ARE VECTOR AUTOREGRESSIONS AN ACCURATE MODEL FOR DYNAMIC ASSET ALLOCATION?

Francisco Peñaranda

CEMFI Working Paper No. 0419

November 2004

CEMFI
Casado del Alisal 5; 28014 Madrid
Tel. (34) 914 290 551. Fax (34) 914 291 056
Internet: www.cemfi.es

We thank Urooj Amjad, Manuel Arellano, Lars P. Hansen, Víctor López, Gabriel Pérez-Quirós, Gonzalo Rubio, Simón Sosvilla, Luis Viceira, and specially Enrique Sentana, for very useful comments and suggestions. We also thank seminar participants at Alicante, Carlos III, CEMFI, ESEM (Venice 2002), FEDEA, LSE, and Pompeu Fabra. We are responsible for any remaining errors.

CEMFI Working Paper 0419
November 2004

ARE VECTOR AUTOREGRESSIONS AN ACCURATE MODEL FOR DYNAMIC ASSET ALLOCATION?

Abstract

Much of the growing literature on tactical and strategic asset allocation uses vector autoregressive models (VAR) for returns and predictors. Since the portfolio advice they generate may be misleading if those models are not an accurate description of reality, we evaluate the implied joint density forecasts of US monthly excess returns on stocks and bonds. From the point of view of an investor who rebalances monthly, a VAR offers a reasonable description of the data, which is not improved upon by richer models. We also study the relevance of considering time-varying risk premia and parameter uncertainty in density forecasts.

JEL Codes: G11, C53.
Keywords: Density forecasts, Parameter uncertainty, Portfolio choice, Probability integral transform, Risk premia.

Francisco Peñaranda
University of Alicante and FMG (LSE)
penaranda@merlin.fae.ua.es

1 Introduction

Much of the growing literature on tactical and strategic asset allocation[1] is based on a Gaussian vector autoregressive (VAR) model for returns and predictors, which provides a simple way of modelling time-varying investment opportunities in terms of changing interest rates and risk premia. Optimal portfolios are then computed for a chosen utility function once the estimation of the VAR parameters has taken place. Campbell et al. (2003) and Lynch (2001) are recent, well-known examples of this approach. These authors conduct classical estimation procedures, and take the resulting parameter estimates as if they were equal to the true parameters. In contrast, other authors explicitly take parameter uncertainty into account by means of Bayesian procedures. Specifically, Kandel and Stambaugh (1996), Stambaugh (1999) and Barberis (2000) study portfolio selection based on Bayesian VAR. This second approach will be the benchmark in this paper.

Unfortunately, the portfolio advice these methods generate may be misleading if the assumed Gaussian VAR does not provide an accurate description of the distribution of returns. The computation of an optimal portfolio requires the whole distribution, not only first and second moments of returns, unless some restrictions are imposed on investors' preferences. In addition, a joint distribution of several assets is required for portfolio advice. For that reason, in this paper we evaluate the joint density forecasts of US monthly excess returns on stocks and bonds that such models generate for the post-war period. The monthly frequency is chosen as the link between the usual high and low rebalancing frequencies of tactical and strategic asset allocations respectively.

To the best of our knowledge, the existing literature has not evaluated the accuracy of density forecasts. Instead, the modelling efforts have been made in two different directions. One is the nonparametric approach advocated in Brandt (1999) and Aït-Sahalia and Brandt (2001), who derive the optimal portfolio choice without a statistical model. The other direction is to explicitly deal with model uncertainty by means of Bayesian model averaging. Such an approach has been followed for instance by Avramov (2002) and Cremers (2002), who limit the model uncertainty to the choice of predictors by taking as given a Gaussian linear model.

This paper tries to complement both these approaches by confronting the density

[1] See for instance Brennan et al. (1997) for their definition and comparison.

forecast properties of a particular model with the real data. Specifically, our evaluation procedure involves out-of-sample checks of real-time density forecasting rules against realizations of returns.[2] The main tool is the probability integral transform (PIT) advocated in Diebold et al. (1998), which is defined as the cumulative distribution function of the density forecast evaluated at the actual realization. If the forecasting model is correct, the PIT should be uniformly and independently distributed over time.

Since investment opportunities might change in more complex ways than the normal VAR allows for, it is not surprising that richer models are also considered in the literature. For instance, Markov-switching VARs have been used to capture the asymmetric behaviour of returns in different market phases (see e.g. Pérez-Quirós and Timmermann (2000, 2001) and Chauvet and Potter (2001), as well as Ang and Bekaert (2002), who study its portfolio implications in an international context). In this respect, we find that an investor who rebalances her portfolio monthly can rely on the standard VAR model, which seems to provide a reasonable description of the data. We also find that richer models do not improve the quality of the density forecasts, and they are more difficult to implement.

Finally, we address two important issues for asset allocation: time-variation in risk premia and parameter uncertainty. Both questions were also studied by Barberis (2000), but only in terms of their portfolio choice implications. In contrast, our paper focuses directly on their density forecast implications, which seem to be robust across different models.

Specifically, we estimate two versions of each model, with constant and time-varying risk premia. It is clear that models with or without time variation in risk premia imply very different density forecasts and this translates into significant effects in portfolio choice as it is well known in the literature. However, the PIT shows that the quality of both forecasts cannot be ranked in general. With respect to constant risk premia, time-varying premia worsen the forecast evaluation for stocks but improve for bonds. Nevertheless, the PIT shows a common feature in stocks and bonds: constant risk premia tend to produce over-pessimistic forecasts.

[2] Evaluation of density forecasts is a very active area nowadays and it is not only useful for asset allocation, but also in other contexts. This is clear in modern models of financial risk such as the Value-at-Risk, where quantiles are computed.

We also consider two forecasting rules for each estimated model: the Bayesian predictive distribution and a plug-in rule. The latter one is easier to compute than the former, but it may not properly account for parameter uncertainty. However, we show that the fully Bayesian approach and the plug-in forecasting rule yield similar monthly density forecasts so that an investor could avoid the complexity of the former.

The rest of the paper is organised as follows. Section 2 contains a description of the data. Section 3 explains the computation of density forecasts in a Bayesian framework, as well as their evaluation. Section 4 defines and applies the standard normal model, while section 5 is devoted to nonlinear dynamics, and specifically Markov-switching models. Finally, section 7 concludes. Details about simulation-based methods and each model implementation are gathered in the appendix.

2 Data Description: Excess Returns and Predictors

US monthly nominal data from January 1954 to December 2001 are taken from CRSP, giving 576 observations. Monthly data is the usual frequency in the previously referred literature on asset allocation. The post-war sample period is used to conform with the period after the Fed-Treasury accord and the presidential election in 1952 after which the Fed stopped pegging interest rates. This fundamentally changed the process of the nominal interest rates. The chosen period for the out-of-sample check of the density forecasts starts with the forecast of January 1975, which gives 324 observations for that check. This period represents enough observations, 252, without letting too few observations for the first estimation. During this period, there was a sharp change in the monetary policy in the U.S. Specifically, Volcker arrived at the Fed in October 1979 and followed a tighter monetary policy by means of monetary aggregate controls that increased the volatility of interest rates. Therefore, it is an interesting period for forecast evaluation.

Traditional assets can be roughly divided into three classes: stocks, bonds, and cash. This paper focuses on excess returns from a stock index and a bond index to keep the analysis simple. In the computation of excess stock returns (ESR), returns are based on the value weighted stock index of the NYSE, NASDAQ and AMEX markets including dividends. The returns are converted into continuously compounded rates in monthly percentage. The other ingredient is the 1-month T-bill return from the US Treasury and Inflation Series (CTI). Again, the series is converted into continuously compounded rates in monthly percentage. Finally, the excess stock return e_t^S is given by the difference between both returns.

A huge fall can be observed in October 1987 in Panel A of Figure 1 as well as a less important fall in August 1998 due to the Russian financial crisis. Apart form that, no sharp changes exist as in the series of bonds. In Table 1, Panel A, there are some descriptive statistics of the stock and bond excess returns for the whole data period and for the two periods given by the out-of-sample check.

(Table 1)

The period from January 1954 to December 1974 corresponds to the first estimation period while the period form January 1975 to December 2001 corresponds to the out-of-sample check. First, the statistics for the whole period are explained. The mean is 0.49%, while the standard deviation is 4.35%, which gives an annualized Sharpe ratio of about 0.4. The coefficient of skewness is -0.77 and the one of kurtosis 2.93, that is, there is asymmetry to the left and fat tails. The Jarque-Bera test rejects the hypothesis of normality with a p-value of 0.00. To approximate the temporal dependence in the series, the autocorrelograms of powers of the demeaned series are shown in Panels B and C of Figure 1. The level does not show autocorrelation, while it can be found only marginally in the second power for the first lag (a p-value of 0.04) even in terms of the Ljung-Box test, see Panel B in Table 1. Regardless, the correlation in the square will be much clearer for bonds.

The period that is left for the out-of-sample check, January 1975 to December 2001, shows a much higher mean and kurtosis than the previous period, while the standard deviation is only slightly higher. The negative skewness also increases. Noteworthy, there is not autocorrelation in the first and second powers during the last period, while the correlation of the square is clear in the first period.

Excess bond returns (EBR) are computed from returns of the 5 year bond from the US Treasury and Inflation Series (CTI), which are converted into continuously compounded rates in monthly percentage. Finally, the excess bond return e_t^B is given by the difference between that return and the continuously compounded 1-month T-bill return.

The time series, Panel A in Figure 2, shows a huge increase in the volatility at the beginning of the 80's due to the commented Fed experiment. An increase in volatility during the first half of the 70's exists due to the OPEC oil crises, but is much less sharp. The mean is 0.10%, while the standard deviation is 1.49%, see Panel A in Table 1, which gives an annualized Sharpe ratio of about 0.2, half the ratio of stocks. The skewness is 0.07 and the kurtosis 4.12, there is asymmetry of the opposite sign to stocks and even fatter tails. The Jarque-Bera test rejects the hypothesis of normality with a p-value of 0.000. As it can be seen in Figure 2, Panels B and C, the level of the series shows autocorrelation at the first lag, the p-value is 0.00. There is a clear dependence in the second power with zero p-values at any lag in terms of the Ljung-Box test, see Panel B in Table 1.

In the out-of-sample period, the mean is much higher while the skewness is no longer positive. The standard deviation and the kurtosis also increase, specially the former. As opposed to the case of stocks, bonds show a higher autocorrelation in the first and second powers during the second period, especially in the level.

Let us turn to predictors now. In our context, they are state variables for risk premia. These variables are in general related to the business cycle and are very persistent. The most popular predictors can be classified into two categories. On the one hand, yield variables such as the dividend price ratio and the price earning ratio are common predictors of stock returns. The dividend yield is one of the predictors that are used in this paper, since it is the most popular for stocks. The dividend yield is studied, for instance, in Fama and French (1988, 1989) and Campbell and Shiller (1988), and its predictive power comes from mean reversion in stock returns. On the other hand, interest rates such as the term premium, the short term rate and the default premium are used as predictors of stock returns, and of bond returns too. The first variable measures the slope in the term structure of interest rates and is used in this paper. The term premium is analysed, for instance, in Fama and French (1989) and Campbell and Shiller (1991).

The dividend yield is derived from the previous value weighted index of stocks without dividends and the corresponding return with dividends, summing up previous year dividends. The dividend yield is often computed using the sum of last year dividends to avoid seasonality. Finally, the logarithm is taken. It can be seen in Panel A of Figure 3 that it is a highly persistent variable.

The term premium is defined as the difference between the continuously compounded monthly yield of a 5 year zero coupon bond, taken form the Fama-Bliss discount bond yields, and the yield of the 1-month risk free asset, obtained from the risk-free rates file. Note that it is computed as a yield spread, not as an excess return. This series, see Panel B in Figure 3, is also persistent but not as much as the divided yield.

These four series, two excess returns and two predictors, will be stacked together in a vector,

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{e}_t \\ \mathbf{z}_t \end{pmatrix}, \quad t = 1, ..., T,$$

where $\mathbf{e}_t$ and $\mathbf{z}_t$ are 2×1 vectors containing the excess returns and the predictors. We will study several models for the dynamics of that vector, with the Bayesian VAR as the baseline model.

Such data are similar to Campbell et al. (2003) but at a different frequency and using less variables. Their chosen frequencies are quarterly and annual since they focus on strategic asset allocation, but the usual frequency in other related papers is monthly data. They use two additional variables, since they model time-variation in both short-term interest rates and risk premia. They add the real short-term rate as a return and the nominal short-term rate as a predictor. Given that the paper estimates and evaluates more complex models, those variables are not included and we focus on risk premia. Brennan et al. (1997) also use stocks, bonds, and bills as their investment set. In a similar spirit, they use the dividend price ratio, and short and long-term interest rates as the predictors. Lynch (2001) uses the dividend price ratio and the long-short yield spread as predictors, but his returns correspond to stock portfolios sorted by size and book-to-market ratios. On the other hand, Barberis (2000) uses only the stock excess return and the dividend yield.

3 Computation and Evaluation of Joint Density Forecasts

This section develops the basic tools of our work. Let us introduce some notation first. $p(\cdot)$ will represent density or mass depending on the associated random variable and $P(\cdot)$ will be the corresponding cumulative distribution function (CDF). $E_{p(\cdot)}[\cdot]$ means expectation with respect to $p(\cdot)$. Regime-switching VARs introduce a latent variable that will be denoted by s_t. In addition, $\mathbf{x}^T = \{\mathbf{x}_t\}_{t=1}^T$ represents a time series with T

observations and $\boldsymbol{\theta}$ is a vector of parameters.

In the following, we have in mind an investor that rebalances her portfolio at the same frequency as our data. The monthly frequency is chosen as the link between the usual high and low rebalancing frequencies of tactical and strategic asset allocations respectively. An optimal portfolio is usually obtained through dynamic programming and the corresponding Bellman equation requires the density forecast $p\left(\mathbf{y}_{T+1} \mid \mathbf{y}^{T}\right)$ to compute the expected continuation value of any portfolio choice. Each model defines a different distribution for $\mathbf{y}_{T+1}$ and therefore a possibly different portfolio advice. That portfolio advice may be misleading if the estimated $p\left(\mathbf{y}_{T+1} \mid \mathbf{y}^{T}\right)$ is not an accurate description of returns and predictors. For that reason, this paper evaluates if a particular model gives a an accurate distribution when faced against data.

Tactical asset allocation computes an optimal portfolio for a short-term investor, while strategic asset allocation does it for a long-term investor. Both portfolios might be different and their difference is called hedging demand. The main role of that difference is hedging against changing investment opportunities, which is irrelevant for a short-term investor. The hedging demand is driven by the dependence structure between returns and predictors and that is why the whole $p\left(\mathbf{y}_{T+1} \mid \mathbf{y}^{T}\right)$ is necessary for strategic asset allocation. On the other hand, $p\left(\mathbf{e}_{T+1} \mid \mathbf{y}^{T}\right)$ is enough for tactical asset allocation.

Empirical results such as the ones in Ang and Bekaert (2002) and Brandt (1999) show that the hedging demand component of a long-term portfolio choice is not very important in practice. Therefore, we can focus on the evaluation of returns density forecasts $p\left(\mathbf{e}_{T+1} \mid \mathbf{y}^{T}\right)$ to simplify the analysis.

