.30%	–
Naive	0.3854	post 60s	46.73%	–
Naive	$\simeq 0$	pre 60s	–	71.13%
Naive	$\simeq 0$	post 60s	–	51.54%
Sophisticated	1	pre 60s	62.83%	–
Sophisticated	0.7350	post 60s	33.16%	–
Sophisticated	1	pre 60s	–	70.45%
Sophisticated	$\simeq 0$	post 60s	–	49.64%
Perf. knowl.	$p_{\Gamma_1,t}$	pre 60s	63.20%	73.19%
Perf. knowl.	$p_{\Gamma_1,t}$	post 60s	58.26%	58.77%

Table 12: The news are computed according to a MS-VAR with the VAR coefficients and the covariance matrix evolving according to a common chain.

6.3 Twisting the transition matrix

In section 6.1 it has been shown that when a large weight is assigned to the *'30s regime*, the ICAPM returns good results. On the other hand, when assuming that agents are fully aware of the underlying model (*perfect knowledge*), the ability of the ICAPM to explain the value and size anomalies vanishes. I argued that this is a direct consequence of the low weight assigned to the *'30s regime* over the second subsample. In the model of section 6.2 stochastic volatilities and VAR coefficients move following the same Markov process. In this case, the statistical properties of the *'30s regime* are naturally linked to the uncertainty prevailing in the stock market and the explanatory power of the ICAPM turns out to be extremely good. This implies that the weight assigned to the *30's regime* by investors is likely to be time-varying. Therefore, the next step would be to consider a model that allows for time-varying weights. In fact, such a model would return a *perfect* fit. However, without any further restriction, this kind of model implies so many degrees of freedom that is completely useless.

A possible solution is to tie the weights to the estimated probabilities. Let agents be aware of the parameters of the Markov-switching model with independent regimes and able to observe the probabilities assigned to the two regimes. However, assume that when forming

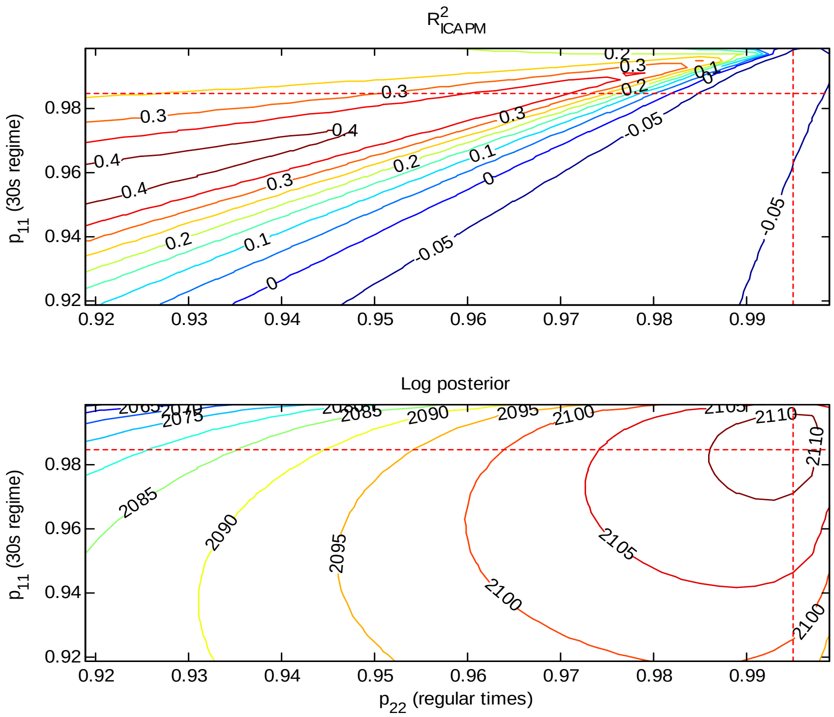

Figure 11: Contour plots of R_i^2 and of the (log) posterior for different values of h_{11}^{ϕ} and h_{22}^{ϕ}, the parameters describing the persistence of the VAR coefficients regimes. All other parameters are kept constant at their posterior mode estimates.

expectations they relay on a transition matrix that can differ from the one arising from the posterior mode estimates. The desired result follows: the weights are time-varying, but they are linked to the estimated probabilities.

