Hierarchical Markov Normal Mixture Models with Applications to Financial Asset Returns

John Geweke[a]*and Giovanni Amisano[b]
[a]*Departments of Economcs and Statistics, University of Iowa, USA*
[b]*European Central Bank, Frankfurt, Germany*
and University of Brescia, Brescia, Italy

July, 2007

Abstract

Motivated by the common problem of constructing predictive distributions for daily asset returns over horizons of one to several trading days, this article introduces a new model for time series. This model is a generalization of the Markov normal mixture model in which the mixture components are themselves normal mixtures, and it is a specific case of an artificial neural network model with two hidden layers. The article characterizes the implications of the model for time series in two ways. First, it derives the restrictions placed on the autocovariance function and linear representation of integer powers of the time series in terms of the number of components in the mixture and the roots of the Markov process. Second, it uses the prior predictive distribution of the model to study the implications of the model for some interesting functions of asset returns. The article uses the model to construct predictive distributions of daily S&P 500 returns 1971-2005, US dollar – UK pound returns 1972-1998, and one- and ten-year maturity bonds 1987-2006. It compares the performance of the model for these returns with ARCH and stochastic volatility models using the predictive likelihood function. The model's performance is about the same as its competitors for the bond returns, better than its competitors for the S&P 500 returns, and much better than its competitors for the dollar-pound returns. In- and out-of-sample validation exercises with predictive distributions identify some remaining deficiencies in the model and suggest potential improvements. The article concludes by using the model to form predictive distributions of one- to ten-day returns during volatile episodes for the S&P 500, dollar-pound and bond return series.

*Correspondence to: John Geweke, Department of Economics, University of Iowa, Iowa City IA 52242, US. E-mail: john-geweke@uiowa.edu

The conditional distributions of future asset returns are important in financial markets, being central to market outcomes like the pricing of derivatives and to reporting tasks like assessing value at risk. This importance is reflected in the applied econometrics literature, in which literally hundreds of articles have been devoted to the application of models to asset returns in scores of financial markets. Many of these studies use the autoregressive conditionally heteroskedastic (ARCH) family of models introduced by Engle (1983) and Bollerslev (1986). Other studies use stochastic volatility (SV) models first developed by Taylor (1986), both in continuous time (Lo (1988)) and discrete time (Jacquier et al. (1994)). A few more have taken an explicitly seminonparametric (SNP) approach (Gallant and Tauchen (1989)).

Asset returns are among the best data in econometrics. They are available daily for many assets and studies that use trade-by-trade data are increasingly common. They are records of market outcomes, and because errors in price records would always have negative consequences for one party to a trade they are very accurate compared with most other economic data. Studies using thousands of observations on asset returns are not uncommon. Section 1.1 introduces four examples of daily returns, used in this article: 35 years of daily returns on the Standard and Poors (S&P) 500 index for U.S. equity markets, over 25 years of daily returns on the dollar-pound exchange rate, and 20 years of daily returns on one- and ten-year maturity bonds.

There is no theoretically compelling parametric model of asset returns, or even a set of such models. (A theoretically compelling model is not the same as one that is convenient, be it for purposes of inference or for pricing derivatives.) Given this fact and the richness of asset return data, the dominance of models with a half-dozen or fewer parameters in the applied econometric asset return literature, and the relative dearth of alternatives like SNP models, is puzzling. The premise of our undertaking is that careful application of models that impose very weak restrictions on asset return dynamics should yield conditional distributions of asset returns that are superior to those using tightly parameterized models, and that premise is confirmed in this article.

In any applied econometric undertaking, the characteristics of the data and the technology to be used for inference are important, if not critical, factors in the choice of model. Section 1.1 describes some of the important characteristics of the asset return data studied here, including those that are the most challenging for the econometrician. Our inference technology is subjective Bayesian using Markov chain Monte Carlo (BMCMC) to recover the posterior distributions from which conditional distributions of asset returns are then constructed. Section 1.2 briefly reviews the aspects of this methodology most important to the conditional distribution modeling exercise.

Our approach in creating a model was to think about generalizations of parametric models that would be natural for BMCMC. We began with Markov normal mixture models, sometimes known as hidden Markov models, which have long been established in statistics (Lindgren (1978); Tyssedal and Tjøstheim (1988)) and econometrics (Chib (1996); Ryden et al. (1998)), but have some clear limitations in modeling asset returns – for example, the combination of absence of serial correlation in conjunction with conditional heteroskedasticity requires large models and the absence of serial correlation is difficult to impose. This led us to a generalization of these models in which the components mixed are non-Gaussian and are themselves mixtures of normal distributions: the hierarchical Markov normal mixture

(HMNM) model. Section 2.1 describes this model which, it turns out, can also be regarded as a restricted Markov normal mixture (MNM) model of much higher order, or as a specific case of an artificial neural network model with two hidden layers (Kuan and White (1992)). The HMNM model is parametric and places restrictions on moments, and Section 2.2 provides some compact theoretical characterizations of these restrictions.

The BMCMC inference technology has implications for applied econometric research that have emerged slowly, and undoubtedly are not yet fully appreciated (at least by us). An important secondary objective of this article is to emphasize these implications and illustrate them in the context of the primary objective of constructing conditional distributions of future asset returns using the HMNM model. One implication is that the properties of the model, including its strengths and limitations in the application at hand, can be well-understood using prior predictive distributions, described in Section 1.2 and applied to our asset return data in Section 2.3. This can be done before setting about the generally much more time-consuming task of formal inference, and as a by-product provides a systematic way to develop substantive, subjective prior distributions for a large number of parameters. This exercise, which can be undertaken for any approach to inference, not just BMCMC, can greatly enhance the productivity of applied econometric research because it can identify poorly specified models before rather than after most of the work is done. Section 2.4 summarizes the BMCMC posterior simulator and documents its computational efficiency, which enables us to carry out in a few days the exercises with thousands of posterior distributions, on which many of the results in the balance of the article are based, that would be conducted over decades in actual applications. Moreover, we are able to undertake this entire exercise using quite a few variants of the HMNM model.

Section 3 evaluates these HMNM models from three perspectives. The first perspective is comparison with some competitors – several ARCH models and a SV model. The comparison is based on predictive likelihood ratios, described in Section 3.1. These ratios have the same theoretically compelling grounding as Bayes factors, but do not have the latter's sensitivity to the prior distribution. The short summary is that for the bond return series the HMNM model performs as well as the strongest competitor, for S&P returns it performs significantly better, and for the dollar-pound return series it performs overwhelmingly better. The concluding section provides a more detailed summary, and the full details are in Section 3.2. That section also makes some less systematic comparisons with affine models and a wider variety of SV models for the S&P 500 return series, and the comparison again favors the HMNM models.

The second approach to evaluation (Section 3.3) summarizes the characteristics of conditional distributions for horizons of one to ten days based on the HMNM model posterior distributions simulated for each trading day for each asset return. In a well-specified model the c.d.f.s of these distributions, evaluated at realized returns, should be uniformly distributed and those for one-day horizons should be independent. The uniform distribution turns out to be an accurate description for one-day horizons, but deteriorates as the horizon lengthens. The one-day c.d.f.s clearly exhibit positive autocorrelation.

The final approach to evaluation (Section 3.4) utilizes the posterior predictive distributions of the characteristics of asset returns identified in Section 1.1. The posterior predictive distribution answers the question, if asset returns were governed by the HMNM, what would

be the values of these characteristics? For most characteristics, asset return series and time periods the answer is consistent with the observed characteristics. However, for the S&P 500 returns and several time periods the posterior distribution of the HMNM model predicts markedly less serial correlation and smaller leverage effects than actually observed, and it cannot account well for the long-memory characteristics of S&P 500 absolute returns in the early 1970s.

Section 4 illustrates how the model and BMCMC approach to inference could be used to construct predictive distributions over one- to ten-day horizons. It focuses on periods of greatest volatility in the asset return series examined, showing how conditional distributions react to strong movements in returns, and studying how they readjust as the volatile period passes. The last section of the article integrates all of its findings to draw conclusions about how the HMNM model of asset returns might be improved, and about the practice of applied econometrics using BMCMC methods.

1 The setting

Modeling asset returns is a central task in risk management in the financial sectors of developed economies. Models provide conditional probability distributions of returns and these distributions, in turn, are important components in decision-making. The universe of models and data sets considered is invariably subjective, and for this if no other reason the probability distributions of returns are subjective. While this is true in all applied econometrics in support of decision-making, the issue in asset returns and risk management is especially critical because (1) probability distributions of extreme events are major drivers of decisions at the heart of the economy, and (2) evidence (i.e., data) about extreme events is, by definition, limited. A focus of this study is the careful derivation of subjective distributions of asset returns from explicit modeling assumptions and evidence provided by data. In particular, it derives the subjective probabilities of extreme returns as they unfold day-by-day in financial crises (Section 4.)

It has long been well understood that the simplest models fail notoriously as descriptions of extreme events. To take the simplest example, the mean of Standard and Poors 500 daily returns (expressed as 100 times log ratio of successive trading day closing prices, the daily percent log return) for the twentieth century is 0.0157 and the standard deviation is 1.036. The lowest daily return, on October 19, 1987, was -22.90, and the associated z-score is -22.12. If returns are assumed to be i.i.d. normal, the probability of a daily return this low, or lower, is 1.031×10^{-108}, and the expected time to recurrence of such an event is 3.59×10^{105} years. (Modern cosmology puts the age of the universe at less than 1.5×10^{10} years.)