3.1 Computation of Joint Density Forecasts

We will compute density forecasts $p\left(\mathbf{y}_{T+1} \mid \mathbf{y}^{T}\right)$ at a given point in time T from statistical models that define $p\left(\mathbf{y}_{T+1} \mid \mathbf{y}^{T}, \boldsymbol{\theta}, s_{T+1}\right)$ and dynamics of the unobservable variable s_{T+1}. This latent variable does not show up in the Gaussian VAR but we will describe a general framework that embeds all the relevant models. The joint distribution of the first two entries of $\mathbf{y}_{T+1}$, $p\left(\mathbf{e}_{T+1} \mid \mathbf{y}^{T}\right)$, is the object of interest.

A Bayesian model is constructed with a prior and a likelihood, which jointly define a posterior for the parameters as[3]

$$p\left(\boldsymbol{\theta} \mid \mathbf{y}^T\right) \propto p(\boldsymbol{\theta})\, p\left(\mathbf{y}^T \middle| \boldsymbol{\theta}\right).$$

The support of the posterior will be denoted by Θ. We use proper priors since they are necessary with regime-switching models, but those priors will not be too informative. In addition, the priors of some parameters will be centred at their maximum likelihood estimators. This type of priors is known as sample priors and has been used by Kandel and Stambaugh (1996) and Avramov (2002) for instance.

We rely on simulation-based methods to get a simulated sample $\{\boldsymbol{\theta}_i\}_{i=1}^{I}$ that approximates the posterior distribution. The basis of the estimation of each model will be a Gibbs sampler jointly with data augmentation. The Gibbs sampler is very popular among the Markov Chain Monte Carlo (MCMC) methods[4] and will be used here. If we can construct a block structure for the parameters such that we can draw from the posterior of each block conditional on the rest of the blocks, then this structure defines a Markov chain that converges to the desired posterior. Data augmentation is a useful tool when we cannot sample from the posterior of $\boldsymbol{\theta}$, but we know how to do it from the posterior of $\left(\boldsymbol{\theta}, s^T\right)$ by adding latent variables s^T to the sampler. In addition, we can usually implement it in terms of a Gibbs sampler. One block would be given by the parameters and the other by the latent variables. The practical implementation of these methods is developed in the appendix.

Once we have the parameters' posterior, we can turn to the computation of density forecasts. We implement two ways of linking both objects. Both forecasting rules start from a Bayesian estimation of the model, but only one is fully Bayesian when computing the density forecast.

In some models, we will find latent variables that should be integrated out to work with Gaussian distributions. Specifically, these latent variables are discrete and take only two values, $s_t = 1, 2$. The Gibbs sampler approximates $p\left(\boldsymbol{\theta}, s^T \mid \mathbf{y}^T\right)$ and we can use its output to approximate

$$p^I\left(\mathbf{y}_{T+1} \middle| \mathbf{y}^T\right) = E_{p(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T)}\left[\sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s_T\right) p\left(\mathbf{y}_{T+1} \middle| \mathbf{y}^T, \boldsymbol{\theta}, s_{T+1}\right)\right], \quad (1)$$

[3]The estimation of every model is going to be conditional on an initial observation $\mathbf{y}_0$ which is omitted from the expressions for simplicity.

[4]See Chib and Greenberg (1996) and the references therein.

where the parameters $\boldsymbol{\theta}$ and the latent variables are integrated out. This will be named integrated density forecast and shows how to take into account estimation risk by means of the posterior. We have used that $p\left(\mathbf{y}_{T+1}|\, \mathbf{y}^T, \boldsymbol{\theta}, s^{T+1}\right) = p\left(\mathbf{y}_{T+1}|\, \mathbf{y}^T, \boldsymbol{\theta}, s_{T+1}\right)$ and $p\left(s_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s^T\right) = p\left(s_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s_T\right)$ in the models that will be estimated. The appendix shows how to approximate (1).

The second forecasting rule is much easier to compute. We only use the posterior mean of the parameters

$$\bar{\boldsymbol{\theta}} = E_{p\left(\boldsymbol{\theta}|\mathbf{y}^T\right)}\left[\boldsymbol{\theta}\right], \quad \bar{s}_T = E_{p\left(s_T|\mathbf{y}^T\right)}\left[s_T\right],$$

not the whole posterior distribution, and simply plug-in this estimators to get a density forecast

$$p^P\left(\mathbf{y}_{T+1}|\, \mathbf{y}^T\right) = \sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \bar{\boldsymbol{\theta}}, \bar{s}_T\right) p\left(\mathbf{y}_{T+1}|\, \mathbf{y}^T, \boldsymbol{\theta}, s_{T+1}\right).$$

This will be named plug-in density forecast. The performance of both approaches will be compared to study if the complexity of the first approach improves the accuracy of density forecasts.

It is trivial to compute $p^I\left(\mathbf{e}_{T+1}|\, \mathbf{y}^T\right)$ and $p^P\left(\mathbf{e}_{T+1}|\, \mathbf{y}^T\right)$ from $p^I\left(\mathbf{y}_{T+1}|\, \mathbf{y}^T\right)$ and $p^P\left(\mathbf{y}_{T+1}|\, \mathbf{y}^T\right)$ respectively. We will compute the mean, standard deviation, and the coefficients of skewness and kurtosis of density forecasts to get an idea of their shape. They require the first four central moments, which can be expressed in terms of noncentral moments. If the particular random variable is normally distributed then the noncentral moments have well known formulas. The density forecasts are not Gaussian, but we can still use those formulas due to the data augmentation approach in the Bayesian estimation. The noncentral moments of a density forecast can be easily expressed as

$$E_{p^I\left(\mathbf{y}_{T+1}|\mathbf{y}^T\right)}\left[\left(e^S_{T+1}\right)^k\right] = \tag{2}$$
$$E_{p(\boldsymbol{\theta},s_T|\mathbf{y}^T)}\left[\sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s_T\right) E_{p\left(\mathbf{y}_{T+1}|\mathbf{y}^T,\boldsymbol{\theta},s_{T+1}\right)}\left[\left(e^S_{T+1}\right)^k\right]\right],$$

for the case of stocks and for a particular power k, which can be approximated with the MCMC output. The Gibbs output gives draws from $p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T\right)$ so that we only have to average the corresponding computations. $p\left(\mathbf{y}_{T+1}|\, \mathbf{y}^T, \boldsymbol{\theta}, s_{T+1}\right)$ is normal in every model and therefore the corresponding moment is easy to compute, while $p\left(s_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s_T\right)$ is discrete and therefore easy to integrate. The counterparts for the plug-in rule do not integrate with respect to $p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T\right)$ but evaluate at the corresponding posterior means

$$E_{p^P\left(\mathbf{y}_{T+1}|\mathbf{y}^T\right)}\left[\left(e_{T+1}^S\right)^k\right] = \sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \bar{\boldsymbol{\theta}}, \bar{s}_T\right) E_{p\left(\mathbf{y}_{T+1}|\mathbf{y}^T, \bar{\boldsymbol{\theta}}, s_{T+1}\right)}\left[\left(e_{T+1}^S\right)^k\right]. \quad (3)$$

3.2 Evaluation of Joint Density Forecasts

We turn to the evaluation of $p^P\left(\mathbf{e}_{T+1} \mid \mathbf{y}^T\right)$ and $p^I\left(\mathbf{e}_{T+1} \mid \mathbf{y}^T\right)$ after showing their computation. In that respect, we will follow the approach in Diebold et al. (1998), jointly with formal tests such as the ones in Wallis (2003). The evaluation is based on the probability integral transform (PIT), which is defined as the density forecast's CDF evaluated at the final realization. Let us think for a moment in a particular scalar random variable x with conditional density $p\left(x_{t+1} \mid x^t\right)$, then the PIT is

$$d_{t+1} = P\left(x_{t+1} \mid x^t\right),$$

and this transformation is uniform and independent over time

$$d_{t+1} \overset{iid}{\sim} U\left(0, 1\right),$$

assuming a nonsingular Jacobian and continuous partial derivatives in this transformation. This property was developed by Rosenblatt (1952) and it has a clear application in density forecast evaluation as Diebold et al. (1998) show.[5] If our density forecast is an accurate description of the true conditional density then its corresponding PIT should be uniformly distributed and independent over time. That is, there should not be a tendency to more or less likely realizations of returns, nor patterns along time of this likeliness.[6]

The PIT is very convenient since we can summarize the performance of a density forecast in only one number. However, our evaluation of interest is a joint distribution and there is not a similar property for the PIT of a joint distribution. Fortunately, Diebold et al. (1998) show how to extend their approach from univariate to multivariate

[5] This application of the PIT was already commented in Dawid (1984) in his "prequential approach" to forecasting and it was applied in Shephard (1994) and Kim et al. (1998) too.

[6] Care must be taken when going from conclusions about PIT's marginal and dynamic behaviour to conclusions about the corresponding return's marginal and dynamic behaviour since they are not linked in general. Nevertheless, failures in PIT's uniformity and independence may suggest directions of model improvement that can be checked once implemented.

density forecasts by means of studying the PITs of marginal and conditional univariate distributions.[7]

The implementation of this method is as follows. The sample period $T = 1, ..., T^{**}$ is split in two parts. The first part, $T = 1, ..., T^{*}$, is used for a first estimation and the second one, $T = T^{*} + 1, ..., T^{**}$, is added observation by observation to the first part to estimate recursively the model and construct one-period ahead density forecasts $\left\{ p^{P}\left(\mathbf{e}_{T+1} | \, \mathbf{y}^{T} \right) \right\}_{T^{*}}^{T^{**}-1}$ and $\left\{ p^{I}\left(\mathbf{e}_{T+1} | \, \mathbf{y}^{T} \right) \right\}_{T^{*}}^{T^{**}-1}$. We evaluate the corresponding CDFs at the realizations of returns for that period $\left\{ \mathbf{e}_{T+1} \right\}_{T^{*}}^{T^{**}-1}$ to compute the PITs. Specifically, T^{*} is December 1974 and therefore the out-of-sample check starts with the forecast of January 1975. There are 324 observations of the previous objects while 252 data observations are left for the first estimation in the recursive procedure.

Bayesian inference implies an updating of density forecasts after each new observation. Therefore the posterior used in each density forecast is not frozen at the first period estimation and we use a rolling approach instead. However, we will only update the parameters' posterior quarterly instead of monthly since the output is similar and much computing time is saved.

The next stage is to compare both series to evaluate the forecasts. There are two excess returns to predict and therefore four PITs to compute. Recall $\mathbf{y}_{T+1} = \left(\mathbf{e}_{T+1}', \mathbf{z}_{T+1}' \right)'$ and $\mathbf{e}_{T+1} = \left(e_{T+1}^{S}, e_{T+1}^{B} \right)'$ so that we can compute

$$
\begin{aligned}
d_{T+1}^{f,S} &= P^{f}\left(e_{T+1}^{S} \mid \mathbf{y}^{T} \right), \quad f = I, P, \\
d_{T+1}^{f,B|S} &= P^{f}\left(e_{T+1}^{B} \mid \mathbf{y}^{T}, e_{T+1}^{S} \right),
\end{aligned}
$$

and similarly $d_{T+1}^{f,B}$ and $d_{T+1}^{f,S|B}$.

Two ways of computing density forecasts will be studied. The fully Bayesian approach was called the integrated rule and its corresponding PITs can be computed as

$$
d_{T+1}^{I,S} = E_{p(\boldsymbol{\theta}, s_T | \mathbf{y}^{T})} \left[\sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^{T}, \boldsymbol{\theta}, s_T \right) P\left(e_{T+1}^{S} \mid \mathbf{y}^{T}, \boldsymbol{\theta}, s_{T+1} \right) \right], \tag{4}
$$

$$
d_{T+1}^{I,B|S} = E_{p\left(\boldsymbol{\theta}, s_T | \mathbf{y}^{T}, e_{T+1}^{S}\right)} \left[\sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^{T}, \boldsymbol{\theta}, s_T \right) P\left(e_{T+1}^{S} \mid \mathbf{y}^{T}, e_{T+1}^{S}, \boldsymbol{\theta}, s_{T+1} \right) \right],
$$

[7]See for instance Diebold, Hahn, and Tay (1999) and Clements and Smith (2000) for an empirical application of this method.

where $P\left(e^S_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s_{T+1}\right)$ and $P\left(e^S_{T+1} \mid \mathbf{y}^T, e^S_{T+1}, \boldsymbol{\theta}, s_{T+1}\right)$ are normal CDFs and therefore easy to compute. The appendix shows how to approximate those expectations using the Gibbs output.

On the other hand, the PITs of the plug-in rule are easily computed as

$$
\begin{aligned}
d^{P,S}_{T+1} &= \sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \bar{\boldsymbol{\theta}}, \bar{s}_T\right) P\left(e^S_{T+1} \mid \mathbf{y}^T, \bar{\boldsymbol{\theta}}, s_{T+1}\right), \\
d^{P,B|S}_{T+1} &= \sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \bar{\boldsymbol{\theta}}, \bar{s}_T\right) P\left(e^S_{T+1} \mid \mathbf{y}^T, e^S_{T+1}, \bar{\boldsymbol{\theta}}, s_{T+1}\right).
\end{aligned} \tag{5}
$$

The PIT's properties imply that if density forecasts are accurate then there are 6 time series should be independent and identically uniform for each forecasting rule,

$$
\begin{aligned}
&d^{f,S}_{t+1}, \quad d^{f,B}_{t+1}, \quad d^{f,S|B}_{t+1}, \quad d^{f,B|S}_{t+1}, \\
&\left(d^{f,S}_{t+1}, d^{f,B|S}_{t+1}\right), \quad \left(d^{f,B}_{t+1}, d^{f,S|B}_{t+1}\right) \overset{iid}{\sim} U(0,1).
\end{aligned}
$$

Therefore, if the evaluation is based on the PITs then we compute only 4 numbers for each density forecast. In our empirical application, the performance of $d^{f,S|B}_{t+1}$, $d^{f,B|S}_{t+1}$ and $\left(d^{f,B}_{t+1}, d^{f,S|B}_{t+1}\right)$ is similar to $d^{f,S}_{t+1}$, $d^{f,B}_{t+1}$ and $\left(d^{f,S}_{t+1}, d^{f,B|S}_{t+1}\right)$, respectively. Therefore, we only show tables and figures concerning $d^{f,S}_{t+1}$, $d^{f,B}_{t+1}$, and $\left(d^{f,S}_{t+1}, d^{f,B|S}_{t+1}\right)$.

There are two features to test for each time series, uniformity and independence and we will use graphs and tests. I will show the CDF and correlograms of the PITs.[8] They will be shown jointly with the 5% critical value of a Kolmogorov-Smirnov test. Unfortunately, that may not be very useful for evaluation of density forecasts. For instance, this test has low power against a bad description of the tails of the distribution since it focuses mainly on the median. Because of that, other tests such as Kupier and Anderson-Darling are computed, which have more power against bad performance in the tails. They will be named empirical distribution function (EDF) tests and the critical values of the three tests will be taken from Stephens (1974).

All of these tests share the independence assumption as a maintained hypothesis. So we should look at them after studying the independence of the corresponding time series. The independence of the PIT can be checked with correlograms of different powers, for instance $\left(d^{f,S}_{t+1} - \bar{d}^{f,S}\right)$ and $\left(d^{f,S}_{t+1} - \bar{d}^{f,S}\right)^2$, where $\bar{d}^{f,S}$ is the corresponding time series

[8] The uniformity of the PIT can be checked by means of a histogram too but histograms can be misleading as they depend on the number of bins that are used.

mean, jointly with the corresponding Barlett confidence intervals. The Ljung-Box statistic is also computed to show a joint tests of different lags.