The top panel of figure 11 shows the contour of R_I^2 for different values of h_{11}^{ϕ} and h_{22}^{ϕ}, the parameters that capture the persistence of the VAR coefficients regimes. The horizontal and vertical lines mark the posterior mode estimates. The values of h_{11}^{ϕ} and h_{22}^{ϕ} span the interval from 0.92 to 1 and form a discretization grid with squares of width 0.02.[11] Values lower than 0.92 are disregarded because associated with values of the posterior that are too low if compared with the mode. In other words there is a limit to the extent to which agents'

[11] Actually, I use the range $0.919 - 0.999$ to avoid having an absorbing state.

beliefs can deviate from what is implied by the data. The lower panel of figure 11 reports the contour of the (log) posterior.[12]

Two results emerge. First, the explanatory power of the ICAPM increases as the persistence of the *regular times regime* decreases. This result is hardly surprising. It mirrors the main finding of the previous sections: the ICAPM delivers good results when agents put some weight on the occurrence of the *30s regime*. Second, the explanatory power of the ICAPM is not monotonic in h_{11}^{ϕ}, i.e. in the persistence of the *'30s regime*. In particular, as the persistence of the regular times regime decreases, too low values of h_{11}^{ϕ} are not desirable. This result reinforces an argument that was presented before: agents seem to pose a relatively large weight on the *'30s regime*, but they are well aware that in the long run both regimes will prevail with a certain probability. The result can also be explained analyzing the statistical properties of the two regimes: the *'30s regime* is not stable, therefore, while it is a valid instrument to form expectations for the *short run*, it is not suitable to predict what is going to happen in the *long run*.

Summarizing, even relatively small deviations from the posterior mode estimates are sufficient to attain good values for R_I^2. On the one hand, the ICAPM asks for a relatively low persistence of the *regular times regime* to give the *'30s regime* a chance to show up, on the other hand, this must come with a decrease in h_{11}^{ϕ} to guard against long run dynamics that do not appear very likely.

7 Conclusions

In this paper I have pointed out that rare events can help explaining the cross section of asset returns, the reason being that they have an important role in shaping expectations.

I have started showing how the ability of the "bad beta, good beta" ICAPM to account

[12] Results for an analogous exercise based on the MS-VAR with common regimes are available upon request.

for the value and size anomalies (Campbell and Vuolteenaho (2004)) depends on including the stock market crash that opened the Great Depression. When excluding the first three years of the sample used by the authors (1928:12-2001:12) the results change substantially, with the explanatory power of the model going to zero. It seems unlikely that the results of Campbell and Vuolteenaho are uniquely determined by the sample choice. It can well be that some valuable information is contained in these early years that saw the occurrence of the dramatic market crash that opened the Great Depression.

Using a time-varying model I have shown that the statistical dynamics of the data were substantially different during the '30s, but they leveled off after that period. However, if the news are computed assuming that agents form expectations according to the time-varying parameters the ICAPM yields poor results. This reinforces the idea that the early years of the sample are important in shaping agents' expectations: Rare events have a very long lasting impact on the way agents form expectations. To test this idea a Markov-switching model has been fitted to the data. The results show that it is in fact possible to isolate a *'30s regime*. Moreover, when trying to maximize the explanatory power of the ICAPM it turns out that a large weight must be assigned to this regime, even if there is little evidence for it occurring again after the first decade. The best specification delivers $R_I^2 = 52.01\%$ when the weight assigned to the *'30s regime* equals one.

Furthermore, I have proposed an alternative model that relies on the distinction between high and low uncertainty times and that represents a substantial improvement over the fixed coefficient model. When agents form expectations taking into account the probability of the two regimes and the law of motion that governs them, the explanatory power of the ICAPM over the subsample of interest (1963:07-2001:12) increases to $R_I^2 = 58.26\%$. Note that the two sets of results go in the same direction because there is no doubt that the early '30s were characterized by high uncertainty. Therefore, the statistical properties of what I called the *'30s regime* are still having a large impact on the way agents form expectations.