While all decision-making is subjective, effective decision-making demands that the models used be well calibrated: that is, if (as is generally the case) the models are predicated on the assumption that the unconditional distributions of past and future events are the same, then predictions of future events should not indicate radically different behavior than has been observed in the past. Section 1.1 recapitulates some key characteristics of asset returns that have been identified in the literature, and shows how they are manifest in the data used in this study. Section 1.2 briefly outlines the approach to subjective distributions

of asset returns.

1.1 Asset returns

The financial decision-making problems motivating this study arise in several contexts and for many different kinds of assets. The focus of the applied work here is on three time series of asset returns, together representative of the returns to which much of the empirical work in the literature has been addressed. In each case the underlying data are closing prices of assets p_t on trading days $t = 1, 2, \ldots$, and the percent log return for day t is $y_t = 100 \cdot \log(p_t/p_{t-1})$.

1. Standard and Poors 500 daily returns, all trading days for 1971 through 2005. The return series y_t is obtained from Wharton Research Data Services.

2. The foreign exchange rate daily returns for the UK pound and the US dollar, all trading days from 1972 through 1998. The price series p_t is obtained from Wharton Research Data Services. A positive return indicates a rise in the value of the dollar relative to the pound.

3. Daily returns on zero coupon bonds, one- and ten-year maturity, all trading days from December 2, 1987 through February 2, 2007. The data source is the Bank for International Settlements (BIS). Since 1996, participating central banks have been reporting their estimates of the zero coupon structure to the BIS. Therefore these data represent an "official" source of information for zero coupon rates. In this way it is possible to abstract from coupon effects and mimic portfolio strategies involving bond instruments. The procedure used to obtain zero coupon rates, as detailed in BIS (2005, section 1), is the Svensson (1994) extension to the Nelson and Siegel (1987) approach which postulates a parametric form for the discount function being used to back out the zero coupon data. Letting $r_t^{(n)}$ the market rate at t of the zero coupon bond expiring at $t+n$ and selling at par, returns on n-maturity $y_t^{(n)}$are computed from prices with the usual formula

$$y_t^{(n)} = 100 \cdot \log\left(p_t^{(n)}/p_{t-1}^{(n)}\right), \;\; p_t^{(n)} = \left(1 + r_t^{(n)}\right)^{-n}.$$

   Subsequently we refer to these series as one-year and ten-year bond returns.

1.1.1 Some functions of the data

For each of these returns the time series is long, and can be summarized by some functions of the series. Subsequently we use these functions in the assessment of model adequacy and calibration by means of prior predictive distributions (Section 2.3) and posterior predictive distributions (Section 3.4). There are 16 functions of interest, organized into four groups. For each asset return series, observed values of the function are organized by five year intervals, as indicated in Tables 1 and 2. Abbreviated descriptions of the functions are in the column headings, where y^n denotes non-overlapping n-day returns.

- Four of the functions are moments, tabulated in Table 1, columns (1)-(4): the coefficient of skewness in one-day returns, the coefficient of excess kurtosis in one-day returns, the coefficient of skewness in ten-day returns, and the coefficient of excess kurtosis in ten-day returns.

- Four of the functions are indicators of the dynamics of returns and are indicated in Table 1 columns (5)-(8): the first-order autocorrelation coefficient for one-, ten-, and fifty-day non-overlapping returns, and the sum of absolute values of the first 200 autocorrelation coefficients.

- Table 2 indicates four aspects of the short-term dynamics of absolute returns in columns (1)-(4): first- and ninth-order autocorrelation coefficients of one-day returns, the ratio of the variance of ten-day returns to one-day returns, and the sample correlation coefficient of y_t and $|y_{t+1}|$, sometimes termed "leverage."

- The last four functions in Table 2 measure some aspects of the long-term dynamics of absolute returns: the GPH (Geweke and Porter-Hudak (1984)) estimate $\widehat{d}$ of the long-memory parameter in column (5), and for the sample autocorrelation function $\widehat{\rho}_i$ of absolute returns $|y_t|$ the functions indicated in columns (6), (7) and (8).

1.1.2 Some challenging characteristics of asset returns

It has long been widely understood that asset return time series are leptokurtic and display substantial persistence in moments of absolute returns. These characteristics are clear in these tables. The function values convey other interesting characteristics if financial asset return series, most not as widely recognized as leptokurtosis and persistence in variance.

1. There is substantial variation in the sample skewness and excess kurtosis coefficients from one five-year period to another, relative to what might be expected from an i.i.d. process (Table 1, columns (1)-(4)). The most pronounced instance is the S&P 500 returns during 1986-90, and there are examples for the other return series as well.

2. The coefficient of skewness for ten-day returns exceeds that for one-day returns in 10 of the 18 cases tabulated (Table 1, columns (1) and (3)). This is in sharp contrast to the behavior of the population coefficient of skewness of an i.i.d. process with finite sixth moment.

3. Using Gaussian i.i.d. asymptotic sampling theory the first order autocorrelation coefficient for returns (Table 1, column (5)) is more than two standard deviations from zero in 8 of the 18 cases. First order autocorrelation coefficients for non-overlapping 10- and 50-day returns (columns (6) and (7)) are often higher in absolute value, but none of them is more than two standard deviations from zero using Gaussian i.i.d. asymptotic sampling theory.

4. Consistent with persistence in volatility, one-day absolute returns exhibit positive first order autocorrelation (Table 2, column (1)). In 6 of the 18 cases, however, the first order autocorrelation for returns (Table 1, column (5)) is even larger in absolute value.

5. The slow decay of the autocorrelation function of absolute stock index returns has been widely noted (Ding et al. (1993); Ryden et al. (1998)**).** That is evident for these return series and time periods as well. The ninth-order autocorrelation coefficient for absolute returns (Table 2, column (2)) is always positive, and it exceeds the first-order autocorrelation coefficient (column (1)) in half of the cases.

6. In two-thirds of the cases the sample variance in the ten-day return is more than 10 times the sample variance in the one-day return (Table 2, column (3)), its expected value in a benchmark i.i.d. model with finite variance.

7. For returns, the sum of absolute values of the first 200 autocorrelation coefficients is about the same in all 18 cases (Table 1, column (8)). For absolute returns (Table 2, column (6)) it varies substantially both across return series and across time periods within series.

8. Sample autocorrelation coefficients for absolute returns show a strong propensity to be positive rather than negative (Table 2, column (8)). In 3 of the 18 cases the first 200 coefficients are all positive. At the other extreme, the first 200 autocorrelation function coefficients of absolute dollar-pound returns in 1987-91 sum nearly to zero.

9. Absolute returns display an apparent long-memory property as indicated by the GPH estimate $\widehat{d}$ of the long memory parameter d (Table 2, column (5)). Conventional non-Bayesian hypotheses tests reject $d = 0$ in favor of $d > 0$ for most series and time periods, and the point estimate indicates nonstationarity in four cases. The statistic $\widehat{d}$ differs substantially from one time period to the next in the case of the S&P 500 returns, as Granger and Ding (1996) have noted.

10. The widely documented leverage effect — negative sample correlation of y_t with $|y_{t+1}|$ — is evident for equity returns as expected (Table 2, column (4)). The leverage interpretation of this sample moment does not extend to foreign exchange and bond returns, and these cases display a mix of positive and negative sample correlations of $|y_t|$ with y_{t+1}, all small in absolute value.

The characteristics of asset return series in Tables 1 and 2 are challenging for model building. As is generally recognized in the literature, a well-calibrated model must be capable of reproducing persistence in volatility, extreme non-Gaussian sampling distributions, and details critical for financial decision-make such as the leverage effect. It must be sufficiently flexible that some of these characteristics are displayed in all cases (e.g., persistence in volatility) while others may or may not be displayed (e.g., the leverage effect). At least as challenging is the variation in many of these sample characteristics from one five-year period to another, for the same time series of asset returns. The model introduced in this article, in Section 3.1, attempts to encompass all of these characteristics.

1.2 Bayesian inference and modeling

This study uses subjective Bayesian methods for inference, model evaluation and decision-making, using simulation to express the requisite distributions. Most of the article is taken

up with the specifics of the model and its application to the data and decision problem discussed in Section 1.1. This section establishes the framework of the procedures used as well as some essential terminology.

The formal motivation for the study is a setting in which decision are made based on an observable history of daily asset returns $\mathbf{y}_T = (y_1, \ldots, y_T)'$. The decision is made based on a loss function $L(\mathbf{a}, \boldsymbol{\omega})$ in which the unobservable vector of interest is $\boldsymbol{\omega} = (y_{T+1}, \ldots, y_{T+f})$, and the action is $\widehat{\mathbf{a}} = \arg\min E[L(\mathbf{a}, \boldsymbol{\omega})]$. The decision could be portfolio allocation, determination of value at risk, or a host of similar tasks related to short-term financial management; subsequent applications in this article concentrate on value at risk.

In any problem with this structure the technical task is finding the appropriate distribution of $\boldsymbol{\omega}$ and representing it in a manner suitable to the determination of $\widehat{\mathbf{a}}$. The distribution of $\boldsymbol{\omega}$ conditions on the information available at the end of period T, taken here to be the observed asset returns $\mathbf{y}_T^o = (y_1^o, \ldots, y_T^o)$ and a model (designated A, for "assumptions") that is capable of producing the distribution $\boldsymbol{\omega}$ conditional on $\mathbf{y}_T^o$. The distinction between the asset returns before they are observed ($\mathbf{y}_T$, a random vector) and the asset returns after they are observed ($\mathbf{y}_T^o$, a vector of constants usually termed data) is central to the concepts used both in the theory and the applications in this study.

The model A provides $p(\boldsymbol{\omega} \mid \mathbf{y}_T^o, A)$ by positing a family of conditional distributions

$$p(y_t \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}_A, A) \quad (t = 1, 2, \ldots) \tag{1}$$

in which $\mathbf{y}_{t-1} = (y_1, \ldots, y_{t-1})'$ and $\boldsymbol{\theta}_A$ is a vector of unobservables, including latent variables as well as parameters. The distributions (1) used in this study are introduced in Section 2.1; both latent variables and potentially large number of parameters are important in these models.