Finally, we compute the tests proposed by Wallis (2003).[9] We apply them to PITs instead of density forecasts themselves since the latter option would require the storage of a huge number of computations. He develops three likelihood ratio tests:[10] unconditional coverage, independence and conditional coverage. The null of the first test is i.i.d. uniformity against i.i.d. sampling from any other distribution, while in the second test we face any i.i.d. distribution against a first order Markov chain with unknown distribution. The final test collapses the previous tests in a joint test and its null hypothesis is i.i.d. uniformity against a first order Markov chain with unknown distribution. We have to choose a particular number of bins to divide the support of the PIT in the implementation of these tests. In that respect, we show results for 3 and 4 bins.

4 Gaussian VAR

We mentioned that much of the growing literature on tactical and strategic asset allocation is based on Gaussian vector autoregressions (VARs) of returns and predictors. Campbell et al. (2003) and Lynch (2001) are well-known examples of this approach using classic inference. They take the corresponding parameter estimates as if they were equal to the true parameters. In contrast, other authors explicitly take parameter uncertainty into account by means of Bayesian inference. Kandel and Stambaugh (1996), Stambaugh (1999) and Barberis (2000) study portfolio selection based on Bayesian VAR. The latter approach will be the benchmark model in this paper.

A VAR is a simple way of introducing time-varying risk premia. Homoskedasticity is often assumed and therefore the model does not show time-varying investment opportunities in terms of volatility. We will study if this particular model is a good description of stock and bond excess returns.

[9] Another approach is proposed in Berkowitz (2001) but it shows very low p-values for all models in general, and therefore it does not help much to draw conclusions. This happens also in Clements and Smith (2000) and they interpret it as higher sensitivity to outliers.

[10] In the same spirit as the interval forecast tests in Christoffersen (1998).

4.1 Econometric Model

Recall that $\mathbf{y}_t$ is a 4×1 vector composed by two excess returns and two predictors. The likelihood of this model is given by a Gaussian VAR,

$$\mathbf{y}_t = \mathbf{a} + \mathbf{B}\mathbf{y}_{t-1} + \mathbf{u}_t, \tag{6}$$

$$\mathbf{u}_t \mid \left(\mathbf{y}^{t-1}, \boldsymbol{\theta}\right) \sim N\left(\mathbf{0}, \Sigma\right), \tag{7}$$

where $\mathbf{a}$ is a 4×1 vector of intercepts, $\mathbf{B}$ is a 4×4 matrix of slope coefficients and $\mathbf{u}_t$ represents the innovations, which are independent and normal. The vector $\boldsymbol{\theta}$ represents the model parameters and it will be defined later. Cross-sectional residual correlation is allowed to take into account feed-back between returns and predictors, but residual homoskedasticity and independence over time are assumed.

The main role of a VAR in the context of asset allocation is modeling time-varying risk premia. Two versions of this model will be estimated and evaluated to study the empirical relevance of that time variation: One with a time-varying conditional mean and another with a constant mean. The latter version imposes the following restrictions on the general model

$$\mathbf{B} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{ze} & \mathbf{B}_{zz} \end{pmatrix}, \tag{8}$$

where $\mathbf{B}$ has been partitioned in terms of $\mathbf{e}_t$ and $\mathbf{z}_t$.

We will use compact notation for the VAR intercepts and slopes. Specifically,

$$\Pi = \begin{pmatrix} \mathbf{a}' \\ \mathbf{B}' \end{pmatrix}, \quad \boldsymbol{\pi} = vec\left(\Pi\right),$$

where the vector $\boldsymbol{\pi}$ is 20×1 for the unrestricted version of the model and 12×1 for the restricted version. The model parameters to estimate are

$$\boldsymbol{\theta} = \left\{\boldsymbol{\pi}, \Sigma\right\},$$

while the likelihood is

$$p\left(\mathbf{y}^T \middle| \boldsymbol{\theta}\right) = \prod_{t=1}^{T} p\left(\mathbf{y}_t \middle| \mathbf{y}^{t-1}, \boldsymbol{\theta}\right),$$

where each $p\left(\mathbf{y}_t|\ \mathbf{y}^{t-1},\boldsymbol{\theta}\right)$ is normal. To close the model, the priors are

$$\boldsymbol{\pi} \sim N\left(\underline{\boldsymbol{\pi}}, \underline{\mathbf{C}}\right),$$

$$\Sigma \sim IW\left(\underline{\nu}, \underline{\mathbf{S}}\right),$$

where $IW(\cdot)$ denotes the inverse Wishart distribution.[11] These priors are informative for continuity with regime switching models where they are needed. We do not restrict the eigenvalues of $\mathbf{B}$ to the interior of the unit circle. We should restrict the VAR if we believe that our model is the true data generating process, since we expect stationary returns and predictors. But if the true data generating process is different, for instance a Markov-switching model, then it may be the case that the estimation of this simple VAR shows nonstationarity although the process is really stationary. More details about the implementation of the model such as specific prior values and corresponding Gibbs sampler can be found in the appendix.

In terms of density forecasts, we can directly rely on normal distributions. Since

$$\mathbf{y}_{T+1}|\left(\mathbf{y}^T,\boldsymbol{\theta}\right) \sim N\left(\mathbf{a}+\mathbf{B}\mathbf{y}_T, \Sigma\right),$$

we have that

$$e^S_{T+1} \mid \left(\mathbf{y}^T,\boldsymbol{\theta}\right) \sim N\left(a^S + \left(\mathbf{b}^S\right)'\mathbf{y}_T, \sigma^{SS}\right),$$

where a^S is the corresponding entry in $\mathbf{a}$, $\left(\mathbf{b}^S\right)'$ is the corresponding row in $\mathbf{B}$, and σ^{SS} is the corresponding entry in Σ. In addition,

$$e^B_{T+1} \mid \left(\mathbf{y}^T, e^S_{T+1}, \boldsymbol{\theta}\right) \sim N\left(a^B + \left(\mathbf{b}^B\right)'\mathbf{y}_T + \frac{\sigma^{SB}}{\sigma^{SS}}u^S_{T+1}, \sigma^{BB} - \frac{\left(\sigma^{SB}\right)^2}{\sigma^{SS}}\right),$$

where the notation is similar to the previous expression. The distributions of $e^B_{T+1} \mid \left(\mathbf{y}^T,\boldsymbol{\theta}\right)$ and $e^S_{T+1} \mid \left(\mathbf{y}^T, e^B_{T+1},\boldsymbol{\theta}\right)$ are defined in a similar way. The required moments can be computed following (2) and (3), while the PITs are described in (4) and (5).

4.2 Estimation and Evaluation

We show the empirical results of the model in two dimensions. First, the estimation is shown for the first and last estimation periods to interpret the model and compare it

[11] The parametrization is such that if $\mathbf{X} \sim IW(\nu, \mathbf{S})$, where $\mathbf{X}$ is a $d \times d$ positive definite matrix, then $E(\mathbf{X}) = \mathbf{S}^{-1}/(\nu - d - 1)$. This is equivalent to $\mathbf{X}^{-1} \sim W(\nu, \mathbf{S})$ with $E\left(\mathbf{X}^{-1}\right) = \nu\mathbf{S}$.

to existing results. Second, the evaluation of density forecasts is shown for a recursive estimation starting at the forecast of January 1975.

4.2.1 Real-Time Model Estimation

Table 2 shows part of the estimation of the two model versions. Specifically, tables and corresponding comments are focused on the posterior mean of some parameters. The estimation is shown for the first and last estimation periods, defined by data up to December 1974 and November 2001.

(Table 2)

The following comments also explain additional results that are not shown in tables. For instance, comments on significance of parameters are based on the comparison of posterior means and standard deviations, although the latter are not displayed in the tables. The posterior distributions of conditional mean parameters are very close to a normal distribution in general.

In the case of stocks, both predictors contribute with a positive coefficient to the conditional mean, as it is found in the literature. But the dividend yield cannot be regarded as significant in the last estimation. It is well known that the dividend yield has lost part of its predictive power during the second half of the 90's. The lagged bond return, which is not shown in the table, has a significant positive coefficient in both estimations. The constant mean version shows a positive mean that is only significant in the last estimation. If we turn to bonds, the dividend yield coefficient changes sign. But the last value cannot be regarded as significant and therefore it has lost predictive power for bonds too. On the other hand, the term premium has a positive coefficient in both estimations, as it is usual in the literature, but only the last estimation is significant. Both lagged excess returns, which are not shown in the table, have a significant coefficient in the last estimation, negative for the stock and positive for the bond. The intercept in the constant mean version is not significant in both estimations. In fact, it is negative in some periods. A well-known feature is that predictors are almost single persistent AR(1)s. This is also found here. During the last years, the estimated value of the autoregressive coefficient of the dividend yield has increased and got closer to 1.

In terms of residual variances of excess returns, there is an increase from one period to the other in the covariance, which is positive, and both variances. The covariance suffers a huge increase in the first estimation years and then becomes more stable. The bond variance increases sharply in the beginning and then falls smoothly. The correlation, computed at the posterior mean values of covariances and variances, has increased from 0.04 to 0.15 in the predictable mean version. The residual covariance between the ESR and the dividend yield is always negative and decreasing. This sign is found in the literature and it is a signal of mean-reversion in stocks. The residual covariance between the EBR and the term premium changes sign form negative to positive and finally becomes negative again. Campbell et al. (2003) found a positive residual covariance, which would be a signal of mean-aversion in bonds, but its value was low. The residual variances in the constant mean version do not show a huge increase with respect to the predictable version. This is expected as the classical R^2 using predictors is low in monthly predictive regressions.

4.2.2 Evaluation of Joint Density Forecasts

The density forecasts from each version of this model are shown in Figures 4 and 5. In the case of stocks and predictable mean version, Panels A and B in Figure 4, the standard deviation increases a little bit around 1987. The mean is unstable around the end of the 70's and the beginning of the 80's, where it is usually negative. The mean is also usually negative during the second half of the 80's. The coefficients of asymmetry and kurtosis are close to zero all the time, see Panels C and D in Figure 4. The EBR density forecast shows an increase in the standard deviation around 1980. The mean is negative around 1979 and 1980, and during the second half of the 90's. The coefficients of asymmetry and kurtosis are also close to zero all the time.

The constant mean version of the model shows a much more stable mean as it can be expected, see Figure 5. It is positive for ESR, Panel A, but has a negative period for EBR during the first half of the 80's, Panel C. Given the big difference between both versions of density forecasts, they will imply different portfolio choices as it is found in the literature.[12]

The corresponding time series means of mean, standard deviation, and coefficients of

[12] See for instance Barberis (2000), although he focuses on long-term buy-and-hold strategies and re-balancing at frequencies lower than 1 year.

asymmetry and kurtosis are 0.67, 4.14, -0.00, and 0.02, respectively, for the ESR's density forecasts with predictable mean. The corresponding EBR's values are 0.13, 1.45, 0.00 and 0.02. In the case of the constant mean version, the values are 0.45, 4.26, -0.00, and 0.02 for ESR, and 0.04, 1.47, -0.00, and 0.02 for EBR. The main difference between both sets of forecasts is that the time series mean of the latter is lower for both assets. But at this stage we do not know if this is due to over-optimism using predictors or over-pessimism with a constant mean. The study of the PITs will help us to answer this question.

PITs' time series show the following features. In the predictable mean version, ESR's PIT is inversely related to the evolution of the conditional mean. That relationship is not so clear for EBR's PIT. In the case of ESR, the constant mean version shows a similar pattern but it seems less variable and systematic.

Some PIT's diagnostics from the predictable mean version are displayed in Figure 6. The CDFs do not show a significant departure form uniformity, while the correlograms exhibit a failure in terms of the second power of the EBR's PIT, see Panel B. This is confirmed by the numbers in Table 3.

(Table 3)

Panel A shows the sample distribution is close to uniform.[13] In Panel B, the ESR's PIT does not exhibit correlation, but the EBR's PIT shows more correlation. The level shows correlation at the first lag, but it is clear in the second power from the second lag. The EDF tests do not show a significant departure from uniformity, see Panel C in Table 3. The statistics are shown with $*$ if it is significant at 5%, $**$ if it is significant at 2.5%, and $***$ if it is significant at 1%. However, the Wallis test, Panel D, gives some concern for ESR with 3 bins, the conditional coverage p-value is 0.01, while has an independence p-value of 0.02 with 4 bins. The joint PIT inherits the correlation in the second power of the EBR and the bad performance of ESR in the Wallis test, mainly in the case of the independence test.

We describe now the corresponding graphical diagnostics for PITs from the constant mean version. In terms of CDFs, they seem to worsen with respect to the predictable version. The CDF is bellow the 45 degrees line, which may be due to a higher mean. Panel A in Table 3 shows that the means of the PITs are around 0.52, while the predictable

[13] The corresponding four parameters for a uniform distribution are 0.50, 0.29, 0.00, and -1.20.

mean versions had a mean of around 0.51. The main difference is that the asymmetry doubles its negative value in the constant mean version. Now we see that the right tail is too thick and the left tail too thin in the constant mean PITs. We commented that time series means of the density forecasts were lower for the case of constant mean and we wondered if that was due to over-pessimistic forecasts. The PIT now provides the answer.

In terms of correlograms, the level correlation in ESR increases, but the second power correlation decreases. This is confirmed by numbers in Panel B of Table 3. The level correlation of EBR increases at the first lag. The EDF tests worsen, Panel C in Table 3, with Anderson-Darling rejected at the 2.5% in the case of joint PIT. Wallis improves for ESR, Panel D, but worsens for EBR because of the independence test. The joint PIT has lower p-values than before. However, there are no rejections at the 2.5%.

We studied the difference between time series of PITs from both versions. The predictable mean version gives lower PITs in general until the beginning of the 90's, 1993 for ESR and 1995 for EBR. This is due to a higher mean in its density forecasts. Then the tendency changes and it usually gives higher PITs because it predicted negative returns. Apart from the mean, the rest of features of the forecasts are similar. The correlation between the PITs ranges from 0.94 to 0.97.

The previous results are based on integrated forecasting rules but a plug-in rule was also evaluated. The difference between time series of PITs from both forecasting rules is low and not systematic for the predictable version, with higher volatility during unstable periods, mainly the beginning of the 80's and the last years. The difference is small in the case of the constant mean version, as it can be expected since the conditional version drives many of the results. All evaluation tests give similar results when applied to both forecasting rules. The correlation between corresponding PITs is 1. Correlations between 4 moments of density forecasts that are computed in this paper are also 1 or very close. The parameter uncertainty per se is not able to create a significant amount of asymmetry or kurtosis at the monthly frequency. Stambaugh (1999) and Barberis (2000) show that estimation risk might be important in asset allocation but only at much lower frequencies.

In summary, we have shown the following relevant results for an investor who rebalances her portfolio monthly. The performance of the PITs from the normal VAR is good in general, so it can be used to compute optimal asset allocations. There is only a failure in the dependence of the second power of the EBR's PIT. On the other hand, it is not clear if constant or time-varying risk premia give better forecasting performance. With respect to constant risk premia, time-varying premia worsen the forecast evaluation for stocks but improve for bonds. Nevertheless, the PIT shows a common feature in stocks and bonds: constant risk premia tend to produce over-pessimistic forecasts. Finally, using the integrated or the plug-in forecasting rules does not matter so the investor can avoid the complexity of the former.