The emphasis posed on the role of rare events should not be interpreted as a claim against the rational expectation assumption. In fact, the best performing model is one in which agents form expectations in a completely rational way. They make use a Markow-switching model to infer the regime-dependent VAR coefficients and the probabilities associated with them. Then they update these probabilities according to a transition matrix. This can be considered a triumph for the rational expectation hypothesis.

Instead, the central message of this paper is that economically motivated models that aim to explain stock market returns should deal with the fact that some events are more important than others in shaping agents' expectations. This aspect becomes particularly important when conducting empirical research, because the researcher is implicitly testing the assumptions made on how agents form expectations. A model could indeed be rejected simply because the expectation mechanism is not captured adequately. This is what seems to happen when the first years of the sample are removed from the data used to compute the news.

The Great Depression can surely be considered a unique event in American history and it started with the devastating market crash that occurred on October 29, 1929, the Black Tuesday. It is hard to establish if the stock market crash was a cause or a consequence of the Great Depression, however there is not doubt that the two events were closely related. Financial markets and economic institutions have evolved significantly since then, but this does not imply that investors should disregard what happened in those years. In fact, the '30s can be considered an emblematic example of how a stock market crisis and a recession can worsen each other. Therefore, if economic theory is looking for a model that links returns to risk, it does not seem to be a good idea to leave out an event like the Great Depression or the stock market crash that came with it: Agents are likely to devote a lot of attention to such kind of events.

References

Alonso, I., 2006, Ambiguity in a Two-Country World, Yale University, mimeo.

Barro, R. J., 2006, Rare Disasters and Asset Markets in the Twentieth Century, *Quarterly Journal of Economics* **121**, 823–866.

Bianchi, F., 2003, The Betas and the Risk Premium: An Econometric Investigation, Bocconi University undergraduate dissertation.

Bianchi, F., H. Mumtaz and P. Surico, 2007, Nominal Yields, Inflation Expectations and Monetary Policy, Bank of England working paper.

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

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

Campbell, J. Y. and T. Vuolteenaho, 2004, Bad Beta, Good Beta, *American Economic Review* **94**, 1249–1275.

Cecchetti, S.G., Pok sang Lam and N. C. Mark, 2000, Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to Be True?, *American Economic Review* **90**(4), 787–805.

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

Cogley, T. and T. J. Sargent, 2006, Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S., *Review of Economic Dynamics* **8**, 262–302.

Cogley, T. and T. J. Sargent, 2007, The Market Price of Risk and the Equity Premium: A Legacy of the Great Depression?, *Journal of Monetary Economics* . forthcoming.

Farhi, E. and X. Gabaix, 2008, Rare Disasters and Exchange Rates, Harvard University and NYU, mimeo.

Friedman, M. and A.J. Schwartz, 1963, *A Monetary History of the United States, 1857-1960*, Princeton University Press, Princeton, New Jersey.

Gabaix, X., 2007, Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance, NBER and NYU working paper.

Gertler, M., 1999, Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication, *Econometric Reviews* **18**(1), 1–73.

Gilboa, I. and D. Schmeidler, 1989, Maxmin Expected Utility with Non-unique Prior, *Journal of Mathematical Economics* **18**(2), 141–153.

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

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

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

Rietz, T. A., 2003, The Equity Risk Premium: A Solution, *Journal of Monetary Economics* **22**, 117–131.

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., D. F. Waggoner and T. Zha, 2006, Methods for Inference in Large Multiple-Equation Markov-Switching Models, Atlanta Fed working paper.

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

A Bayesian algorithms

A.1 Time-Varying model

A.1.1 Priors

VAR coefficients

The prior for the VAR coefficients is obtained via a fixed coefficients VAR model estimated over the sample 1973:01 to 1973:12. Φ_0 is therefore set equal to

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

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 (7) is as follows:

$$\ln h_0 \sim N(\ln \mu_0, I_6)$$

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 $var(\hat{\phi}^{OLS}) \times 10^{-4}$ 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 $i = 1..6$ 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.1.2 Simulating the Posterior Distributions

Time-Varying VAR

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 (1999).

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 (1999) 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.