Because $\boldsymbol{\omega} = (y_{T+1}, \ldots, y_{T+f})$, the conditional distribution $p(\boldsymbol{\omega} \mid \mathbf{y}_T^o, \boldsymbol{\theta}_A, A)$ is implicit in (1) and it is generally accessible by means of simulations

$$y_{T+s}^{(m)} \sim p\left(y_{T+s} \mid y_1^o, \ldots, y_T^o, y_{T+1}^{(m)}, \ldots, y_{T+s-1}^{(m)}, \boldsymbol{\theta}_A, A\right) \quad (s = 1, \ldots, f) \tag{2}$$

for $m = 1, \ldots, M$. In order to move from $p(\boldsymbol{\omega} \mid \mathbf{y}_T^o, \boldsymbol{\theta}_A, A)$ to $p(\boldsymbol{\omega} \mid \mathbf{y}_T^o, A)$ the model must also provide the posterior density $p(\boldsymbol{\theta}_A \mid \mathbf{y}_T^o, A)$ and make it accessible. Since $p(\mathbf{y}_T \mid \boldsymbol{\theta}_A, A)$ is available from (1), the missing component essential to the posterior distribution is the prior distribution $p(\boldsymbol{\theta}_A \mid A)$; such a prior distribution is implicit in any representation $p(\boldsymbol{\omega} \mid \mathbf{y}_T^o, A)$ and, indeed, in the action $\widehat{\mathbf{a}}$. Section 2.1 introduces the prior distributions used in this study. These provide prior densities of the form $p(\boldsymbol{\theta}_A \mid \boldsymbol{\lambda}_A, A)$, $\boldsymbol{\lambda}_A$ being a vector of fixed hyperparameters specified before proceeding. (To reduce notational clutter we will often absorb $\boldsymbol{\lambda}_A$ into A, which denotes the entire specification of the model.)

If it is possible to simulate from the posterior density

$$\boldsymbol{\theta}_A^{(m)} \sim p(\boldsymbol{\theta}_A \mid \mathbf{y}_T^o, A) \propto p(\boldsymbol{\theta}_A \mid A)\, p(\mathbf{y}_T^o \mid \boldsymbol{\theta}_A, A) \tag{3}$$

for $m = 1, \ldots, M$, then it is possible to simulate $\boldsymbol{\omega}^{(m)} \sim p(\boldsymbol{\omega} \mid \mathbf{y}_T^o, A)$. Iteration m of the simulator consists of (3) followed by (2) with $\boldsymbol{\theta}_A^{(m)}$ in place of $\boldsymbol{\theta}_A$. While (2) is generally straightforward, the posterior simulator (3) is not. Section 2.4 outlines the posterior simulator used, with details relegated to an on-line appendix.

In summary, the model A has three components: a prior distribution $\boldsymbol{\theta}_A \mid A$ for unobservable entities, including latent variables as well as parameters; a distribution $\mathbf{y}_T \mid (\boldsymbol{\theta}_A, A)$ of observable asset returns conditional on the unobservables $\boldsymbol{\theta}_A$; and a conditional distribution $\boldsymbol{\omega} \mid (\mathbf{y}_T, \boldsymbol{\theta}_A, A)$ of the vector of interest. The posterior density in (3) is implicit in the first two of these three, and given a posterior simulator the distribution $\boldsymbol{\omega} \mid (\mathbf{y}_T^o, A)$ relevant for decision-making is accessible.

These procedures are essential to decision-making, but they are also entirely formal. They do not directly address to following kinds of questions, which are essential to good applied econometrics.

1. What is a reasonable specification of the prior distribution $\boldsymbol{\theta}_A \mid A$? In particular, how should the vector of hyperparameters $\boldsymbol{\lambda}_A$ in the prior density $p(\boldsymbol{\theta}_A \mid \boldsymbol{\lambda}_A, A)$ be chosen?

2. Most models are sufficiently complicated that their implications for observable outcomes of interest $\mathbf{y}_T$ are not immediately apparent. Is it possible to understand these implications in reasonably short order, and in particular before constructing the posterior simulator (3), which is usually the most resource-intensive part making $\boldsymbol{\omega} \mid (\mathbf{y}_t^o, A)$ accessible?

3. Is the model capable of accounting for important summary properties of the data $\mathbf{y}_T^o$?

The first question arises generally in applying Bayesian methods. The second two questions fall under the general heading "specification analysis," and are important in good applied econometric work regardless of approach. In many of these approaches specification analysis must be undertaken following inference, "specification tests" usually involving estimates in one or more models. If inference is time and resource consuming, as is often the case, this presents a practical problem: the work must be completed before the applied econometrician starts to get good answers to these three questions.

A complete model changes these circumstances in important ways that work to the advantage of the applied econometrician, however. The model A provides a probability density for the observable $\mathbf{y}_T$,

$$p(\mathbf{y}_T \mid A) = \int_{\Theta_A} p(\boldsymbol{\theta}_A \mid A)\, p(\mathbf{y}_T \mid \boldsymbol{\theta}_A, A)\, d\boldsymbol{\theta}_A,$$

the prior predictive distribution (Box (1980); Lancaster (2004), Section 2.4.2; Geweke (2005), Section 8.3.1). This distribution can be accessed by means of the simulator

$$\boldsymbol{\theta}_A^{(m)} \sim p(\boldsymbol{\theta}_A \mid \boldsymbol{\lambda}_A, A), \;\; \mathbf{y}_T^{(m)} \sim p\left(\mathbf{y}_T^{(m)} \mid \boldsymbol{\theta}_A, A\right). \tag{4}$$

for $m = 1, \ldots, M$. Since the model provides a prior predictive distribution for $\mathbf{y}_T$, it provides such a distribution for any function of $\mathbf{y}_T$, including the functions of interest $z_j = h_j\left(\mathbf{y}_T^{(m)}\right)$ described in Section 1.1.1. These distributions are immediately accessible by means of the subsidiary computations $z_j^{(m)} = h_j\left(\mathbf{y}_T^{(m)}\right)$ $(m = 1, \ldots, M)$.

The prior predictive distribution provides a strategy for addressing all three of the questions posed above. For any given setting of the prior hyperparameter vector $\boldsymbol{\lambda}_A$, the prior predictive simulator (4) provides the distribution of the selected observable outcomes of interest z_j, thereby addressing question 2. As illustrated in Section 1.1.1 the functions h_j are quite general and can be constructed to address issues like stability across time periods and the relationship between single- and multi-day returns.

The distributions $z_j \mid (\boldsymbol{\lambda}_A, A)$ enable the applied econometrician to evaluate alternative settings of the prior hyperparameter vector $\boldsymbol{\lambda}_A$ in terms of the prior distribution they induce on important characteristics of asset returns. This exercise addresses question 1, which can be a significant barrier to entry for applied Bayesian work: hyperparameters of prior distributions are often driven by technical considerations, whereas what constitutes a reasonable model – and this includes the prior distribution – is much more readily conceived in terms of the prior distributions of the characteristics z_j.

Such exercises with prior predictive distributions can substantially increase one's understanding of the model, before incurring the costs of constructing a posterior simulator or even, for that matter, collecting data. The immediate outcome may be negative – that is, the applied econometrician may conclude that no setting of the prior hyperparameter vector $\boldsymbol{\lambda}_A$ expresses a suitable prior distribution for characteristics z_j. More important, the exercise may indicate why the model is incapable of expressing prior beliefs for the z_j. This, in turn, can spur consideration of new or alternative models, without incurring a large commitment to the failed model – a positive longer-run outcome.

The prior predictive distribution for single characteristic z_j is easy to represent graphically for one or for two characteristics at a time. Question 3 may then be addressed by indicating, on the same graphs, the observed value of the characteristics z_j^o, and this can be done for several time series of financial asset returns. Section 2.3 provides examples. Using standard smoothing methods it is possible to approximate densities of the form $p\left(z_j^o \mid A\right)$ or $p\left(z_i^o, z_j^o \mid A\right)$, but this level of formality is usually not required when the prior predictive distribution is used as described here – to gain insight into the implications of the model.

2 The model

The model proposed in this study is motivated by several considerations. First, in financial market decision-making the full distribution of asset returns, as opposed to a few moments, is critical. Second, asset return data is plentiful and of very high quality, relative to other economic data. Third, there is no theoretically compelling tightly parameterized model for the distribution of returns that has proved competitive in forecasting. These facts strongly suggest consideration of models for asset returns that are inherently flexible, and suggest that there is no reason to be concerned about large numbers of parameters *per se*.

Mixture models provide flexibility in describing distributions and are particularly useful in conjunction with Bayesian inference by means of posterior simulation. In the simple normal mixture model (e.g. MacLachlan and Peel (2000**,** Chapter 3), Geweke (2005, Section 6.4)) the observable variable of interest y_t occupies one of m discrete latent states denoted $s_t \in \{1, \ldots, m\}$:

$$y_t \mid (\mathbf{x}_t, s_t = i) \sim N(\boldsymbol{\beta}'\mathbf{x}_t + \phi_i, h^{-1} \cdot h_j^{-1}).$$

The covariate vector $\mathbf{x}_t$ is deterministic, and in this study there is a single covariate $x_t = 1$. The process $\{s_t\}$ in the simple normal mixture model is i.i.d. with $P(s_t = i) = p_i$ $(i = 1, \ldots, m)$, and $\sum_{i=1}^{m} p_i = 1$. The precision parameters h, on the one hand, and $h_1, \ldots, h_m$, on the other, are identified only up to a factor of proportionality. This lack of identification turns out to be useful in specifying the prior, to which we return in Section 2.1.2. The parameters β_1 and $\phi_1, \ldots, \phi_m$ are identified by requiring $E(\varepsilon_t) = 0$, equivalent to $\mathbf{p}'\boldsymbol{\phi} = 0$ where $\mathbf{p}' = (p_1, \ldots, p_m)$ and $\boldsymbol{\phi}' = (\phi_1, \ldots, \phi_m)'$.