5 Markov-Switching VAR

The previous model was mainly concerned with time-variation in risk premia. Now we study models that add time-varying risk by enriching the Gaussian VAR. This can be done with GARCH, stochastic volatility, or regime-switching models. Hamilton and Susmel (1994) explore ARCH versus regime-switching effects in variance with U.S. monthly data. They conclude that ARCH effects are not very important at the monthly frequency after taking into account regime-switching.

In addition, a clear asymmetry is found in asset correlations during bull and bear markets. See for instance the case of stocks in Ang and Chen (2002). Regime-switching models are well suited to separate the asymmetric behaviour of returns in different states as it is pointed out in Ramchad and Susmel (1998) and Ang and Bekaert (2002), who also apply these models to asset allocation.

For both reasons, we will estimate and evaluate a Markov-switching model.[14] We will check the robustness of results through alternative regime-switching models at the end of this section.

5.1 Econometric Model

A Markov-switching VAR introduces an unobserved discrete variable that follows a Markov chain and whose outcome defines the values of the vector of VAR parameters. The Markov chain represents persistence in economic regimes and differentiates this model from a simple Gaussian mixture. Unobservability of the state is a convenient approach to make real-time forecasts.

[14] Hamilton (1989) developed the reference Markov-switching model, while Kim and Nelson (1998) showed its implementation in a Bayesian framework to macroeconomic indicators. Pérez-Quirós and Timmermann (2000, 2001), Chauvet and Potter (2001), and Bansal and Zhou (2002) are some examples in finance.

Only two regimes are allowed, so that the regime variable s_t can take only two values, say 1 or 2. The unobservable regime follows a binary Markov chain,

$$p\left(s_t \mid s^{t-1}, \mathbf{y}^{t-1}, \boldsymbol{\theta}\right) = p\left(s_t \mid s_{t-1}, \boldsymbol{\theta}\right), \tag{9}$$
$$p\left(s_t = j \mid s_{t-1} = i, \boldsymbol{\theta}\right) = p_{ij}, \quad i, j = 1, 2.$$

The specific model that we will estimate and evaluate can be expressed with the same VAR as (6), i.e. the conditional mean parameters are kept constant in both regimes,[15] plus a different distribution of the perturbation. Instead of (7), now we have

$$\mathbf{u}_t \mid \left(\mathbf{y}^{t-1}, s_t = j, \boldsymbol{\theta}\right) \sim N\left(\mathbf{0}, \Sigma_j\right), \quad j = 1, 2 \tag{10}$$

where the distribution of $\mathbf{u}_t$ is expressed conditional on the regime. This is our way of introducing time-varying risk. The vector $\boldsymbol{\theta}$ represents the model parameters and its specific components will be shown later.

Our Markov-switching VAR has 4 variables, but it is not usual to find this type of VARs with more than 2 or 3 variables. For instance, Pérez-Quirós and Timmermann (2000, 2001) apply the model to 1 or 2 series of stock returns, without introducing predictors' dynamics. Ang and Bekaert (2001) apply it to 3 returns (short-term rate, bonds and stocks) but their only predictor is one of the returns, the short term rate. We will show that the joint modelling of returns and predictors is crucial in the identification of regimes.

Two versions of this model will be estimated and evaluated again, facing time-varying against constant risk premia. The case of constant risk premia is defined with the same restrictions on $\mathbf{B}$ as in the Gaussian model (8).

The parameters to estimate are now

$$\boldsymbol{\theta} = \left\{\left(p_{11}, p_{22}\right), \boldsymbol{\pi}, \left(\Sigma_1, \Sigma_2\right)\right\},$$

and the likelihood is

$$p\left(\mathbf{y}^T \middle| \boldsymbol{\theta}\right) = \prod_{t=1}^{T} \sum_{s_t} p\left(s_t \mid y^{t-1}, \boldsymbol{\theta}\right) p\left(\mathbf{y}_t \middle| \mathbf{y}^{t-1}, \boldsymbol{\theta}, s_t\right),$$

[15] This model does not allow asymmetry in returns. A model with switches in the conditional mean was estimated in a previous version of this paper and there was not a clear evidence of switching in any parameter.

where $p\left(s_t \mid \mathbf{y}^{t-1}, \boldsymbol{\theta}\right)$ is given by the filter in Hamilton (1989) and the component $p\left(\mathbf{y}_t \middle| \mathbf{y}^{t-1}, \boldsymbol{\theta}, s_t\right)$ is Gaussian. The priors have the following structure,

$$p_{11} \sim B\left(\underline{n}_{11}, \underline{n}_{12}\right), \quad p_{22} \sim B\left(\underline{n}_{22}, \underline{n}_{21}\right),$$
$$\boldsymbol{\pi} \sim N\left(\underline{\boldsymbol{\pi}}, \underline{\mathbf{C}}\right),$$
$$\Sigma_1 \sim IW\left(\underline{\nu}_1, \underline{\mathbf{S}}_1\right), \quad \Sigma_2 \middle| \Sigma_1 \sim TIW_{\mathcal{I}(\sigma_{2,11} > \sigma_{1,11})}\left(\underline{\nu}_2, \underline{\mathbf{S}}_2\right)$$

where $B(\cdot,\cdot)$ means a beta distribution[16] and $TIW_{\mathcal{I}(\cdot)}(\cdot,\cdot)$ is a truncated inverse Wishart with $\mathcal{I}(\cdot)$ defining the truncation region. The prior of Σ_2 given Σ_1 is truncated to region where the first entry of Σ_2 is higher that the first entry of Σ_1. The priors on p_{11} and p_{22} are independent, but the priors on Σ_1 and Σ_2 are defined jointly to identify the regimes during the MCMC. Informative priors are needed with these type of models because it could be the case that some regimes are not visited in some iterations of the Gibbs sampler. More details about the implementation of the model such as specific prior values and corresponding Gibbs sampler can be found in the appendix.

In terms of density forecasts, latent variables imply that we cannot directly rely on normal distributions. Nevertheless, we are back to the Gaussian world conditioning on the latent variables. Since

$$\mathbf{y}_{T+1} \mid \left(\mathbf{y}^T, \boldsymbol{\theta}, s_{T+1} = j\right) \sim N\left(\mathbf{a} + \mathbf{B}\mathbf{y}_T, \Sigma_j\right),$$

we have that

$$e^S_{T+1} \mid \left(\mathbf{y}^T, \boldsymbol{\theta}, s_{T+1} = j\right) \sim N\left(a^S + \left(\mathbf{b}^S\right)' \mathbf{y}_T, \sigma_j^{SS}\right),$$

where a^S is the corresponding entry in $\mathbf{a}$, $\left(\mathbf{b}^S\right)'$ is the corresponding row in $\mathbf{B}$, and σ_j^{SS} is the corresponding entry in Σ_j. In addition,

$$e^B_{T+1} \mid \left(\mathbf{y}^T, e^S_{T+1}, \boldsymbol{\theta}, s_{T+1} = j\right) \sim N\left(a^B + \left(\mathbf{b}^B\right)' \mathbf{y}_T + \frac{\sigma_j^{SB}}{\sigma_j^{SS}} u^S_{T+1}, \sigma_j^{BB} - \frac{\left(\sigma_j^{SB}\right)^2}{\sigma_j^{SS}}\right).$$

The distributions of $e^B_{T+1} \mid \left(\mathbf{y}^T, \boldsymbol{\theta}, s_{T+1} = j\right)$ and $e^S_{T+1} \mid \left(\mathbf{y}^T, e^B_{T+1}, \boldsymbol{\theta}, s_{T+1} = j\right)$ are defined in a similar way. The required moments can be computed following (2) and (3), while the PITs are described in (4) and (5).

[16] The parametrization is such that if $x \sim B(\alpha, \beta)$ then $E(x) = \alpha / (\alpha + \beta)$.

5.2 Estimation and Evaluation

We show again the empirical results of the model in two dimensions, estimation of parameters and evaluation of real-time density forecasts.

5.2.1 Real-Time Model Estimation

Table 4 shows posterior means of some parameters of the two model versions. Again, the estimation is shown for first and last estimation periods, December 1974 and November 2001, and there will be comments that refer to some computations that are not shown in tables.

(Table 4)

The values of transition probabilities imply persistent regimes. Both regimes are similarly persistent in the first estimation, but the high-volatility regime persistence decreases in the last estimation. The persistence of the low-volatility regime decreases at the beginning but afterwards increases so that ends at a similar level, while the other regime persistence is decreasing along the whole period. The posterior means for the predictable mean version would imply a steady state probability for the low-volatility regime that goes from 0.51 to 0.77 during the data period.

The values of conditional mean parameters are similar to the Gaussian VAR model. Perhaps it is strange to see now a negative intercept for bonds in the first estimation of the constant mean version, but it is not significant.

In terms of residual variances, recall that our only identifying assumption was that the second regime had a higher first entry. The rest of entries are left unrestricted and the model identifies a high-volatility regime for both assets. In the first estimation, stocks approximately double its variance form one regime to the other. Bond variance is approximately eight times higher, but its values are much lower than stock's ones. The covariance is not really significant in both regimes. In the second estimation, switching is about three times for stocks and six times for bonds. All the values are higher than the first estimation, especially the high-volatility regime. The stock variance has increased along the sample period in both regimes. The bond variance in the first regime grows smoothly in general but suffers a jump at the beginning of the 90's. The corresponding second regime variance jumps at the beginning of the 80's and the 90's.

About covariances, only the low-volatility regime one can be regarded as significantly positive. In fact, the corresponding correlations in each regime, computed at posterior means of parameters, are 0.19 versus 0.12 for the predictable mean version. The first regime covariance starts at a negative value but becomes positive and grows very soon, while the second regime covariance begins growing and becomes stable afterwards. About residual covariances between returns and predictors, the covariance between ESR and dividend yield is negative in both regimes and both periods. In the final estimation, this covariance turns to be about three times more negative in the high-volatility regime. On the other hand, the covariance between EBR and term premium is negative in the first regime, but the sign is not clear in the second. The high-volatility regime covariance is significantly positive in the constant mean version.

Finally, the probability of the high-volatility regime from last estimation can be seen in Panel A of Figure 7, jointly with the NBER recession index. The high-volatility regime is usually associated with recessions and that is a noteworthy contribution of this paper. This model identifies regimes that are closely related to the business cycle without using macroeconomic variables such as GDP and/or time-varying transition probabilities. For instance, Pérez-Quirós and Timmermann (2000, 2001) introduce that variation as dependence of transition probabilities on the composite leading indicator. We do not need that model because we work with a joint VAR of excess returns and predictors, not with a VAR of excess returns only.

5.2.2 Evaluation of Joint Density Forecasts

Density forecasts from each model version are shown in Figures 8 and 9. The case of stocks and predictable mean version is displayed in Panels A and B of Figure 8. The standard deviation shows more action now, while the mean shows a bit lower variation. The coefficient of asymmetry is still close to zero all the time, but the coefficient of kurtosis varies through time and increases during the 90's. The EBR density forecast changes more compared to the Gaussian VAR, see Panels C and D of Figure 8. There is more variation in the standard deviation compared to the mean, while the mean shows smaller peaks. The coefficient of asymmetry is again close to zero all the time, but kurtosis now changes very sharply and is higher than ESR's kurtosis. The constant mean version of the model, Figure 9, does not need additional comments.

In terms of numbers, the time series means of mean, standard deviation, and coefficients of asymmetry and kurtosis are 0.71, 4.20, -0.00, and 0.36, respectively, for ESR's density forecasts with predictable mean. The corresponding EBR's values are 0.10, 1.48, 0.00 and 1.97. In the case of the constant mean version, values are 0.48, 4.32, -0.00, and 0.47 for ESR, and 0.01, 1.50, -0.00, and 2.01 for EBR. The main difference between both versions of forecasts is that the time series mean is lower for both assets. Therefore, the previous result found for the Gaussian VAR model is not specific to that model. In both versions, the main difference with the previous Gaussian VAR forecasts is the increase in kurtosis, especially in EBR. The correlation between means from the predictable mean version of both models is around 0.99 for ESR, and around 0.85 for EBR. These values are similar for other two regime-switching models that were also evaluated. They are briefly commented at the end of this section.

Now we turn to the study of PITs. The time series of PITs do not show many differences with respect to the Gaussian VAR. The correlation between corresponding PITs from each model is 1 for ESR and around 0.97-0.98 for EBR. Again, these values are similar for the additional regime-switching models that were also evaluated. If we compute the time series of differences of PITs from the predictable mean versions of both models then we can see a similar pattern to the difference between the predictable and constant mean versions of the Gaussian model. That time series exhibits a negative value before the 90's in general and a positive value afterwards. In fact, the difference between PITs can also be related with the difference in mean of density forecasts. Regardless, the constant mean version shows a lower difference, there is a higher range for EBR again, and does not have a clear sign.

Figure 10 shows some PITs' graphical diagnostics from the predictable mean version. The CDFs do not show a significant departure form uniformity, although the EBR in Panel B looks worse that the Gaussian VAR. It has a fatter right tail, which translates also into the joint PIT in Panel C. The noteworthy feature is the lack of correlation in the second power of EBR, although there is still some in the joint PIT. We commented there is not a direct relation between marginal and dynamic properties of PITs and returns, but in this case a richer specification of the conditional variance of returns has removed the correlation in the second power of the corresponding PIT.

The corresponding numbers and tests can be found in Table 5. Panel A shows some descriptive statistics that are close to its uniform counterparts, while Panel B shows the commented lack of second power correlation in EBR. EDF tests of EBR, Panel C, worsen with respect to Gaussian VAR, specifically the Kuiper test of the joint PIT is rejected at the 2.5%. The Wallis test, Panel D, shows a p-value of only 0.001 for ESR with 4 bins due to the independence test. EBR's p-values decrease, mainly due to the uniformity test. The unconditional coverage p-value for EBR with 4 bins is only 0.012.

(Table 5)

The corresponding graphical diagnostics of PITs from the constant mean version are as follows. In terms of CDFs they seem to worsen with respect to the predictable version. The correlograms do not require additional comments. Table 5 shows the corresponding tests too. First, looking at Panel A, PITs mean from the constant mean version is a bit higher and its asymmetry is more negative, as it was the case in the Gaussian VAR model. Therefore, this feature of PITs is not specific to the normal model. EDF tests worsen, Panel C, with Anderson-Darling rejected at 1% in the case of EBR and joint PIT. Wallis improves for ESR, Panel D, but worsens for EBR leading to a clear rejection with 4 bins. The joint PIT has lower p-values than before, with rejections of conditional coverage at 1%.

In terms of comparing time series of PITs from the two versions, results are similar to the Gaussian VAR. In terms of comparing corresponding PITs from the two forecasting rules, there is not a big difference. One of the new features in the Markov-switching model is kurtosis. The correlation between kurtosis of both forecasting rules is 1 too.

To summarise, we have shown the following relevant results for an investor that rebalances her portfolio monthly. The main difference with density forecasts from the Gaussian VAR is an increase in kurtosis, specially for bonds. The Markov-switching model has removed the dependence in the second power of EBR's PIT, but has worsen other dimensions with respect to the Gaussian VAR such as uniformity of EBR's PIT. This fact reinforces the previous statement that the Gaussian VAR model can be used to compute optimal asset allocations, being much easier to implement. Finally, the relation between PITs from two risk premia versions and the relation between PITs from two forecasting rules are common to the Gaussian and Markov-switching models.