A.2 Markov-Switching model

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 (1999) 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 heteroschedasticity 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:

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

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 autoregressive elements of B_0 are equal to the AR(1) coefficients, while all the other elements are set to zero. As in Sims and Zha (1998), the variance of the prior distribution is specified by a number of hyperparameters that pin down N_0. The choice of hyperparameters implies a fairly loose prior for the VAR coefficients. Let λ be a (5×1) vector containing the hyperparameters. The diagonal elements of $inv(N_0)$ 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 $\lambda_0 = 1$, $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 0.5$ and $\lambda_4 = 1$.

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

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

  with $V_0 = 9$.

3. Sampling H^{ϕ} and H^{Σ}:

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

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

  We choose $a_{ii}^{\Sigma} = 20$, $a_{ij}^{\Sigma} = 1$, $a_{ii}^{\phi} = 80$, and $a_{ij}^{\phi} = 1$. Note that the prior on H^{ϕ} is very tight and it implies that the VAR coefficients regimes are very persistent.

- 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}.

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 11 as:

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

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 heteroschedasticity.

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

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 heteroschedasticity: 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:

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

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, $\widetilde{B} = vec(\beta)'$, $\widetilde{X} = [bdiag(x_1'), ..., bdiag(x_T')]$ 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:

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

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

An alternative way to obtain a draw from the posterior of the VAR coefficients is based on the Kalman filter. 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 an high volatility regime is in place. For a detailed description of the Kalman filter see Kim and Nelson (1999).

C Computing the news in the perfect knowledge case

This appendix describes how to compute the news when agents form expectations according to a MS model. The underlying assumption is that they have *perfect knowledge* of the model.

Remember that unexpected returns can be decomposed according to the following formula:

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

Our goal is to obtain:

$$N_{DR,t+1} = (E_{t+1} - E_t)\sum_{j=0}^{\infty}\rho^j r_{t+1+j}$$

Let's start considering:

$$E_{t+1}\sum_{j=0}^{\infty}\rho^j r_{t+1+j} \tag{18}$$

The first term of this summation is simply:

$$E_{t+1}r_{t+1} = e_1\left(\left(\Phi_1 p_{\Phi_1,t+1} + \Phi_2 p_{\Phi_2,t+1}\right)[y_t', 1]' + u_{t+1}\right) = r_{t+1}$$

where $p_{\Phi_1,t+1}$ and $p_{\Phi_2,t+1}$ are the probabilities for the two regimes at time $t+1$. To get forecasts of the remaining addends, agents update these probabilities according to the transition matrix H^{Φ}. Therefore the remaining terms can be expressed as:

$$\begin{aligned} E_{t+1}r_{t+i} &= e_1\left(\widehat{\Phi}_{t+2}\widehat{\Phi}_{t+3}...\widehat{\Phi}_{t+i}\right)[y_{t+1}', 1]' \\ &= e_1\prod_{j=2}^{i}\widehat{\Phi}_{t+i}[y_{t+1}', 1]' \end{aligned}$$

where

$$\begin{aligned} \widehat{\Phi}_{t+i} &= \Phi_1\widehat{p}_{\Phi_1,t+i} + \Phi_2\widehat{p}_{\Phi_2,t+i} \\ \begin{bmatrix} \widehat{p}_{\Phi_1,t+i} \\ \widehat{p}_{\Phi_2,t+i} \end{bmatrix} &= \left(H^{\Phi}\right)^{i-1}\begin{bmatrix} p_{\Phi_1,t+1} \\ p_{\Phi_2,t+1} \end{bmatrix} \\ i &= 2, ...I \end{aligned}$$

where I is arbitrarily large. Later on it will become why it is possible to disregard the terms dated $I+1,\ I+2, ...$