An extension of this model widely used with time series is the Markov normal mixture model (Lindgren (1978)), which maintains the same structure except that the latent states s_t are not independent. Instead

$$P\ [s_t = j \mid s_{t-1} = i,\ s_{t-u}\ (u > 1)] = p_{ij} \quad (i = 1, \ldots, m;\ j = 1, \ldots, m)\,. \tag{5}$$

Thus s_t evolves as a first order, m-state Markov chain, characterized by the $m \times m$ transition matrix $\mathbf{P} = [p_{ij}]$. This matrix, in turn accounts for the serial dependence properties of Markov normal mixture models. Several characteristics of $\mathbf{P}$ are important subsequently in developing some properties of the model proposed in this study.

1. In view of the fact that $\sum_{j=1}^{m} p_{ij} = 1$ $(i = 1, \ldots, m)$, at least one of the eigenvalues of $\mathbf{P}$ is 1; and since $\mathrm{P}(s_{t+u} = j \mid s_t = i) = [\mathbf{P}^u]_{ij}$, no eigenvalue can exceed one in modulus.

2. If all $m - 1$ remaining eigenvalues of $\mathbf{P}$ have modulus strictly less than one, $\mathbf{P}$ is said to be irreducible. There is then a unique vector $\boldsymbol{\pi}$ with the properties $\boldsymbol{\pi}'\mathbf{P} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}'\mathbf{e}_n = 1$. (Here, and throughout, $\mathbf{e}_n$ denotes an $n \times 1$ vector of units.) It can be shown that $\pi_i \in (0, 1)$ $(i = 1, \ldots, m)$.

3. A transition matrix that has two or more real eigenvalues of unit modulus is reducible: that is, it is impossible to move between certain states no matter how many periods elapse. If there are complex eigenvalues of modulus unity they must occur in complex pairs, and such a transition matrix renders the process s_t strictly periodic. Simple examples of reducible and periodic transition matrices are

$$\mathbf{P} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 0.7 & 0.3 \\ 0.0 & 0.4 & 0.6 \end{bmatrix} \quad \text{and} \quad \mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix},$$

   respectively. Neither property is reasonable for the latent state s_t in these models, and the prior distributions introduced in Section 2.1.2 assign probability zero to such transition matrices $\mathbf{P}$.

Given (5), if $P(s_t = i) = p_{ti}$ and $\mathbf{p}_t = (p_{t1}, \ldots, p_{tm})'$, then $\mathbf{p}'_{t+u} = \mathbf{p}'_t\mathbf{P}^u$. If $\mathbf{P}$ is irreducible and $\mathbf{p}_1 = \boldsymbol{\pi}$, then $\mathbf{p}_t = \boldsymbol{\pi}$ for all $t > 0$; $\boldsymbol{\pi}$ is the invariant distribution of s_t. In the special case $\mathbf{P} = \mathbf{e}_m\boldsymbol{\pi}'$, the states s_t are serially independent. If $\varepsilon_t = y_t - \boldsymbol{\beta}'\mathbf{x}_t$ is stationary then $P(s_t = i) = \boldsymbol{\pi}_i$ and $E(\varepsilon_t) = 0$ is equivalent to $\boldsymbol{\pi}'\boldsymbol{\phi} = \mathbf{0}$.

This section combines the Markov normal mixture and normal mixture models in a new model, the hierarchical Markov normal mixture (HMNM) model. As described in Section

2.1, this model can be interpreted as either a generalization of a smaller Markov mixture model or a restriction of a much larger Markov mixture model. This extension accommodates several characteristics of interest in financial asset return modeling in a tidy and practical way relative to the conventional Markov normal mixture model. Section 2.2 develops these and some other population properties of the HMNM model. The characteristics of this (or any other) model are best understood through their implications for observable characteristics of financial asset returns. Section 2.3 develops the implications for the characteristics discussed in Section 1.1 using prior predictive distributions. Section 2.4 briefly describes the BMCMC posterior simulator.

2.1 The complete model

Consider a generalization of the Markov normal mixture model in which the distribution of the variable of interest y_t, conditional on the state s_t and the covariate vector $\mathbf{x}_t$, is a normal mixture. As a simple example, suppose that there are two states. In the first state the distribution is a simple mixture of $N(-0.5, 1)$ $(p = 2/3)$ and $N(1, 0.5^2)$ $(p = 1/3)$. In the second state the distribution is a simple mixture of $N(0.8, 0.7^2)$ $(p = 1/2)$ and $N(-0.8, 0.7^2)$ $(p = 1/2)$. The transition matrix is

$$\mathbf{P} = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}. \tag{6}$$

This is a simple instance of the HMNM model. In this particular example the mean of $\varepsilon_t = y_t - \boldsymbol{\beta}'\mathbf{x}_t$ is zero in each state and ε_t is therefore serially uncorrelated, but it is not serially independent. Conditional on $s_{t1} = 1$ the distribution of ε_t is skewed, and conditional on $s_{t1} = 2$ it is symmetric. If the rows of $\mathbf{P}$ are proportional (unlike this example) then ε_t is serially independent, but the distribution of ε_t maintains the same flexibility as in the normal mixture model – in particular, leptokurtosis does not imply serial dependence as is the case in some models of volatility persistence like Gaussain GARCH models.

The HMNM model may be regarded as a restricted case of a Markov normal mixture model. In the preceding example this Markov normal mixture model has four states, component distributions $N(-0.5, 1)$, $N(1, 0.5^2)$, $N(0.8, 0.7^2)$, $N(-0.8, 0.7^2)$, and transition matrix

$$\mathbf{P}^* = \begin{bmatrix} 0.40 & 0.20 & 0.20 & 0.20 \\ 0.40 & 0.20 & 0.20 & 0.20 \\ 0.20 & 0.10 & 0.35 & 0.35 \\ 0.20 & 0.10 & 0.35 & 0.35 \end{bmatrix}. \tag{7}$$

To generalize this example define the 1×2 latent state vector $\mathbf{s}_t$, with *persistent component* $s_{t1} \in \{1, \ldots, m_1\}$ and *transitory component* $s_{t2} \in \{1, \ldots, m_2\}$. The persistent component behaves just like the latent state in the Markov normal mixture model: s_{t1} depends only on $s_{t-1,1}$, with $P(s_{t1} = j \mid s_{t-1,1} = i) = p_{ij}$. In the example the matrix $[p_{ij}]$ is (6). Conditional on s_{t1} the transitory component behaves just like the latent state in the normal mixture model: $P(s_{t2} = k \mid s_{t1} = j) = r_{jk}$. In the example $m_1 = m_2 = 2$, $r_{11} = 2/3$, $r_{12} = 1/3$, and $r_{21} = r_{22} = 1/2$.

In the generalization the HMNM model may be regarded as an $m_1 m_2$-state MNM model, with a special structure for the $m_1 m_2 \times m_1 m_2$ transition matrix $\mathbf{P}^*$, reflected in

(7): if $[(i-1)/m_2] = [(k-1)/m_2]$ then $p^*_{ij} = p^*_{kj}$ $(j = 1, \ldots, m_1 m_2)$ and $p^*_{ji}/p^*_{jk} = p^*_{\ell i}/p^*_{\ell k}$ $(j, \ell = 1, \ldots, m_1 m_2)$. These restrictions turn out to be critical for practical applied work. Whereas the matrix $\mathbf{P}^*$ has $m_1 m_2 (m_1 m_2 - 1)$ parameters, the matrices $\mathbf{P}$ and $\mathbf{R}$ have $m_1 (m_1 + m_2 - 1)$ parameters. As discussed in Section 2.4 using the unrestricted Markov normal mixture model increases computation time by a factor of roughly m_2. Section 3.2 provides evidence that values of m_1 and m_2 on the order of 4 or 5 are optimal for daily financial asset returns, and also provides formal comparison of the HMNM and MNM models.

The HMNM model may also be regarded as an artificial neural network (Kuan and White (1992)) with observed inputs $\mathbf{x}_t$ $(t = 1, \ldots, T)$ and i.i.d. $N(0,1)$ unobserved input vector $\mathbf{z} = (z_1, \ldots, z_T)'$. The first hidden layer selects states $s_{11}, \ldots, s_{T1}$, the second hidden layer next selects states $s_{21}, \ldots, s_{T2}$, and the output vector is then $y_t = \boldsymbol{\beta}'\mathbf{x}_t + \phi_i + \psi_{ij} + (h_i \cdot h_{ij})^{-1/2} z_t$ $(t = 1, \ldots, T)$ where $i = s_{t1}$ and $j = s_{t2}$ $(t = 1, \ldots, T)$.

2.1.1 The conditional distribution of observables

Turning to a full specification of the HMNM model, for each of $t = 1, \ldots, T$ periods there are three kinds of variables. The observable random variable y_t denotes an asset return of interest. The $k \times 1$ vector $\mathbf{x}_t$ denotes deterministic variables for period t, such as an intercept (used in subsequent applications in this study) or indicators for days of the week (not used). The latent 1×2 vector $\mathbf{s}_t = (s_{t1}, s_{t2})$ denotes the persistent and transitory states in period t. The three sets of variables can be expressed compactly in the $T \times 1$ vector $\mathbf{y} = (y_1, \ldots, y_T)'$, the $T \times k$ matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]'$, and the $T \times 2$ matrix $\mathbf{s} = [\mathbf{s}_1', \ \ldots, \ \mathbf{s}_T']'$.