5.3 Alternative Regime-Switching Models

Two alternative regime-switching models were evaluated, which may be called independent-switching and predictor-switching modes. The former is an independent mixture of two regimes and therefore it may be seen as a restricted version of the Markov-switching model. It only introduces fat tails with respect to the Gaussian model. On the other hand, the predictor-switching model defines directly mixture weights as functions of predictors by means of a normal CDF. These functions avoid filtering an unobservable regime. Both models are much easier to estimate than a Markov-switching model so it is interesting to study their performance.

Their corresponding results are not shown to shorten the extension of the paper,[17] apart from their probabilities of the high-volatility regime in Figure 7. The independent-switching model identifies regimes that are not closely related to the business cycle. The behaviour of the probabilities in the predictor-switching model lies in between the other two regime-switching models. That is, they are not as smooth as the Markov-switching ones but are closer to the business cycle than in the independent-switching case. The main difference with the previous models is that the end of the sample is labelled as a low volatility regime although it is a recession for the NBER index. This is due to the strong dividend yield effect in that probability.

These additional two models create higher kurtosis than the Markov-switching in general. However, kurtosis decreases during the 90's in the predictor-switching model because of dividend yieldb4s evolution. The evaluation of these models is worse than the previous ones in general. Both models fail to give a uniform PIT in the case of bonds, with too many realizations on both tails of density forecasts. The predictor-switching model is able to improve dynamics of second powers of bonds' PIT.

After the evaluation of these additional models, we can conclude that previous advices about density forecasts are robust enough. The Gaussian VAR model can be used to compute optimal asset allocations and the complexity of a fully Bayesian approach can be avoided.

[17] They are available upon request from the author.

6 Conclusions

Much of the growing literature on tactical and strategic asset allocation is based on a Gaussian vector autoregression (VAR) of returns and predictors. Unfortunately, the portfolio advice from that literature may be misleading if the estimated Gaussian VAR is not an accurate description of the distribution of returns. For that reason, in this paper we evaluate its joint density forecasts of US monthly excess returns of stocks and bonds during the post-war period. The evaluation method relies on the probability integral transforms (PITs) of the density forecasts. If the model accurately describes the distribution of returns then the PITs should be uniform and independent over time.

We show that an investor who rebalances her portfolio monthly can rely on the standard VAR model since it provides a reasonable description of the data. Richer models do not improve the quality of density forecasts and they are more difficult to implement. The Markov-switching VAR increases the kurtosis of density forecasts, especially in the case of bonds, but does not improve the evaluation results in general. On the other hand, a noteworthy feature of its estimation is that the regime variable tracks the business cycle without the introduction of macroeconomic variables.

The accuracy of the Gaussian VAR explains why the portfolio choices in Brandt (1999), which do not rely in a statistical model of returns, are close to the portfolios computed using a Gaussian VAR model. For instance, in the case of models with time-varying risk premia, the correlation of the PITs is 1 for the stocks and approximately 0.98 for bonds. This is mainly driven by the correlation in risk premia, approximately 0.99 for stocks and 0.85 for bonds.

In addition, two important issues for asset allocation are addressed: time-variation in risk premia and parameter uncertainty. The conclusions about both issues are robust across different models. In terms of time variation in risk premia, versions of models with or without that time variation imply very different density forecasts, translating into significant effects in portfolio choice as it is well known. However, it is not clear if constant or time-varying risk premia give better forecasting performance. With respect to constant risk premia, time-varying premia worsen the forecast evaluation for stocks but improve for bonds. Nevertheless, the PIT shows a common feature in stocks and bonds: constant risk premia tend to produce over-pessimistic forecasts.

When we study the relevance of parameter uncertainty, the fully Bayesian approach and the plug-in forecasting rule give similar density forecasts. This uncertainty does not create much asymmetry or kurtosis in the forecasts. In fact, he correlation between the corresponding descriptive statistics of forecasts and PITs are close to 1. Therefore, an investor that rebalances monthly could avoid the complexity of the fully Bayesian approach.

Although we have relied on particular returns and predictors, we could extend our work to study other returns, predictors and/or countries. For instance, one could analyse portfolios constructed through size and book-to-market criteria as in Lynch (2001). It would also be interesting to introduce a short-term interest rate and its corresponding time-variation. Additionally, it might be worth to evaluate GARCH and stochastic volatility models. Chacko and Viceira (2003) study asset allocation under the latter.

The results of this paper are relevant mainly for an investor that rebalances her portfolio each month and has a short-term perspective. But it may be the case that the investor's rebalancing frequency is lower and/or her investment horizon is longer, and such extensions can be incorporated in the current evaluation method. The first issue can be handled by using returns and density forecasts at lower frequencies while he second issue requires the evaluation of the joint density forecast of returns and predictors. The analysis of rebalancing at lower frequencies might be more interesting since existing empirical results show that the hedging demand component of a long-term portfolio choice is not very important in practice.

Finally, another interesting avenue for future research would be the application of our evaluation methodology to Bayesian models where priors are obtained from an asset pricing model. Pástor (2000), Pástor and Stambaugh (2000), Tan and Zhou (2004), and Avramov (2003) study portfolio choice in that context.

References

Ahït-Sahalia, and M.W. Brandt (2001): Variable Selection and Portoflio Choice, *Journal of Finance* 56, 1297-1351.

Ang, A., and G. Bekaert (2002): International Asset Allocation with Regime Shifts, *Review of Financial Studies* 15, 1137-1187.

Ang, A., and J. Chen (2002): Asymmetric Correlations of Equity Portfolios, *Journal of Financial Economics* 63, 443-494.

Avramov, D. (2002): Stock Return Predictability and Model Uncertainty, *Journal of Financial Economics* 64, 423-458.

Avramov, D. (2003): Stock Return Predictability and Asset Pricing Models, forthcoming in *Review of Financial Studies.*

Bansal, R., and H. Zhou (2002): Term Structure of Interest Rates with Regime Shifts, *Journal of Finance* 57, 1997-2043.

Barberis, N. (2000): Investing in the Long Run when Returns are Predictable, *Journal of Finance* 55, 225-264.

Berkowitz, J. (2001): Testing Density Forecasts, with Applications to Risk Management, *Journal of Business and Economic Statistics* 19, 465-474.

Brennan, M.J., E.S. Schwartz, and R. Lagnado (1997): Strategic Asset Allocation, *Journal of Economic Dynamics and Control* 21, 1377-1403.

Campbell, J.Y., Y.L. Chan, and L.M. Viceira (2003): A Multivariate Model of Strategic Asset Allocation, *Journal of Financial Economics* 67, 41-80.

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

Campbell, J.Y., and R.J. Shiller (1991): Yield Spreads and Interest Rate Movements: A Bird's Eye View, *Review of Economic Studies* 58, 495-514.

Carter, C.K., and R. Kohn (1994): On Gibbs Sampling for State Space Models, *Biometrika* 81, 541-553.

Chacko, G., and L. Viceira (2003): Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets, forthcoming in *Review of Financial Studies.*

Chauvet, M., and S. Potter (2001): Nonlinear Risk, *Macroeconomic Dynamics* 5, 621-646.

Chib, S., and E. Greenberg (1996): Markov Chain Monte Carlo Simulation Methods in Econometrics, *Econometric Theory* 12, 409-431.

Christoffersen, P. (1998): Evaluating Interval Forecasts, *International Economic Review* 39, 841-862.

Clements, M.P., and J. Smith (2000): Evaluating the Forecast Densities of Linear and Non-linear Models: Applications to Output Growth and Unemployment, *Journal of Forecasting* 19, 255-276.

Cremers, K.J.M. (2002): Stock Return Predictability: A Bayesian Model Selection Perspective, forthcoming *Review of Financial Studies* 15, 1223-1249.

Dawid, A.P. (1984): Statistical Theory: The Prequential Approach, *Journal of the Royal Statistical Society* (Series A) 147, 278-292.

Diebold, F.X., T.A. Gunther, and A. S. Tay (1998): Evaluating Density Forecasts with Applications to Financial Risk Management, *International Economic Review* 39, 863-883.

Diebold, F.X., J. Hahn, and A.S. Tay (1999): Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns of Foreign Exchange, *Review of Economics and Statistic*s 81, 661-673.

Fama, E.F., and K.R. French (1988): Dividend Yields and Expected Stock Returns, *Journal of Financial Economics* 22, 3-27.

Fama, E.F., and K.R. French (1989): Business Conditions and Expected Returns on Stocks and Bonds, *Journal of Financial Economics* 25, 23-49.

Hamilton, J.D. (1989): A New Approach to the Economic Analysis of Nonstationay Time Series and the Business Cycle, *Econometrica* 57, 357-384.

Hamilton, J.D., and R. Susmel (1994): Autoregressive Conditional Heteroskedasticity and Changes in Regime, *Journal of Econometrics* 64, 307-333.

Kandel, S., and R.F. Stambaugh (1996): On the Predictability of Stock Returns: An Asset-Allocation Perspective, *Journal of Finance* 51, 385-424.

Kim, C.-J., and C.R. Nelson (1998): Business Cycle Turning Points, a New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching, *Review of Economics and Statisitics* 80, 188-201.

Kim, S., N. Shephard, and S. Chib (1998): Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies* 65, 361-393.

Lynch, A. (2001): Portfolio Choice and Equity Characteristics: Characterizing the

Hedging Demands Induced by Return Predictability, *Journal of Financial Economics* 62, 67-130.

Pástor, L. (2000): Portfolio Selection and Asset Pricing Models, *Journal of Finance* 55, 179-223.

Pástor, L. and R.F. Stambaugh (2000): Comparing Asset Pricing Models: An Investment Perspective, *Journal of Financial Economics* 56, 335-381.

Pérez-Quirós, G., and A. Timmermann (2000): Firm Size and Cyclical Variations in Stock Returns, *Journal of Finance* 55, 1229-1263.

Pérez-Quirós, G., and A. Timmermann (2001): Business Cycle Asymmetries in Stock Returns: Evidence from Higher Order Moments and Conditional Densities, *Journal of Econometric*s 103, 259-306.

Ramchand, L., and R. Susmel (1998): Cross Correlations Across Major International Markets, *Journal of Empirical Finance* 5, 397-416.

Rosenblatt, M. (1952): Remarks on Multivariate Transformations, *Annals of Mathematical Statistics* 23, 470-472.

Shephard, N. (1994): Partial Non-Gaussian State Space, *Biometrika* 81, 115-131.

Stambaugh, R.F. (1999): Predictive Regressions, *Journal of Financial Economics* 54, 375-421.

Stephens, M.A. (1974): EDF Statistics for Goodness of Fit and Some Comparisons, *Journal of the American Statistical Association* 69, 730-737.

Tu, J., and G. Zhou (2004): Data-Generating Process Uncertainty: What Difference Does it Make in Portfolio Decisions?, *Journal of Financial Economics* 72, 385-421.

Wallis, K.F. (2003): Chi-squared Tests of Interval and Density Forecasts, and the Bank of England's Fan Charts, *International Journal of Forecasting*, 19, 165-175.

Appendix

A MCMC Implementation

The initial condition of the MCMC is taken from the maximum likelihood estimator (MLE) of the parameters. The first model is a Gaussian VAR and its MLE is easy to compute. The Markov-switching model introduces latent variables and its MLE is computed using the EM algorithm. The corresponding Gibbs samplers draw $\left\{\boldsymbol{\theta}_i, s_i^T\right\}_{i=1}^{I}$ from the posterior $p\left(\boldsymbol{\theta}, s^T \mid \mathbf{y}^T\right)$, where we are discarding the first $\left\{\boldsymbol{\theta}_i, s_i^T\right\}_{i=1}^{I^*}$ draws for convergence reasons. The convergence of all chains was checked and did not show problems. In the case of the Gaussian VAR, we run 101000 iterations in the computation of the posterior distribution and burn the first 1000. Given its computational cost, we run 51000 iterations and burn the first 1000 for the Markov-switching model.

We can compute the different objects of interest from Gibbs output. For instance, the density forecast (1) at any point $\mathbf{y}_{T+1}$ can be approximated by

$$p^I\left(\mathbf{y}_{T+1} \mid \mathbf{y}^T\right) \simeq \frac{1}{I}\sum_{i=1}^{I}\left[\sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}_i, s_{T,i}\right) p\left(\mathbf{y}_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}_i, s_{T+1}\right)\right].$$

The approximation error can be computed from standard time series methods. Expectations in the computation of PITs in (4).are approximated by the corresponding sample mean from the Gibbs output. For instance,

$$d_{T+1}^{I,S} \simeq \frac{1}{I}\sum_{i=1}^{I}\sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}_i, s_{T,i}\right) P\left(e_{T+1}^S \mid \mathbf{y}^T, \boldsymbol{\theta}_i, s_{T+1}\right).$$

The conditional PIT is computed under the approximation $p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T, e_{T+1}^S\right) \simeq p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T\right)$ so that we can use directly Gibbs output.

B Gaussian VAR

The specific values of the prior hyperparameters are as follows.[18] We choose $\underline{\boldsymbol{\pi}} = \mathbf{0}$ and $\underline{\mathbf{C}}=100\mathbf{I}$, where $\mathbf{I}$ is the corresponding identity matrix. Therefore, the prior on $\boldsymbol{\pi}$ is not very informative. On the other hand, the prior on Σ will be computed from its MLE

[18] Each modelb4s priors are available upon request from the author.

at the first estimation period, up to December 1974. Specifically,

$$\underline{\nu} = \underline{T}, \quad \underline{\mathbf{S}} = \left(\underline{T}\hat{\Sigma}\right)^{-1},$$

where $\underline{T} = 25 \simeq T/10$. This prior represents a small sample centred at the MLE.

We use a small approximation in the computations since we do not reestimate the model monthly. We do it each quarterly, that is, if we use $p\left(\boldsymbol{\theta} \mid \mathbf{y}^{T}\right)$ at period T then we use the approximations $p\left(\boldsymbol{\theta} \mid \mathbf{y}^{T+1}\right) \simeq p\left(\boldsymbol{\theta} \mid \mathbf{y}^{T}\right)$ and $p\left(\boldsymbol{\theta} \mid \mathbf{y}^{T+2}\right) \simeq p\left(\boldsymbol{\theta} \mid \mathbf{y}^{T}\right)$ in the following two periods. Afterwards, we estimate $p\left(\boldsymbol{\theta} \mid \mathbf{y}^{T+3}\right)$ and the previous procedure is repeated again. This does not represent a significant change in the final computations and it really decreases the required time for the recursive out-of-sample check.

Let us define some notation before describing the Gibbs sampler of this model. Defining

$$\mathbf{x}_t = \begin{pmatrix} 1 \\ \mathbf{y}_{t-1} \end{pmatrix},$$

we can write the sample data in matrix notation at a particular point in time T as

$$\mathbf{Y} = \mathbf{X}\Pi + \mathbf{U},$$

where $\mathbf{Y}$ is $T \times 4$ and $\mathbf{X}$ is $T \times 5$. The following notation will be convenient too,

$$\mathbf{y} = vec\left(\mathbf{Y}\right), \quad \widetilde{\mathbf{X}} = \mathbf{I}_4 \otimes \mathbf{X},$$

where $\mathbf{I}_4$ is the identity matrix of order 4. In the case of constant risk premia, the matrix $\widetilde{\mathbf{X}}$ takes the form

$$\widetilde{\mathbf{X}} = \begin{pmatrix} \mathbf{I}_2 \otimes \ell_T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \otimes \mathbf{X} \end{pmatrix}$$

where ℓ_T is a $T \times 1$ vector of ones.