Now consider:

$$E_t\sum_{j=0}^{\infty}\rho^j r_{t+1+j} \tag{19}$$

The terms of this summation can be computed according to the following formula:

$$\begin{aligned} E_t r_{t+i} &= e_1 \left(\widetilde{\Phi}_{t+1} ... \widetilde{\Phi}_{t+i} \right) [y_t', 1]' \\ &= e_1 \prod_{j=1}^{i} \widetilde{\Phi}_{t+i} y_{t+1} \end{aligned}$$

where

$$\begin{aligned} \widetilde{\Phi}_{t+i} &= \Phi_1 \widetilde{p}_{\Phi_1,t+i} + \Phi_2 \widetilde{p}_{\Phi_2,t+i} \\ \begin{bmatrix} \widetilde{p}_{\Phi_1,t+i} \\ \widetilde{p}_{\Phi_2,t+i} \end{bmatrix} &= \left(H^{\Phi} \right)^i \begin{bmatrix} p_{\Phi_1,t} \\ p_{\Phi_2,t} \end{bmatrix} \\ i &= 1, ...I \end{aligned}$$

There a two differences between the summations (18) and (19). The first one is trivial: in one case the first term of the summation is observed, while in the other a forecast must be used. The second difference is more subtle, but also more important: the probabilities assigned to the two regimes are different, especially for the first terms of the summations. However, as i becomes larger and larger, both $\widetilde{p}_{t+i}$ and $\widehat{p}_{t+i}$ converge to their stationary distributions. Therefore all the terms after I, with I large enough, are the same in the two summations.

To obtain $N_{CF,t+1}$ use:

$$N_{CF,t+1} = u_{t+1} + N_{DR,t+1}$$

Note that:

$$u_{t+1} = y_{t+1} - \left(\Phi_1 p_{\Phi_1,t+1} + \Phi_2 p_{\Phi_2,t+1} \right) y_t \neq r_{t+1} - E_t r_{t+1}$$

because the expectation error have now two components: one is due to a revision in the probabilities assigned to the two regimes, the other reflects the incoming news.

The *sophisticated specification* is obtained replacing the starting probabilities with the time-invariant weights assigned to the two regimes. Finally in the *naive case* there is no need to go through this computationally intensive exercise because the formulas of section 2.2 apply once Γ has been defined as:

$$\Gamma = w_{\Gamma_1}\Gamma_1 + (1 - w_{\Gamma_1})\,\Gamma_2$$

D Additional graphs

This appendix reports additional graphs.

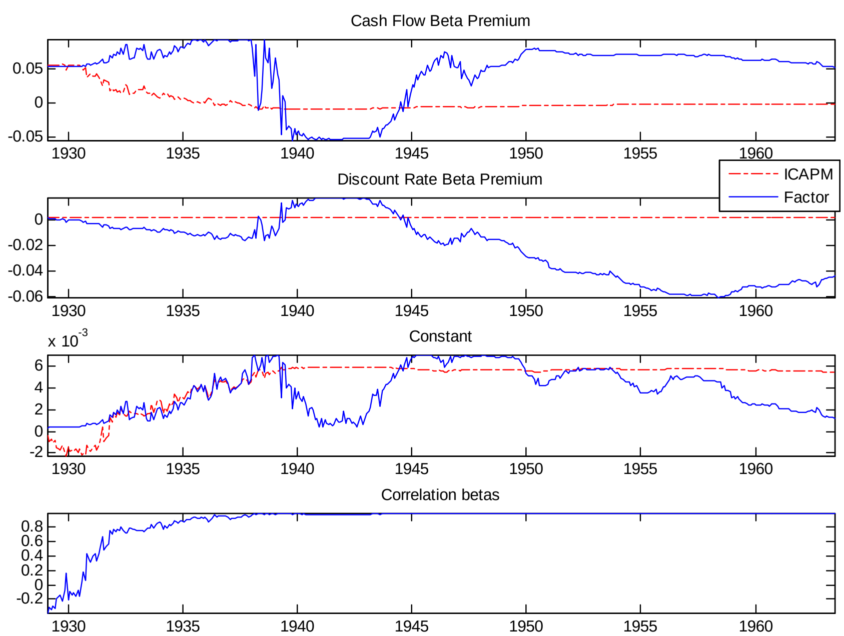

Figure 12: Fixed coeefficient VAR. Evolution of the premia as the sample size shortens. For example: 1940:12 means that the VAR has been estimated over the sample 1940:12-2001:12. The last plot describes the evolution of the correlation between the betas.