Let $\mathbf{s}^1 = (s_{11}, \ldots, s_{T1})'$ denote all T persistent states. Then

$$p\left(\mathbf{s}^1 \mid \mathbf{X}\right) = \pi_{s_{11}} \prod_{t=2}^{T} p_{s_{t-1,1},s_{t1}} = \pi_{s_{11}} \prod_{i=1}^{m_1} \prod_{j=1}^{m_1} p_{ij}^{T_{ij}}, \tag{8}$$

where T_{ij} is the number of transitions from persistent state i to j in $\mathbf{s}^1$. The $n \times n$ Markov transition matrix $\mathbf{P}$ is irreducible and aperiodic, and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{m_1})'$ is the unique stationary distribution of $\{s_{t1}\}$. For subsequent reference define the $m_1 \times 1$ vectors $\mathbf{p}_i = \left(p_{i1}, \ldots p_{i_{m_1}}\right)'$, so that $\mathbf{P}' = [\mathbf{p}_1, \ldots, \mathbf{p}_{m_1}]$.

Let $\mathbf{s}^2 = (s_{12}, \ldots, s_{T2})'$ denote all T transitory states. Then

$$p\left(\mathbf{s}^2 \mid \mathbf{s}^1, \mathbf{X}\right) = \prod_{t=1}^{T} r_{\mathbf{s}_t} = \prod_{i=1}^{m_1} \prod_{j=1}^{m_2} r_{ij}^{U_{ij}}, \tag{9}$$

where U_{ij} is the number of occurrences of $\mathbf{s}_t = (i, j)$ $(t = 1, \ldots, T)$. For subsequent reference let the $m_2 \times 1$ vectors $\mathbf{r}_i$ denote the transitory state probabilities $\mathbf{r}_i = (r_{i1}, \ldots, r_{im_2})'$ within each persistent state $(i = 1, \ldots, m_1)$ and define the $m_1 \times m_2$ matrix $\mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_{m_1}]'$.

The observables y_t depend on the latent states $\mathbf{s}_t$ and the deterministic variables $\mathbf{x}_t$. Conditional on $\mathbf{s}_t = (i, j)$,

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \phi_i + \psi_{ij} + \varepsilon_t; \quad \varepsilon_t \sim N\left[0, (h \cdot h_i \cdot h_{ij})^{-1}\right]. \tag{10}$$

For subsequent reference define the $m_1 \times 1$ vector $\boldsymbol{\phi} = \left(\phi_1, \ldots, \phi_{m_1}\right)'$, the $m_2 \times 1$ vectors $\boldsymbol{\psi}_i = \left(\psi_{i1}, \ldots, \psi_{im_2}\right)'$ $(i = 1, \ldots, m_1)$ and the $m_1 m_2 \times 1$ vector $\boldsymbol{\psi} = \left(\boldsymbol{\psi}_1', \ldots, \boldsymbol{\psi}_{m_1}'\right)'$. Further define the $m_1 \times 1$ vector $\mathbf{h} = (h_1, \ldots, h_{m_1})'$ and the $m_1 \times m_2$ matrix $\mathbf{H} = [h_{ij}]$.

For $t = 1, \ldots, T$ let $\mathbf{d}_t^1$ be an $m_1 \times 1$ vector of dichotomous variables in which $d_{ti}^1 = 1$ if $s_{t1} = i$ and $d_{ti}^1 = 0$ otherwise. Similarly define the $m_2 \times 1$ vector $\mathbf{d}_t^2$ with $d_{tj}^2 = 1$ if $s_{t2} = j$ and $d_{tj}^2 = 0$ if not. Let $\mathbf{d}_t = \mathbf{d}_t^1 \otimes \mathbf{d}_t^2$ $(t = 1, \ldots, T)$. In this notation (10) becomes

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\phi}'\mathbf{d}_t^1 + \boldsymbol{\psi}'\mathbf{d}_t + \boldsymbol{\varepsilon}_t \qquad (t = 1, \ldots, T)\,. \tag{11}$$

2.1.2 The prior distribution

The parameters $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ in (11) can be identified by means of two conventions about state means. The first convention is that the unconditional mean of the persistent states is $\mathbf{0}$, which is equivalent to $\boldsymbol{\pi}'\boldsymbol{\phi} = 0$. Let the $m_1 \times (m_1 - 1)$ matrix $\mathbf{C}_0$ be the orthonormal complement of $\boldsymbol{\pi}$: that is, $\boldsymbol{\pi}'\mathbf{C}_0 = \mathbf{0}'$ and $\mathbf{C}_0'\mathbf{C}_0 = \mathbf{I}_{m_1 - 1}$. Define the $(m_1 - 1) \times 1$ vector $\widetilde{\boldsymbol{\phi}} = \mathbf{C}_0'\boldsymbol{\phi}$ and note that $\boldsymbol{\phi} = \mathbf{C}_0\widetilde{\boldsymbol{\phi}}$ because $\boldsymbol{\pi}'\boldsymbol{\phi} = 0$.

The second convention is that the unconditional mean of the transitory states within each permanent state is $\mathbf{0}$, which is equivalent to $\boldsymbol{\psi}_i'\mathbf{r}_i = 0$ $(i = 1, \ldots, m_1)$. Let $\mathbf{C}_j$ be an $m_2 \times (m_2 - 1)$ orthonormal complement of $\mathbf{r}_j$, define the $(m_2 - 1) \times 1$ vectors $\widetilde{\boldsymbol{\psi}}_j' = \mathbf{C}_j'\boldsymbol{\psi}_j$, and note that $\boldsymbol{\psi}_j = \mathbf{C}_j\widetilde{\boldsymbol{\psi}}_j$ $(j = 1, \ldots, m_1)$. Construct the $m_1 m_2 \times m_1 (m_2 - 1)$ block diagonal matrix $\mathbf{C} = \text{Blockdiag}\left[\mathbf{C}_1, \ldots, \mathbf{C}_{m_1}\right]$ and the $m_1 (m_2 - 1) \times 1$ vector $\widetilde{\boldsymbol{\psi}} = \left(\widetilde{\boldsymbol{\psi}}_1', \ldots, \widetilde{\boldsymbol{\psi}}_{m_1}'\right)'$. Then $\boldsymbol{\psi} = \mathbf{C}\widetilde{\boldsymbol{\psi}}$, and substituting in (11),

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \widetilde{\boldsymbol{\phi}}'\mathbf{C}_0'\mathbf{d}_t^1 + \widetilde{\boldsymbol{\psi}}'\mathbf{C}'\mathbf{d}_t + \varepsilon_t.$$

From this point forward in developing the prior distribution, $\widetilde{\boldsymbol{\phi}}$ and $\widetilde{\boldsymbol{\psi}}$ replace $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$.

A proper prior distribution for the parameters $\boldsymbol{\beta}$, $\widetilde{\boldsymbol{\phi}}$, $\widetilde{\boldsymbol{\psi}}$, $\mathbf{P}$, $\mathbf{R}$, h, $\mathbf{h}$ and $\mathbf{H}$, combined with the specifications (8), (9) and (10), provides a probability distribution for any observable $(y_1, \ldots, y_T)'$. The following family of conditionally conjugate prior distributions is convenient, for reasons discussed in Section 2.4. Each prior distribution is described with reference to hyperparameters, each denoted with an underbar. The hyperparameters are numbers that must be chosen prior to inference; Section 2.3 describes the values used in the applied work in this study and how these numbers were chosen.

Given the restrictions on $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$, $E\left(y_t \mid \mathbf{x}_t, A\right) = \boldsymbol{\beta}'\mathbf{x}_t$. The prior distribution of $\boldsymbol{\beta}$ is Gaussian,

$$\boldsymbol{\beta} \sim N\left(\underline{\boldsymbol{\beta}}, \underline{\mathbf{H}}_{\beta}^{-1}\right); \tag{12}$$

since $x_t = 1$ in subsequent applications, this involves the choice of the two numbers $\underline{\beta}$ and $\underline{h}_{\beta}$.

The prior distribution of $\mathbf{p}_i$ the i'th row of $\mathbf{P}$, is

$$\mathbf{p}_i \sim Beta_{m_1}\left[\underline{p}_{i1}^*, \ldots, \underline{p}_{im_1}^*\right], \tag{13}$$

where $\underline{p}^*_{ij} = \delta_{ij}\underline{p}_0 + (1-\delta_{ij})\,\underline{p}_1$. The multivariate $Beta_n(a_1,\ldots,a_n)$ distribution, also known as the Dirichlet distribution, has p.d.f.

$$p_n(x_1,\ldots,x_n) \propto \prod_{i=1}^{n} x_i^{(a_i-1)}.$$

The prior distributions (13) are independent across the rows i of $\mathbf{P}$. They require the specification of two hyperparameters, $\underline{p}_0$ and $\underline{p}_1$. Loosely speaking the prior predictive distribution of the model will display substantial serial persistence for $\{y_t\}$ when $\underline{p}_0 \gg \underline{p}_1$; Section 2.3 returns to a more precise elucidation of this relationship.

The prior distribution of $\mathbf{r}_i$, the i'th row of $\mathbf{R}$, is

$$\mathbf{r}_i \sim Beta_{m_2}(\underline{r},\ldots,\underline{r}). \tag{14}$$

These prior distributions are independent across the rows of $\mathbf{R}$.