The Gibbs sampler that we use to approximate the posterior distributions is based on the following conditional posteriors of parameter blocks.

- Sampling $\boldsymbol{\pi}$:

Since the likelihood implies the following kernel for $\boldsymbol{\pi}$

$$p\left(\mathbf{y}^{T} \mid \boldsymbol{\theta}\right) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{y} - \widetilde{\mathbf{X}}\boldsymbol{\pi}\right)' \mathbf{V}^{-1}\left(\mathbf{y} - \widetilde{\mathbf{X}}\boldsymbol{\pi}\right)\right\},$$

$$\mathbf{V}^{-1} = \Sigma^{-1} \otimes \mathbf{I}_T,$$

we can sample $\boldsymbol{\pi}$ as

$$\boldsymbol{\pi} \mid \left(\mathbf{y}^T, \Sigma\right) \sim N\left(\bar{\boldsymbol{\pi}}, \bar{\mathbf{C}}\right),$$
$$\bar{\boldsymbol{\pi}} = \bar{\mathbf{C}}\left[\underline{\mathbf{C}}^{-1}\underline{\boldsymbol{\pi}} + \left(\widetilde{\mathbf{X}}\right)'\mathbf{V}^{-1}\mathbf{y}\right],$$
$$\bar{\mathbf{C}} = \left[\underline{\mathbf{C}}^{-1} + \left(\widetilde{\mathbf{X}}\right)'\mathbf{V}^{-1}\widetilde{\mathbf{X}}\right]^{-1}.$$

Obviously, this is only a compact way of expressing the terms, because the dimension of $\mathbf{V}^{-1}$ would make it infeasible. In the case of a predictable mean, the computations are much easier.

- Sampling Σ:

The corresponding kernel from the likelihood can be expressed as

$$p\left(\mathbf{y}^T \mid \boldsymbol{\theta}\right) \propto \left|\Sigma^{-1}\right|^{\frac{T}{2}} \exp\left\{-\frac{1}{2}tr\left(\Sigma^{-1}\mathbf{U}'\mathbf{U}\right)\right\},$$

which implies the conditional posterior

$$\Sigma \mid \left(\mathbf{y}^T, \boldsymbol{\pi}\right) \sim IW\left(\bar{\nu}, \bar{\mathbf{S}}\right),$$
$$\bar{\nu} = \underline{\nu} + T, \quad \bar{\mathbf{S}} = \left[\underline{\mathbf{S}}^{-1} + \mathbf{U}'\mathbf{U}\right]^{-1},$$
$$\mathbf{U} = \mathbf{Y} - \mathbf{X}\Pi.$$

C Markov-Switching VAR

We choose $\underline{n}_{ij} = 1$, for all i and j, and $\underline{\boldsymbol{\pi}} = \mathbf{0}$ and $\underline{\mathbf{C}}=100\mathbf{I}$. On the other hand, the prior on Σ_1 and Σ_2 will be computed from its MLE of each version of the model for the first estimation period, up to December 1974,

$$\underline{\nu}_j = \underline{T}, \quad \underline{\mathbf{S}}_j = \left(\underline{T}\hat{\Sigma}_j\right)^{-1}, \quad j = 1, 2,$$

where $\underline{T} = 25$ again.

We reestimate the model quarterly. If we use $p\left(\boldsymbol{\theta}, s^T \mid \mathbf{y}^T\right)$ at period T then we use approximations at $T+1$ and $T+2$ to take advantage of the Gibbs from T and decrease the number of required computations. We use the exact expression (2) at T, but we use

$$E_{p^I\left(\mathbf{y}_{T+2}|\mathbf{y}^{T+1}\right)}\left[\left(e_{T+2}^S\right)^k\right] =$$

$$E_{p(\boldsymbol{\theta},s_T|\mathbf{y}^{T+1})}\left[\sum_{s_{T+1}} p\left(s_{T+1} \mid \boldsymbol{\theta}, s_T\right) \sum_{s_{T+2}} p\left(s_{T+2} \mid \boldsymbol{\theta}, s_{T+1}\right) E_{p\left(\mathbf{y}_{T+2}|\mathbf{y}^{T+1},\boldsymbol{\theta},s_{T+2}\right)}\left[\left(e_{T+2}^S\right)^k\right]\right]$$

at period $T+1$, where the only approximation is $p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T\right) \simeq p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^{T+1}\right)$. The sums over regimes can be directly computed by means of powers of the transition probabilities matrix. A similar computation is applied at period $T+2$. Afterwards, we estimate $p\left(\boldsymbol{\theta}, s^{T+3} \mid \mathbf{y}^{T+3}\right)$ and the previous procedure is repeated again. We evaluate the previous expressions at the corresponding posterior means in the case of the plug-in rule.

The estimation is based on Kim and Nelson (1998), where data augmentation with regimes s^T is the key element to return to the Gaussian world. The Gibbs sampler that we use to approximate the posterior distributions is based on the following conditional posteriors of parameter blocks.

- Sampling s^T:

The sample regimes are drawn following Carter and Khon (1994) and Shephard (1994) who advocate the use of a multimove sampler instead of a singlemove one. The required distribution is

$$p\left(s^T \mid \mathbf{y}^T\right) \propto p\left(s_T \mid \mathbf{y}^T\right) \prod_{t=1}^{T-1} p\left(s_t \mid \mathbf{y}^t, s_{t+1}\right),$$

$$p\left(s_t = j \mid \mathbf{y}^t, s_{t+1}\right) \propto p\left(s_t = j \mid \mathbf{y}^t\right) p\left(s_{t+1} \mid s_t = j\right)$$

where $\boldsymbol{\theta}$ has been suppressed from all the conditioning sets for ease of notation. To implement that sampler, we must run the filter developed by Hamilton (1989) before to compute $p\left(s_t \mid \mathbf{y}^t\right)$.

- Sampling (p_{11}, p_{22}):

The sampler of probabilities is based on

$$p\left(p_{11}, p_{22} \mid \mathbf{y}^T, s^T, \boldsymbol{\pi}, (\Sigma_1, \Sigma_2)\right) = p\left(p_{11}, p_{22} \mid s^T\right) \propto p\left(p_{11}, p_{22}\right) p\left(s^T \mid p_{11}, p_{22}\right).$$

Therefore, we can draw p_{11} from

$$p_{11} \mid \left(\mathbf{y}^T, s^T, \boldsymbol{\pi}, (\Sigma_1, \Sigma_2)\right) \sim B\left(\bar{n}_{11}, \bar{n}_{12}\right),$$

$$\bar{n}_{11} = \underline{n}_{11} + n_{11}, \quad \bar{n}_{12} = \underline{n}_{12} + n_{12},$$

where n_{ij} is the number of transitions from $s_t = i$ to $s_{t+1} = j$. This distribution ignores the dependence of $p\left(s_1 \mid p_{11}, p_{22}\right)$ on transition probabilities and focuses on $p\left(s_t \mid s_{t-1}\right)$, which is a small approximation that greatly simplifies the computations.

Given the independence of priors, by a similar argument,

$$p_{22} \mid \left(\mathbf{y}^T, s^T, \boldsymbol{\pi}, (\Sigma_1, \Sigma_2)\right) \sim B\left(\bar{n}_{22}, \bar{n}_{21}\right)$$

$$\bar{n}_{22} = \underline{n}_{22} + n_{22}, \quad \bar{n}_{21} = \underline{n}_{21} + n_{21}.$$

- Sampling $\boldsymbol{\pi}$:

By means of regimes, the likelihood can be decomposed as

$$\exp\left\{-\frac{1}{2}\left(\mathbf{y}_1 - \widetilde{\mathbf{X}}_1\boldsymbol{\pi}\right)' \mathbf{V}_1^{-1}\left(\mathbf{y}_1 - \widetilde{\mathbf{X}}_1\boldsymbol{\pi}\right)\right\} \exp\left\{-\frac{1}{2}\left(\mathbf{y}_2 - \widetilde{\mathbf{X}}_2\boldsymbol{\pi}\right)' \mathbf{V}_2^{-1}\left(\mathbf{y}_2 - \widetilde{\mathbf{X}}_2\boldsymbol{\pi}\right)\right\},$$

which implies that the conditional posterior is

$$\boldsymbol{\pi} \mid \left(\mathbf{y}^T, s^T, (p_{11}, p_{22}), (\Sigma_1, \Sigma_2)\right) \sim N\left(\bar{\boldsymbol{\pi}}, \bar{\mathbf{C}}\right),$$

$$\bar{\boldsymbol{\pi}} = \bar{\mathbf{C}}\left[\underline{\mathbf{C}}^{-1}\underline{\boldsymbol{\pi}} + \left(\widetilde{\mathbf{X}}_1\right)' \mathbf{V}_1^{-1}\mathbf{y}_1 + \left(\widetilde{\mathbf{X}}_2\right)' \mathbf{V}_2^{-1}\mathbf{y}_2\right],$$

$$\bar{\mathbf{C}} = \left[\underline{\mathbf{C}}^{-1} + \left(\widetilde{\mathbf{X}}_1\right)' \mathbf{V}_1^{-1}\widetilde{\mathbf{X}}_1 + \left(\widetilde{\mathbf{X}}_2\right)' \mathbf{V}_2^{-1}\widetilde{\mathbf{X}}_2\right]^{-1}.$$

- Sampling (Σ_1, Σ_2):

The distribution of $\mathbf{u}_t$ given s^T is

$$\mathbf{u}_t \mid s_t = j \overset{iid}{\sim} N\left(\mathbf{0}, \Sigma_j\right), \quad j = 1, 2,$$

conditioning on the rest of parameters and the data.

This sampler requires dividing data by means of regime variable $s_t = 1$ and 2. We can define $\mathbf{U}_j = \mathbf{Y}_j - \mathbf{X}_j \Pi$, $j = 1, 2$, for data that are assigned to regime j in one particular iteration of the Gibbs sampler. This splits the likelihood information for each regime's parameters in each regime. If all realizations of regimes are equal in a particular iteration, the other regime's parameters are sampled directly from the prior.

The residual variance in the first regime is sampled as

$$\Sigma_1 \mid \left(\mathbf{y}^T, s^T, (p_{11}, p_{22}), \boldsymbol{\pi}\right) \sim IW\left(\bar{\nu}_1, \bar{\mathbf{S}}_1\right),$$
$$\bar{\nu}_1 = \underline{\nu}_1 + T_1, \quad \bar{\mathbf{S}}_1 = \left[\underline{\mathbf{S}}_1^{-1} + \mathbf{U}_1'\mathbf{U}_1\right]^{-1},$$

and the variance in the second regime follows a truncated distribution given the previous draw

$$\Sigma_2 \mid \left(\mathbf{y}^T, s^T, (p_{11}, p_{22}), \boldsymbol{\pi}, \Sigma_1\right) \sim TIW_{\mathcal{I}(\sigma_{2,11} > \sigma_{1,11})}\left(\bar{\nu}_2, \bar{\mathbf{S}}_2\right),$$
$$\bar{\nu}_2 = \underline{\nu}_2 + T_2, \quad \bar{\mathbf{S}}_2 = \left[\underline{\mathbf{S}}_2^{-1} + \mathbf{U}_2'\mathbf{U}_2\right]^{-1}.$$

The truncated inverse Wishart is sampled using an accept-reject method via the inverse Wishart.

D Tables

Table 1. Descriptive Statistics of Excess Stock and Bond Returns.

Descriptive statistics are shown for three periods. The first estimation period, 252 observations, the out-of-sample period, 324 observations, and the whole period. Each series' distribution is described by its mean, standard deviation, and its coefficients of skewness and kurtosis. Finally, its dynamic properties are described by Ljung-Box test's p-values of level and second power of demeaned series. This test is shown for 1, 3, and 6 lags.

		1954/01-1974/12		1975/01-2001/12		1954/01-2001/12	
		ESR	EBR	ESR	EBR	ESR	EBR
Panel A. Descriptive Statistics.							
Mean		0.347	0.011	0.608	0.169	0.494	0.100
Stan. Deviation		4.053	1.176	4.572	1.692	4.351	1.489
Coef. of Skewness		-0.423	0.076	-0.971	-0.001	-0.772	0.068
Coef. of Kurtosis		0.865	3.013	3.873	3.515	2.928	4.118
Panel B. Ljung-Box Test. P-values of 1, 3, and 6 Lags.							
Level	1	0.043	0.753	0.403	0.004	0.079	0.003
	3	0.151	0.953	0.331	0.008	0.252	0.008
	6	0.022	0.410	0.392	0.015	0.310	0.041
Second Power	1	0.000	0.139	0.557	0.015	0.041	0.000
	3	0.000	0.031	0.946	0.000	0.065	0.000
	6	0.000	0.001	0.998	0.000	0.188	0.000

Table 2. Gaussian VAR Estimation. Posterior Means of some Parameters.

The Gaussian VAR is defined by (6) and (7). Posterior means of some parameters are shown for estimations at December 1974 and November 2001. Results are shown at each date for time-varying and constant risk premia versions of the model. In the former version, first parameters are slopes of excess stock return (ESR) and excess bond return (EBR) with respect to dividend yield (DY) and term premium (TP). In the latter version, intercepts are shown. Finally, residual variances and covariances of ESR and EBR are shown.

		1974/12		2001/11	
		Time-varying	Constant	Time-varying	Constant
Π	ESR,DY	2.620	0.333	0.827	0.485
	ESR,TP	1.052		0.478	
	EBR,DY	0.897	0.012	-0.008	0.102
	EBR,TP	0.196		0.192	
Σ	ESR,ESR	15.169	16.716	18.333	19.104
	ESR,EBR	0.166	0.324	0.944	1.130
	EBR,EBR	1.382	1.410	2.105	2.240

Table 3. PITs of Gaussian VAR.

The model is defined by (6) and (7). Three PITs are studied for two versions of the model. Panel A shows four descriptive statistics of each PIT. Panel B studies each PIT's autocorrelation by means of Ljung-Box test. Panel C shows some empirical distribution function (EDF) tests. Finally, Panel D shows Wallis tests of unconditional coverage (UC), independence (IN), and conditional coverage (CC).

		Time-varying			Constant		
		ESR	EBR	Joint	ESR	EBR	Joint
Panel A. Descriptive Statistics.							
Mean		0.506	0.508	0.505	0.524	0.523	0.522
Stan. Deviation		0.299	0.286	0.291	0.285	0.286	0.285
Coef. of Skewness		-0.074	-0.051	-0.047	-0.181	-0.116	-0.137
Coef. of Kurtosis		-1.308	-1.110	-1.202	-1.155	-1.090	-1.121
Panel B. Ljung-Box Test. P-values of 1, 3, and 6 Lags.							
Level	1	0.375	0.030	0.293	0.834	0.006	0.012
	3	0.221	0.186	0.289	0.865	0.049	0.077
	6	0.046	0.228	0.082	0.668	0.102	0.128
Second Power	1	0.890	0.092	0.135	0.251	0.087	0.025
	3	0.172	0.000	0.000	0.081	0.000	0.000
	6	0.023	0.000	0.000	0.001	0.000	0.000
Panel C. EDF Tests. Statistic and Significance.							
Kolmogorov-Smir.		0.875	0.761	0.643	1.273	0.995	1.384*
Kuiper		1.520*	1.163	1.120	1.700*	1.288	1.787*
Anderson-Darling		1.039	0.996	1.158	1.929*	1.933*	3.274**
Panel D. Wallis Tests. P-values of UC, IN, and CC.							
3 Bins	UC	0.046	0.796	0.292	0.294	0.213	0.156
	IN	0.034	0.154	0.025	0.099	0.029	0.137
	CC	0.011	0.309	0.034	0.115	0.031	0.098
4 Bins	UC	0.587	0.972	0.826	0.116	0.381	0.056
	IN	0.018	0.536	0.447	0.230	0.035	0.355
	CC	0.039	0.768	0.634	0.128	0.049	0.131

Table 4. Markov-Switching VAR Estimation. Posterior Means of Some Parameters.

The Markov-switching VAR model is defined by (6), (10), and (9). Posterior means of some parameters are shown for estimations at December 1974 and November 2001. Results are shown at each date for time-varying and constant risk premia versions of the model. First, transition probabilities are shown. In the time-varying risk premia version, first parameters are slopes of excess stock return (ESR) and excess bond return (EBR) with respect to dividend yield (DY) and term premium (TP). In the second version, intercepts are shown. Finally, residual variances and covariances of ESR and EBR are shown for each regime.

		1974/12		2001/11	
		Time-varying	Constant	Time-varying	Constant
p_{11}		0.932	0.939	0.951	0.943
p_{22}		0.928	0.937	0.838	0.838
Π	ESR,DY	2.791	0.339	0.677	0.600
	ESR,TP	0.965		0.383	
	EBR,DY	0.370	-0.050	-0.052	0.083
	EBR,TP	0.155		0.190	
Σ_1	ESR,ESR	11.367	12.000	13.470	12.934
	ESR,EBR	-0.321	-0.315	0.726	0.719
	EBR,EBR	0.352	0.345	1.036	1.030
Σ_2	ESR,ESR	20.251	22.655	36.058	38.133
	ESR,EBR	0.762	1.063	1.750	2.364
	EBR,EBR	2.665	2.690	5.936	5.921

Table 5. PITs of Markov-Switching VAR.

The model is defined by (6), (10), and (9). Three PITs are studied for two versions of the model. Panel A shows four descriptive statistics of each PIT. Panel B studies each PIT's autocorrelation by means of Ljung-Box test. Panel C shows some empirical distribution function (EDF) tests. Finally, Panel D shows Wallis tests of unconditional coverage (UC), independence (IN), and conditional coverage (CC).

		Time-varying			Constant		
		ESR	EBR	Joint	ESR	EBR	Joint
Panel A. Descriptive Statistics.							
Mean		0.504	0.518	0.509	0.522	0.536	0.528
Stan. Deviation		0.299	0.304	0.301	0.288	0.305	0.296
Coef. of Skewness		-0.068	-0.089	-0.071	-0.182	-0.148	-0.158
Coef. of Kurtosis		-1.341	-1.343	-1.330	-1.205	-1.329	-1.267
Panel B. Ljung-Box Test. P-values of 1, 3, and 6 Lags.							
Level	1	0.169	0.003	0.120	0.988	0.001	0.006
	3	0.258	0.028	0.302	0.875	0.013	0.050
	6	0.090	0.097	0.245	0.719	0.040	0.103
Second Power	1	0.950	0.752	0.584	0.426	0.613	0.183
	3	0.212	0.305	0.063	0.168	0.630	0.124
	6	0.020	0.397	0.009	0.004	0.629	0.007
Panel C. EDF Tests. Statistic and Significance.							
Kolmogorov-Smir.		0.965	1.259	1.282	1.242	1.601*	1.782**
Kuiper		1.632*	1.817*	2.013**	1.717*	1.798*	2.009**
Anderson-Darling		1.023	1.828*	1.957*	1.676*	3.982***	4.107***
Panel D. Wallis Tests. P-values of UC, IN, and CC.							
3 Bins	UC	0.137	0.068	0.120	0.133	0.006	0.002
	IN	0.033	0.238	0.054	0.547	0.253	0.265
	CC	0.025	0.092	0.035	0.311	0.015	0.007
4 Bins	UC	0.118	0.011	0.012	0.079	0.006	0.006
	IN	0.012	0.372	0.280	0.062	0.003	0.043
	CC	0.008	0.051	0.039	0.028	0.000	0.003

E Figures

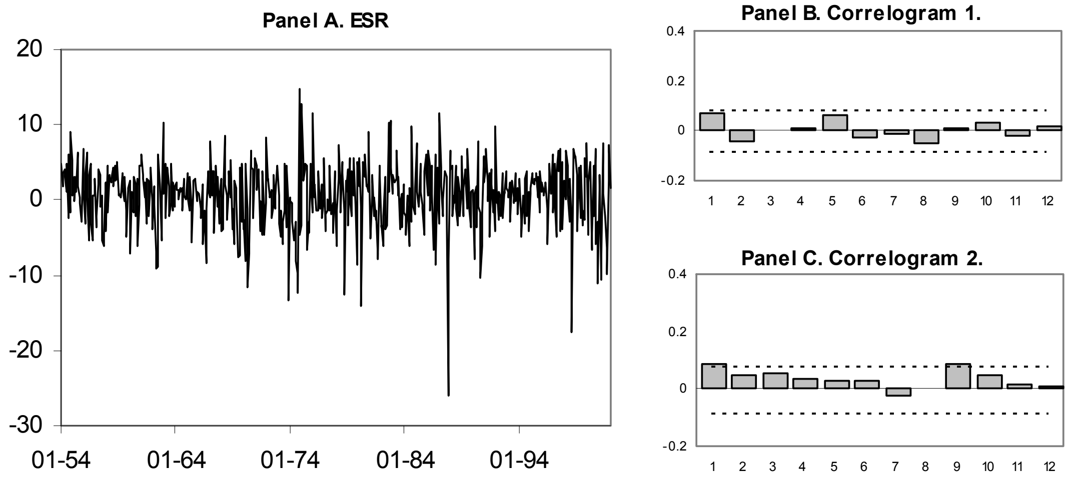

Figure 1. Time Series and Correlograms of Excess Stock Return. Panel A shows ESR's time series in monthly percentage. Panel B shows autocorrelation in the series' level from 1^{st} to 12^{th} lag jointly with 5% critical values (dotted lines). Panel C shows autocorrelation in the second power of demeaned series.

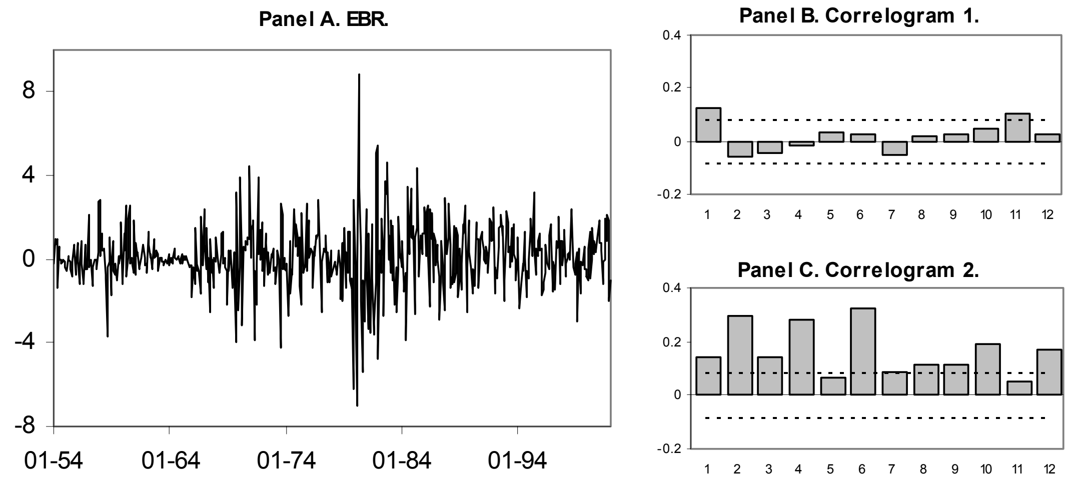

Figure 2. Time Series and Correlograms of Excess Bond Return. Panel A shows EBR's time series in monthly percentage. Panel B shows autocorrelation in series' level from 1^{st} to 12^{th} lag jointly with 5% critical values (dotted lines). Panel C shows autocorrelation in the second power of demeaned series.

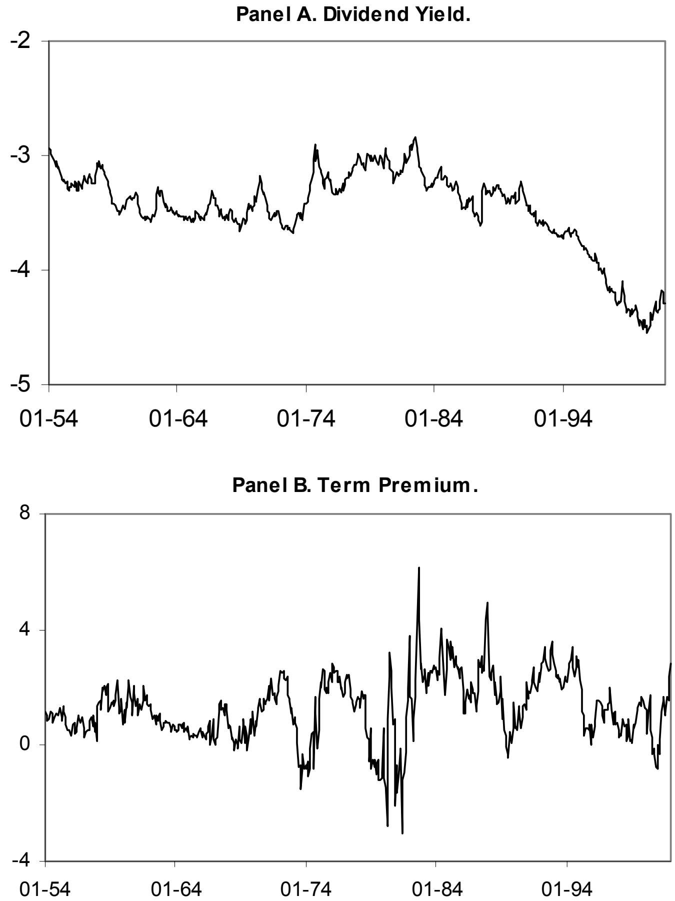

Figure 3. Time Series of Predictors. Panel A shows dividend yield's time series. Panel B shows term premium's time series.

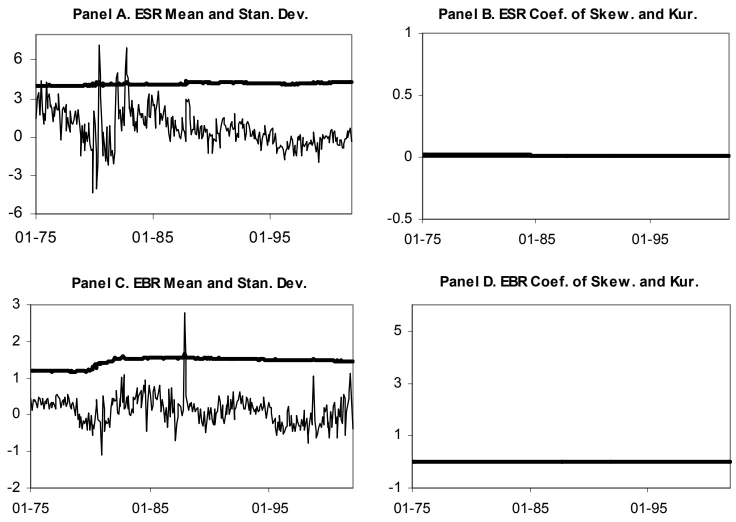

Figure 4. Density Forecasts from Gaussian VAR with Time-Varying Risk Premia. Panel A shows ESR's mean and standard deviation (thick line). Panel B shows ESR's coefficients of asymmetry and kurtosis (thick line). Panel C shows EBR's mean and standard deviation (thick line). Panel D shows EBR's coefficients of asymmetry and kurtosis (thick line).

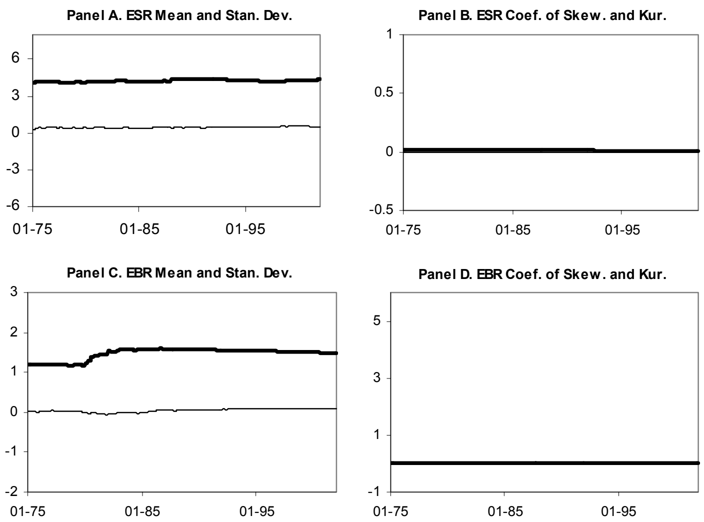

Figure 5. Density Forecasts from Gaussian VAR with Constant Risk Premia. Panel A shows ESR's mean and standard deviation (thick line). Panel B shows ESR's coefficients of asymmetry and kurtosis (thick line). Panel C shows EBR's mean and standard deviation (thick line). Panel D shows EBR's coefficients of asymmetry and kurtosis (thick line).

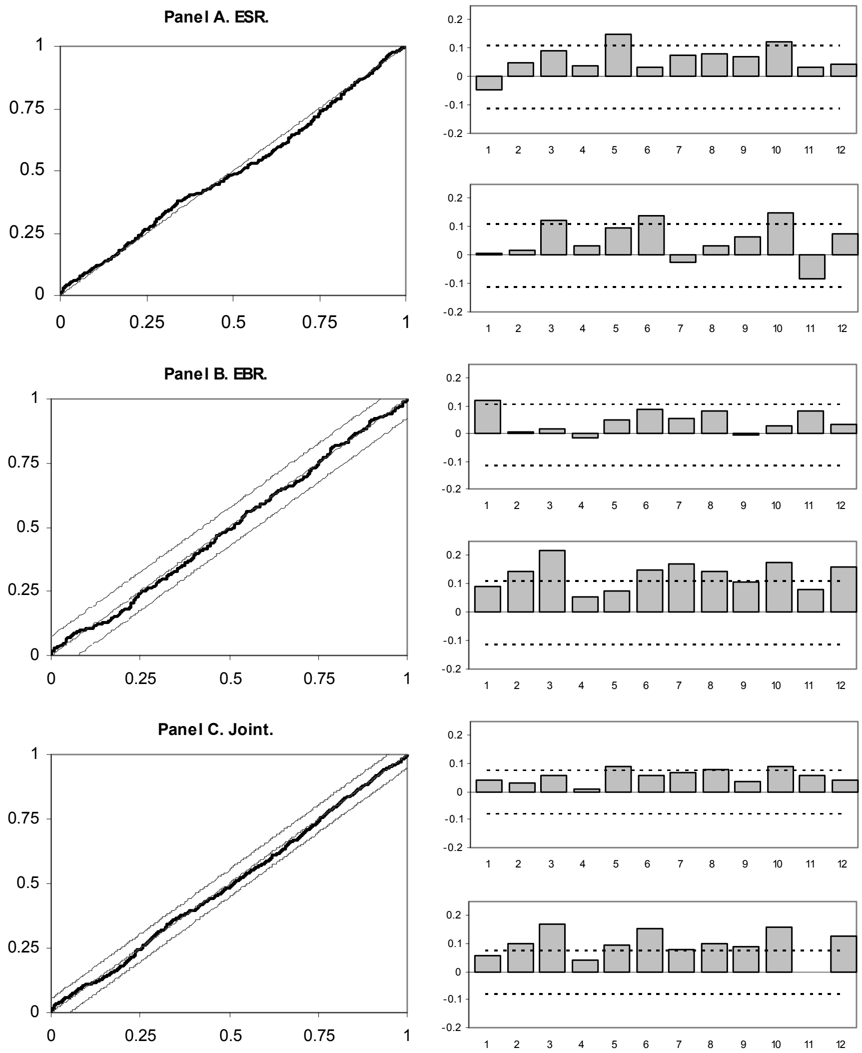

Figure 6. CDF and Correlograms of PITs from Gaussian VAR with Time-Varying Risk Premia. The left figure in each panel is a particular PIT's CDF (thick line) jointly with Kolmogorov-Smirnov 5% critical values. Both figures on the right show autocorrelation in level (top figure) and second power of demeaned PIT from 1^{st} to 12^{th} lag jointly with 5% critical values (dotted lines). Panel A shows those figures for ESR's PIT $d_{t+1}^{f,S}$, Panel B for EBR's PIT $d_{t+1}^{f,B}$, and Panel C for joint PIT $\left(d_{t+1}^{f,S}, d_{t+1}^{f,B|S}\right)$.