The prior distributions of the precision parameters in h, $\mathbf{h}$ and $\mathbf{H}$ are independent gamma,

$$\underline{s}^2 h \sim \chi^2(\underline{\nu}); \tag{15}$$

$$\underline{\nu}_1 h_j \sim \chi^2(\underline{\nu}_1)\ (\ j=1,\ldots,m_1); \tag{16}$$

$$\underline{\nu}_2 h_{ij} \sim \chi^2(\underline{\nu}_2)\ (\ i=1,\ldots,m_1;\ j=1,\ldots,m_2). \tag{17}$$

The prior means of all the precision parameters in $\mathbf{h}$ and $\mathbf{H}$ are one, and their variances are $2/\underline{\nu}_1$. The restriction on the mean of each of these parameters resolves the identification question apparent in (10), but does so only in the prior distribution: that is, even if the distribution of y_t conditional on $\mathbf{x}_t$ and $\mathbf{s}_t$ were known to be a specific instance of the HMNM model, there would still be uncertainty about h, h_i, and h_{ij}, and that uncertainty would be specified by (15)-(17).

The prior distributions of $\tilde{\boldsymbol{\phi}}$ and $\widetilde{\boldsymbol{\psi}}$ are Gaussian,

$$\tilde{\boldsymbol{\phi}} \mid h \sim N\left[\mathbf{0}, \left(\underline{h}_\phi h\right)^{-1}\mathbf{I}_{m_1-1}\right], \tag{18}$$

$$\tilde{\boldsymbol{\psi}}_j \mid (h_j,h) \sim N\left[\mathbf{0}, \left(\underline{h}_\psi h\cdot h_j\right)^{-1}\mathbf{I}_{m_2-1}\right]\ (\ j=1,\ldots,m_1), \tag{19}$$

all m_1+1 components being independent. In the case of $\tilde{\boldsymbol{\phi}}$, precision is scaled relative to h. This relative precision is important in establishing the set of densities for y_t that is reasonable under the prior. For example, when $\underline{h}_\phi$ is large then the probability of a bimodal or multimodal distribution is small, but as $\underline{h}_\phi \to 0$ this probability approaches 1. The scaling for the precision of the $\widetilde{\boldsymbol{\psi}}_j$ reflects similar considerations. (Geweke (2005), Section 6.4.2 provides additional detail on these points.)

The priors are invariant with respect to the particular choices of the orthonormal complements $\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{m_1}$. To establish this fact, for any $m_1 \times 1$ vector $\mathbf{f}$ consider $\mathbf{f}'\boldsymbol{\phi}$ and set $\mathbf{f} = [\mathbf{C}_0\ \boldsymbol{\pi}]\,\mathbf{f}^* = \mathbf{C}_0\mathbf{f}_1^* + \boldsymbol{\pi} f_2^*$. From (18),

$$\begin{aligned}\mathbf{f}'\boldsymbol{\phi} &= \mathbf{f}_1^{*\prime}\mathbf{C}_0'\boldsymbol{\phi} + f_2^*\boldsymbol{\pi}'\boldsymbol{\phi} = \mathbf{f}_1^{*\prime}\mathbf{C}_0'\boldsymbol{\phi} \sim N\left[0, \left(\underline{h}_\phi h\right)^{-1}\mathbf{f}_1^{*\prime}\mathbf{C}_0'\mathbf{C}_0\mathbf{f}_1^*\right] \\ &= N\left[0, \left(\underline{h}_\phi h\right)^{-1}\mathbf{f}_1^{*\prime}\mathbf{f}_1^*\right] = N\left[0, \left(\underline{h}_\phi h\right)^{-1}\mathbf{f}_1'\mathbf{f}_1\right]\end{aligned}$$

because $\mathbf{f}_1^* = \mathbf{C}_0'\mathbf{f}_1$. Exactly the same argument applies to the $\boldsymbol{\psi}_j$.

The prior distribution consists of the $m_1\,(m_2+4)+2$ components (12)-(19). This distribution, together with the conditional distributions (8), (9) and (10), provides a complete model of asset returns $\{y_t\}$, as described in Section 1.2. In this distribution the latent states $\mathbf{s}_t$ are entirely exchangeable: the prior density evaluated at a given matrix $\mathbf{s}$ remains the same under any of the $m_1!m_2!$ permutations of the states. The prior distribution thereby incorporates the fact that the states do not have substantive interpretations, but rather are a technical device to provide substantial flexibility in modeling $\{y_t\}$. (Section 2.4 returns to the implications for the posterior distribtuion and BMCMC.) Of the 11 hyperparameters $\underline{\beta}$, $\underline{h}_\beta$, $\underline{s}^2$, $\underline{\nu}$, $\underline{\nu}_1$, $\underline{\nu}_2$, $\underline{p}_0$, $\underline{p}_1$, $\underline{r}$, $\underline{h}_\phi$, $\underline{h}_\psi$, the first four must take account of the location and scale of $\{y_t\}$. (They, or their equivalent, would need to be specified in the textbook i.i.d. normal model for $\{y_t\}$, but of course that model is entirely inadequate for financial asset returns.) The other 7 prior hyperparameters, together with the numbers of states m_1 and m_2, govern the degree of persistence and shape of the distribution in $\{y_t\}$. Section 2.3 provides numerical values of the hyperparameters for this study, using the functions of interest detailed in Section 1.1 and prior predictive distributions used as described in Section 1.2.

2.2 Properties of the model: some theory

In many settings no-arbitrage conditions imply that financial asset returns y_t should be nearly serially uncorrelated. Especially in the presence of transactions costs the condition of no serial correlation will only be an approximation, but approximations can be very useful in simplifying models and improving their forecasting performance. Absence of serial correlation is a straightforward restriction in the HMNM model.

Theorem 1 *Conditional on* $\mathbf{x}_t$ $(t=1,\ldots,T)$*, the observables* y_t $(t=1,\ldots,T)$ *are serially uncorrelated if* $\boldsymbol{\phi}=\mathbf{0}$*. Suppose further that* $\mathbf{P}$ *is irreducible and aperiodic and its eigenvalues are distinct. Then the observables* y_t *are serially uncorrelated if and only if* $\boldsymbol{\phi}=\mathbf{0}$*.*

Proof. Using the methods of Ryden et al. (1998) for the Markov normal mixture model,

$$cov\left(y_t, y_{t-s} \mid \mathbf{x}_1,\ldots,\mathbf{x}_T\right) = \boldsymbol{\phi}'\,\mathbf{B}^{s'}\boldsymbol{\Pi}\,\boldsymbol{\phi} = \boldsymbol{\phi}'\boldsymbol{\Pi}\,\mathbf{B}^{s}\,\boldsymbol{\phi}\ \ (s=1,2,\ldots)\,, \tag{20}$$

where $\boldsymbol{\Pi}$ =$diag(\boldsymbol{\pi}), \mathbf{B}=\mathbf{P}-\mathbf{e}_{m_1}\boldsymbol{\pi}'$, which establishes sufficiency.

If the eigenvalues of $\mathbf{P}$ are distinct then $\mathbf{P}$ is diagonable and it has spectral decomposition $\mathbf{P}=\mathbf{Q}^{-1}\boldsymbol{\Lambda}\mathbf{Q}$, where the matrix $\boldsymbol{\Lambda}$ =diag $(\lambda_1,\ldots,\lambda_{m_1})$ contains the ordered eigenvalues λ_j of $\mathbf{P}$, $|\lambda_1|\geq|\lambda_2|\geq|\lambda_3|\geq\ldots|\lambda_{m_1}|$. The matrix $\mathbf{Q}$ has orthogonal columns and we may take

$$\begin{aligned}\mathbf{Q} &= \left[\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_m\right]' = \left[\boldsymbol{\pi},\mathbf{q}_2,\ldots,\mathbf{q}_m\right]',\\ \mathbf{Q}^{-1} &= \left[\mathbf{q}^1,\mathbf{q}^2,\ldots,\mathbf{q}^m\right] = \left[\mathbf{e}_m,\mathbf{q}^2,\ldots,\mathbf{q}^m\right].\end{aligned}$$

If $\mathbf{P}$ is also irreducible and aperiodic then $\lambda_1=1>|\lambda_2|$ and we may write

$$\mathbf{B} = \mathbf{Q}^{-1}\boldsymbol{\Lambda}\mathbf{Q} - \mathbf{q}^1\mathbf{q}_1' = \mathbf{Q}^{-1}\widetilde{\boldsymbol{\Lambda}}\mathbf{Q} \tag{21}$$

where $\widetilde{\mathbf{\Lambda}} = \text{diag}\,(0, \lambda_2, \ldots, \lambda_{m_1})$. From (20) absence of serial correlation is equivalent to

$$\boldsymbol{\phi}'\mathbf{\Pi}\mathbf{Q}^{-1}\widetilde{\mathbf{\Lambda}}^s\mathbf{Q}\boldsymbol{\phi} = 0 \quad (s = 1, 2, \ldots)\,.$$

The first element of $\mathbf{Q}\boldsymbol{\phi}$ is $\mathbf{q}_1'\boldsymbol{\phi} = \boldsymbol{\pi}'\boldsymbol{\phi} = 0$, and so

$$\boldsymbol{\phi}'\mathbf{\Pi}\mathbf{Q}^{-1}\mathbf{\Lambda}^s\mathbf{Q}\boldsymbol{\phi} = 0 \quad (s = 1, 2, \ldots)\,. \tag{22}$$

Define the $m_1 \times m_1$ matrix

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_{m_1} \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_{m_1}^2 \\ \vdots & \vdots & & \vdots \\ \lambda_1^{m_1} & \lambda_2^{m_1} & \cdots & \lambda_{m_1}^{m_1} \end{bmatrix},$$

whose determinant is $\left(\prod_{i=1}^{m_1}\lambda_i\right)\prod_{i<j}^{m_1}(\lambda_i - \lambda_j) \neq 0$ (Rao (1965), p 28). Let $\mathbf{A} = \mathbf{D}^{-1}$ and let $\delta_{i,j}$ denote the Kronecker delta function; then

$$\sum_{s=1}^{m_1} a_{is}\lambda_j^s = \delta_{i,j} \implies \sum_{i=1}^{m_1}\sum_{s=1}^{m_1} a_{is}\mathbf{\Lambda}^s = \mathbf{I}_{m_1} \quad (i = 1, \ldots, m_1)\,,$$

and from (22)

$$\sum_{i=1}^{m_1}\sum_{s=1}^{m_1} a_{is}\boldsymbol{\phi}'\mathbf{\Pi}\mathbf{Q}^{-1}\mathbf{\Lambda}^s\mathbf{Q}\boldsymbol{\phi} = \boldsymbol{\phi}'\mathbf{\Pi}\boldsymbol{\phi} = \sum_{i=1}^{m}\phi_i^2\pi_i = 0.$$

■

Thus all that is required to impose absence of serial correlation is to omit the vector $\boldsymbol{\phi}$ from the model. This is the limiting case $\underline{h}_\phi \to \infty$ in the specification of the prior distribution (18), which may therefore be used to express the condition that serial correlation is small rather than that it is absent.

With a sufficiently large number of states, m_1 and m_2, the HMNM model is quite flexible and can accommodate many kinds of persistence in higher moments, even in the absence of serial correlation. Yet the structure of the model imposes systematic restrictions on these moments. These properties can be expressed in terms of a linear representation of the $p \times 1$ vector of mixed powers $\mathbf{z}_t^{(p)} = (y_t^1, \ldots, y_t^p)'$.

The representation involves the $m \times m$ Markov transition matrix $\mathbf{P}$, and the state-specific moments

$$\mu_j^h = E\left(y_t^h \mid s_t = j\right) \qquad (j = 1, \ldots, m;\ h = 1, \ldots, 2p)\,. \tag{23}$$

In the case of the HMNM model $m = m_1$, and the moments (23) are known functions of the model parameters. The representation pertains more generally to any Markov mixture model in which the first $2p$ state-specific moments are finite. It simplifies the notation to take $\boldsymbol{\beta} = \mathbf{0}$; in the subsequent applications in which $x_t = 1$, this amounts to removing the unconditional mean.

For given $p > 0$ define the $p \times 1$ vector of mixed powers $\mathbf{z}_t^{(p)\prime} = (y_t^1, \ldots, y_t^p)$. Take the $p \times 1$ vectors of moments

$$\boldsymbol{\mu}_j^{(p)} = E\left(\mathbf{z}_t^{(p)} \mid s_t = j\right) = \left(\mu_j^1, \ldots, \mu_j^p\right)' \qquad (j = 1, \ldots, m)$$

and arrange them in the $p \times m$ matrix

$$\mathbf{M}^{(p)} = \left[\boldsymbol{\mu}_1^{(p)}, \ldots, \boldsymbol{\mu}_m^{(p)}\right].$$

Denote the second central moments of $\mathbf{z}_t^{(p)}$,

$$\mathbf{R}_j^{(p)} = E\left\{\left(\mathbf{z}_t^{(p)} - \boldsymbol{\mu}_j^{(p)}\right)\left(\mathbf{z}_t^{(p)} - \boldsymbol{\mu}_j^{(p)}\right)' \mid s_t = j\right\} \qquad (j = 1, \ldots, m). \tag{24}$$

Theorem 2 *Suppose that in a Markov mixture model with m states and irreducible and aperiodic transition matrix $\mathbf{P}$, the stationary distribution is $\boldsymbol{\pi}$. Suppose further that the m moment matrices $\mathbf{R}_j^{(p)}$ $(j = 1, \ldots, m)$ are finite. Then the unconditional mean of $\mathbf{z}_t^{(p)}$ is*

$$E\left(\mathbf{z}_t^{(p)}\right) = \boldsymbol{\mu}^{*(p)} = \mathbf{M}^{(p)}\boldsymbol{\pi}.$$

The instantaneous variance matrix is

$$\begin{aligned}\boldsymbol{\Gamma}_0^{(p)} &= E\left(\mathbf{z}_t^{(p)} - \boldsymbol{\mu}^{*(p)}\right)\left(\mathbf{z}_t^{(p)} - \boldsymbol{\mu}^{*(p)}\right)' \\ &= \sum_{j=1}^{m} \pi_j \left(\mathbf{R}_j^{(p)} + \boldsymbol{\mu}_j^{(p)}\boldsymbol{\mu}_j^{(p)\prime}\right) - \boldsymbol{\mu}^{*(p)}\boldsymbol{\mu}^{*(p)\prime}.\end{aligned} \tag{25}$$

The dynamic covariance matrices are

$$\begin{aligned}\boldsymbol{\Gamma}_u^{(p)} &= E\left(\mathbf{z}_t^{(p)} - \boldsymbol{\mu}^{*(p)}\right)\left(\mathbf{z}_{t-u}^{(p)} - \boldsymbol{\mu}^{*(p)}\right)' \\ &= \mathbf{M}^{(p)}\mathbf{B}^{u\prime}\boldsymbol{\Pi}\mathbf{M}^{(p)\prime} \qquad (u = 1, 2, 3, \ldots)\end{aligned} \tag{26}$$

where $\mathbf{B} = \mathbf{P} - \mathbf{e}_m\boldsymbol{\pi}'$, $\mathbf{B}^u = \mathbf{P}^u - \mathbf{e}_m\boldsymbol{\pi}'$, and $\boldsymbol{\Pi} = diag\,(\pi_1, \ldots, \pi_m)$.

Proof. The instantaneous variance matrix $\boldsymbol{\Gamma}_0^{(p)}$ is immediately attained by considering

$$\begin{aligned}\boldsymbol{\Gamma}_0^{(p)} &= E\left[\mathbf{z}_t^{(p)} - \boldsymbol{\mu}^{*(p)}\right]\left[\mathbf{z}_t^{(p)} - \boldsymbol{\mu}^{*(p)}\right]' = E\left(\mathbf{z}_t^{(p)}\mathbf{z}_t^{(p)\prime}\right) - \boldsymbol{\mu}^{*(p)}\boldsymbol{\mu}^{*(p)\prime} \\ &= \sum_{j=1}^{m} \pi_j \left[\mathbf{z}_t^{(p)}\mathbf{z}_t^{(p)\prime} \mid s_t = j\right] - \boldsymbol{\mu}^{*(p)}\boldsymbol{\mu}^{*(p)\prime} \\ &= \sum_{j=1}^{m} \pi_j \left(\mathbf{R}_j^{(p)} + \boldsymbol{\mu}_j^{(p)}\boldsymbol{\mu}_m^{(p)\prime}\right) - \boldsymbol{\mu}^{*(p)}\boldsymbol{\mu}^{*(p)\prime}.\end{aligned}$$

The dynamic covariance matrices $\mathbf{\Gamma}_u^{(p)}$ $(p > 0)$ are obtained by conditioning on s_t and s_{t-u}, exploiting serial independence of observables after conditioning on the states, and then by marginalizing out the states:

$$\begin{aligned}
\mathbf{\Gamma}_u^{(p)} &= cov\left(\mathbf{z}_t^{(p)}, \mathbf{z}_{t-u}^{(p)}\right) = E\left(\mathbf{z}_t^{(p)}\mathbf{z}_{t-u}^{(p)\prime}\right) - \boldsymbol{\mu}^{*(p)}\boldsymbol{\mu}^{(p)\prime} \\
&= \sum_{j=1}^{m}\sum_{i=1}^{m} E\left(\mathbf{z}_t^{(p)}\mathbf{z}_{t-u}^{(p)\prime} \mid s_t = j, s_{t-u} = i\right)\left[\mathbf{P}^u\right]_{ij}\,\pi_i - \mathbf{M}^{(p)}\boldsymbol{\pi}\boldsymbol{\pi}'\mathbf{M}^{(p)\prime} \\
&= \sum_{j=1}^{m}\sum_{i=1}^{m} E\left(\mathbf{z}_t^{(p)} \mid s_t = j\right)\,E\left(\mathbf{z}_{t-u}^{(p)\prime} \mid s_t = i\right)\left[\mathbf{P}^u\right]_{ij}\,\pi_i - \mathbf{M}^{(p)}\boldsymbol{\pi}\boldsymbol{\pi}'\mathbf{M}^{(p)\prime} \\
&= \sum_{j=1}^{m}\sum_{i=1}^{m} \boldsymbol{\mu}_j^{(p)}\boldsymbol{\mu}_i^{(p)\prime}\left[\mathbf{P}^u\right]_{ij}\,\pi_i - \boldsymbol{\mu}^{(p)}\mathbf{e}_m'\mathbf{\Pi}\mathbf{M}^{(p)\prime} = \mathbf{M}^{(p)}\mathbf{B}^{u\prime}\mathbf{\Pi}\mathbf{M}^{(p)\prime},
\end{aligned}$$

where $\mathbf{B}^u = \left(\mathbf{P} - \mathbf{e}_m\boldsymbol{\pi}'\right)^u = \mathbf{P}^u - \mathbf{e}_m\boldsymbol{\pi}'$. ■

The geometric decay in u evident in (26) is that of an autoregressive process of finite order. However, this pattern does not extend to (25), which suggests that $\mathbf{z}_t^{(p)}$ might be represented as the sum of such a process and a serially uncorrelated process that is uncorrelated with it.