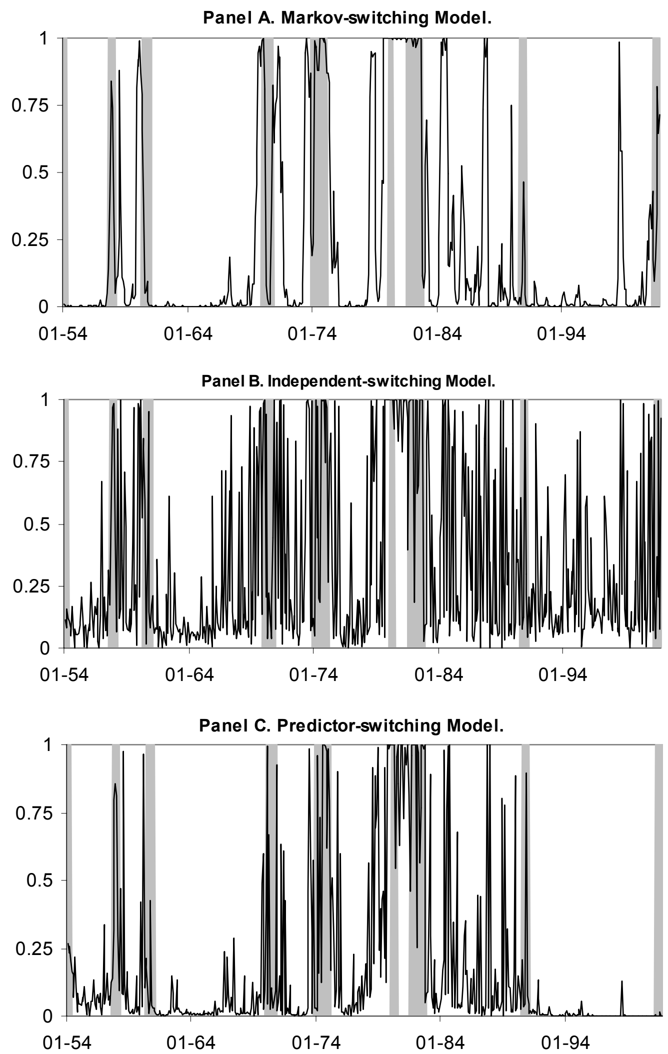

Figure 7. Probability of High-Volatility Regime and NBER Recession Index. Shaded regions are labelled as recessions by NBER. Probabilities are taken from last estimation, November 2001, of each model's time-varying risk premia version. Panel A shows probability from Markov-switching VAR, Panel B from Independent-switching VAR, and Panel C from Predictor-switching VAR.

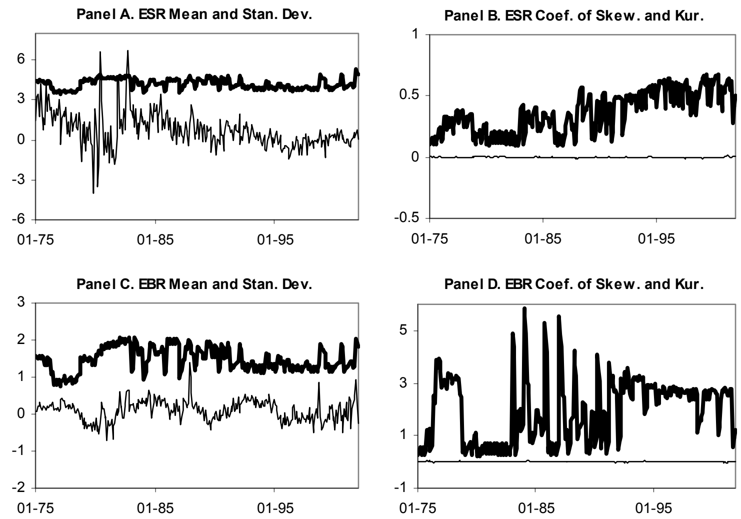

Figure 8. Density Forecasts from Markov-Switching VAR with Time-Varying Risk Premia. Panel A shows ESR's mean and standard deviation (thick line). Panel B shows ESR's coefficients of asymmetry and kurtosis (thick line). Panel C shows EBR's mean and standard deviation (thick line). Panel D shows EBR's coefficients of asymmetry and kurtosis (thick line).

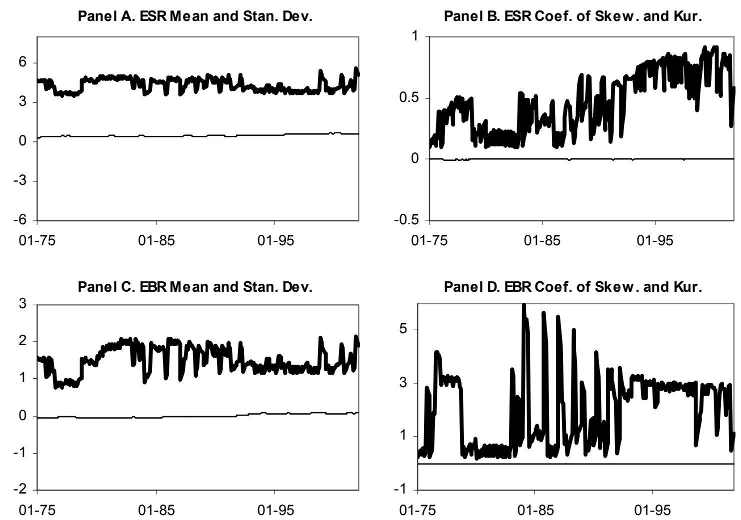

Figure 9. Density Forecasts from Markov-Switching VAR with Constant Risk Premia. Panel A shows ESR's mean and standard deviation (thick line). Panel B shows ESR's coefficients of asymmetry and kurtosis (thick line). Panel C shows EBR's mean and standard deviation (thick line). Panel D shows EBR's coefficients of asymmetry and kurtosis (thick line).

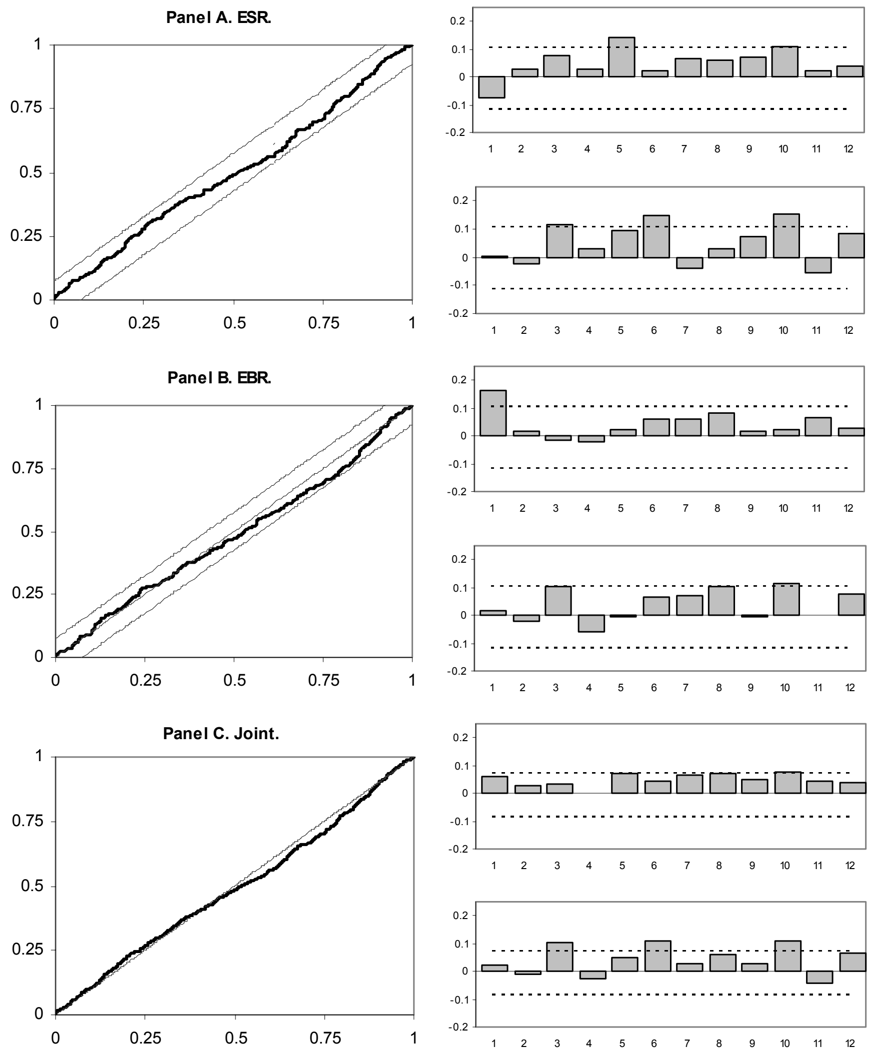

Figure 10. CDF and Correlograms of PITs from Markov-Switching VAR with Time-Varying Risk Premia. The left figure in each panel is a particular PIT's CDF (thick line) jointly with Kolmogorov-Smirnov 5% critical values. Both figures on the right show autocorrelation in level (top figure) and second power of demeaned PIT from 1^{st} to 12^{th} lag jointly with 5% critical values (dotted lines). Panel A shows those figures for ESR's PIT $d_{t+1}^{f,S}$, Panel B for EBR's PIT $d_{t+1}^{f,B}$, and Panel C for joint PIT $\left(d_{t+1}^{f,S}, d_{t+1}^{f,B|S}\right)$.

CEMFI WORKING PAPERS

0201 *Claudio Michelacci*: "Low returns in R&D due to the lack of entrepreneurial skills".

0202 *Jesús Carro and Pedro Mira*: "A dynamic model of contraceptive choice of Spanish couples".

0203 *Claudio Michelacci and Javier Suarez*: "Incomplete wage posting".

0204 *Gabriele Fiorentini, Enrique Sentana and Neil Shephard*: "Likelihood-based estimation of latent generalised ARCH structures".

0205 *Guillermo Caruana and Marco Celentani*: "Career concerns and contingent compensation".

0206 *Guillermo Caruana and Liran Einav*: "A theory of endogenous commitment".

0207 *Antonia Díaz, Josep Pijoan-Mas and José-Víctor Ríos-Rull*: "Precautionary savings and wealth distribution under habit formation preferences".

0208 *Rafael Repullo*: "Capital requirements, market power and risk-taking in banking".

0301 *Rubén Hernández-Murillo and Gerard Llobet*: "Patent licensing revisited: Heterogeneous firms and product differentiation".

0302 *Cristina Barceló*: "Housing tenure and labour mobility: A comparison across European countries".

0303 *Víctor López Pérez*: "Wage indexation and inflation persistence".

0304 *Jesús M. Carro*: "Estimating dynamic panel data discrete choice models with fixed effects".

0305 *Josep Pijoan-Mas*: "Pricing risk in economies with heterogenous agents and incomplete markets".

0306 *Gabriele Fiorentini, Enrique Sentana and Giorgio Calzolari*: "On the validity of the Jarque-Bera normality test in conditionally heteroskedastic dynamic regression models".

0307 *Samuel Bentolila and Juan F. Jimeno*: "Spanish unemployment: The end of the wild ride?".

0308 *Rafael Repullo and Javier Suarez*: "Loan pricing under Basel capital requirements".

0309 *Matt Klaeffling and Victor Lopez Perez*: "Inflation targets and the liquidity trap".

0310 *Manuel Arellano*: "Modelling optimal instrumental variables for dynamic panel data models".

0311 *Josep Pijoan-Mas*: "Precautionary savings or working longer hours?".

0312 *Meritxell Albertí, Ángel León and Gerard Llobet*: "Evaluation of a taxi sector reform: A real options approach".

0401 *Andres Almazan, Javier Suarez and Sheridan Titman*: "Stakeholders, transparency and capital structure".

0402 *Antonio Diez de los Rios*: "Exchange rate regimes, globalisation and the cost of capital in emerging markets".

0403 *Juan J. Dolado and Vanessa Llorens*: "Gender wage gaps by education in Spain: Glass floors vs. glass ceilings".

0404 *Sascha O. Becker, Samuel Bentolila, Ana Fernandes and Andrea Ichino*: "Job insecurity and children's emancipation".

0405 *Claudio Michelacci and David Lopez-Salido*: “Technology shocks and job flows”.

0406 *Samuel Bentolila, Claudio Michelacci and Javier Suarez*: “Social contacts and occupational choice”.

0407 *David A. Marshall and Edward Simpson Prescott*: “State-contingent bank regulation with unobserved actions and unobserved characteristics”.

0408 *Ana Fernandes*: “Knowledge, technology adoption and financial innovation”.

0409 *Enrique Sentana, Giorgio Calzolari and Gabriele Fiorentini*: “Indirect estimation of conditionally heteroskedastic factor models”.

0410 *Francisco Peñaranda and Enrique Sentana*: “Spanning tests in return and stochastic discount factor mean-variance frontiers: A unifying approach”.

0411 *F. Javier Mencía and Enrique Sentana*: “Estimation and testing of dynamic models with generalised hyperbolic innovations”.

0412 *Edward Simpson Prescott*: “Auditing and bank capital regulation”.

0413 *Víctor Aguirregabiria and Pedro Mira*: “Sequential estimation of dynamic discrete games”.

0414 *Kai-Uwe Kühn and Matilde Machado*: “Bilateral market power and vertical integration in the Spanish electricity spot market”.

0415 *Guillermo Caruana, Liran Einav and Daniel Quint*: “Multilateral bargaining with concession costs”.

0416 *David S. Evans and A. Jorge Padilla*: “Excessive prices: Using economics to define administrable legal rules”.

0417 *David S. Evans and A. Jorge Padilla*: “Designing antitrust rules for assessing unilateral practices: A neo-Chicago approach”.

0418 *Rafael Repullo*: “Policies for banking crises: A theoretical framework”.

0419 *Francisco Peñaranda*: “Are vector autoregressions and accurate model for dynamic asset allocation?”