Theorem 3 *Suppose that the Markov transition matrix* $\mathbf{P}$ *has spectral decomposition* $\mathbf{P} = \mathbf{Q}^{-1}\mathbf{\Lambda}\mathbf{Q}$ *defined in the proof of Theorem 1. Let* $\lambda_1, \ldots, \lambda_r$ *be the distinct eigenvalues in the open unit interval associated with at least one column of* $\mathbf{Q}'$ *not contained in the null space of* $\mathbf{M}^{(p)}$*. Then* $\mathbf{z}_t^{(p)}$ *can be decomposed as the sum of two vector processes,*

$$\mathbf{z}_t^{(p)} = \mathbf{v}_t^{(p)} + \boldsymbol{\eta}_t^{(p)}.$$

The process $\boldsymbol{\eta}_t^{(p)}$ *is uncorrelated with* $\mathbf{v}_{t+u}^{(p)}$ $(u = 0, \pm 1, \pm 2, \ldots)$ *and is itself serially uncorrelated, var* $\left(\boldsymbol{\eta}_t^{(p)}\right) = \sum_{j=1}^{m}\pi_j\mathbf{R}_j^{(p)}$*. The process* $\mathbf{v}_t^{(p)}$ *has vector autoregressive representation*

$$\mathbf{v}_t^{(p)} = \sum_{u=1}^{r}\alpha_u\mathbf{v}_{t-u}^{(p)} + \boldsymbol{\omega}_t^{(p)}$$

in which the coefficients α_u $(u = 1, \ldots, r)$ *are scalars. The roots of the generating polynomial* $1 - \sum_{u=1}^{r}\alpha_u z^u$ *are* $\lambda_1, \ldots, \lambda_r$*.*

Proof. Adopt the notation in the proof of Theorem 1. From (21), $\mathbf{B}^u = \sum_{j=2}^{m}\lambda_j^u\mathbf{q}^j\mathbf{q}_j'$. Substituting in (26),

$$\mathbf{\Gamma}_u^{(p)} = \sum_{j=2}^{m}\lambda_j^u\mathbf{M}^{(p)}\mathbf{q}_j\mathbf{q}^{j\prime}\mathbf{M}^{(p)\prime} = \sum_{j=2}^{r+1}\lambda_j^u\mathbf{A}_j' \qquad (u = 1, 2, 3, \ldots)$$

where

$$\mathbf{A}_j' = \sum_{h\in H_j}\mathbf{M}^{(p)}\mathbf{q}_h\mathbf{q}^{h\prime}\mathbf{M}^{(p)\prime}, \;\; H_j = \left\{h : \mathbf{q}_h'\mathbf{P} = \lambda_j\mathbf{q}_h', \;\mathbf{M}^{(p)}\mathbf{q}_h \neq \mathbf{0}\right\}..$$

Observe that r is the number of distinct eigenvalues of $\mathbf{P}$ with modulus in the open unit interval associated with as least one column of $\mathbf{Q}'$ not in the column null space of $\mathbf{M}^{(p)}$. In other words, r can be less than $m-1$ because some eigenvalues are equal to zero (as in the compound Markov model interpreted as having $m = m_1 m_2$ states), because some eigenvalues are repeated, or because some eigenvalues are associated with columns of $\mathbf{Q}'$ all in the column null space of $\mathbf{M}^{(p)}$.

Define now a stochastic process $\mathbf{v}_t^{(p)}$ with autocovariances $\tilde{\mathbf{\Gamma}}_u^{(p)} = \sum_{j=2}^{r+1} \lambda_j^u \mathbf{A}_j' \;\; (u > 0)$ and $\tilde{\mathbf{\Gamma}}_0^{(p)} = \sum_{j=2}^{r+1} \mathbf{A}_j'$. Then for $u > 0$, $\tilde{\mathbf{\Gamma}}_u^{(p)} = \mathbf{\Gamma}_u^{(p)}$, while

$$\tilde{\mathbf{\Gamma}}_0^{(p)} = \sum_{j=2}^{r+1} \mathbf{A}_j' = \sum_{j=1}^{m} \boldsymbol{\mu}_j^{(p)} \boldsymbol{\mu}_j^{\prime(p)} \pi_j - \boldsymbol{\mu}^{*(p)} \boldsymbol{\mu}^{*(p)\prime}.$$

Notice that the matrix $\mathbf{\Gamma}_0^{(p)} - \tilde{\mathbf{\Gamma}}_0^{(p)} = \sum_{j=1}^{m} \mathbf{R}_j^{(p)} \boldsymbol{\pi}_j$ is positive (semi) definite, since each $\mathbf{R}_j^{(p)}$ is a variance matrix (24).

Given that there are r distinct eigenvalues of $\mathbf{P}$, $\lambda_2, \ldots, \lambda_{r+1}$, with modulus in the open unit interval, contributing to the determination of $\mathbf{\Gamma}_u^{(p)} = \tilde{\mathbf{\Gamma}}_u^{(p)}$, there exists a unique set of constants $\alpha_1, \ldots, \alpha_r$ such that

$$\lambda_j^r - \sum_{i=1}^{r} \alpha_i \lambda_j^{r-i} = 0 \qquad (j = 2, \ldots, r+1)\,.$$

The coefficients $\alpha_1, \ldots, \alpha_r$ determine a degree r polynomial whose roots are $\lambda_2^{-1}, \ldots, \lambda_r^{-1}$. Thus for all $u > r$,

$$\begin{aligned} \tilde{\mathbf{\Gamma}}_u^{(p)} - \sum_{i=1}^{r} \alpha_i \tilde{\mathbf{\Gamma}}_{u-i}^{(p)} &= \sum_{j=2}^{r+1} \lambda_j^u \mathbf{A}_j' - \sum_{i=1}^{r} \alpha_i \sum_{j=2}^{r+1} \lambda_j^{u-i} \mathbf{A}_j' \\ &= \sum_{j=2}^{r+1} \left(\lambda_j^u - \sum_{i=1}^{r} \alpha_i \lambda_j^{u-i} \right) \mathbf{A}_j' = \mathbf{0}. \end{aligned}$$

The autocovariance function of $\left\{ \mathbf{v}_t^{(p)} \right\}$ therefore satisfies the Yule-Walker equations for a VAR(r) process with coefficient matrices $\alpha_i \mathbf{I}_{np} \quad (i = 1, \ldots, r)$. ■

It follows from Theorem 3 that $\mathbf{z}_t^{(p)}$ has a VARMA(r, r) representation

$$\left(\mathbf{I}_p - \sum_{u=1}^{r} \alpha_i \mathbf{I}_p L^u \right) \mathbf{z}_t^{(p)} = \left(\mathbf{I}_p - \sum_{u=1}^{r} \mathbf{B}_i^{(p)} L^u \right) \boldsymbol{\nu}_t^{(p)}.$$

2.3 Implications for observables

The prior hyperparameters of the complete model (Section 2.1), together with the values of m_1 and m_2, provide a prior distribution over observable functions of interest z_j through the prior predictive densities $p\,(z_j \mid A)$. Ideas about reasonable models are usually formulated in terms of these observable functions, rather than in terms of parameters or hyperparameters.

Simulation from the prior predictive distributions of the z_j using alternative values of the hyperparameters allows the econometrician to select prior hyperparameters so as to reflect prior beliefs about these observable functions of interest. In our experience, this exercise also deepens one's understanding of the model. As a by-product, it may reveal limitations in the ability of the model to account for observed behavior of the functions of interest.

Simulation of observable functions of interest from the prior predictive distribution can be a key strategy in a successful program of applied econometric research. Prior predictive simulators are typically quick to code and execute, compared with the coding and execution of either posterior simulators or algorithms for non-Bayesian estimation methods. Exercises with these simulators can quickly reveal some important properties of a new model, including deficiencies, and suggest modifications of the model, all before undertaking major investments in formal inference. They may even indicate that a particular type of modelling should be abandoned before incurring these costs. On the other hand, while prior predictive simulation can be quick, informative and suggest major changes at an early stage of research, it does not have the power of formal posterior inference in revealing advantages and disadvantages of alternative models and providing tools for decision-making.

A formal analysis of the model through its implications for the function $\mathbf{z} = (z_1, \ldots, z_J)'$ would require handling a J-dimensional multivariate distribution, and would ultimately focus on $p(\mathbf{z}^o \mid A)$; Geweke (2007b) explores this approach. The procedures here eschew this method in favor of procedures that are less formal and less demanding of time and resources. Formal analysis is left to the posterior distribution (Sections 3 and 4). All of the analyses presented here uses $M = 10,000$ iterations from the prior distribution. Analyses were conducted for several alternative values of m_1 and m_2, but only the case $m_1 = m_2 = 4$ is studied in this section. Section 3.2 provides evidence from posterior distributions on choices of m_1 and m_2, and that evidence shows that $m_1 = m_2 = 4$ is a relevant illustrative case.

2.3.1 Prior predictive distributions of the functions of interest

Our prior predictive analysis utilized alternative values of the prior distribution hyperparameters, but in the interests of conserving space all the tables and figures in this section, and all of the analysis in the balance of the study, employ only the selected values indicated in Table 3. The prior predictive analysis here uses two representations of $p(\mathbf{z} \mid A)$ that we have found useful in understanding the complete model, choosing the values of the prior hyperparameters, and checking for indications that the complete model might not account for characteristics of the data such as the ten enumerated in Section 1.1.2.

The first representation is tabular. Tables 4 and 5 indicate the univariate cumulative prior distribution functions for the 16 observable functions of interest detailed in Section 1.1.1, each evaluated at the observed values of these functions presented in the corresponding Tables 1 and 2. For example the sample skewness coefficient -4.859 and sample kurtosis coefficient 84.693 of the S&P 500 returns in 1986-1990 (fourth line, Table 1, columns (1)-(2)) correspond to the respective prior c.d.f. evaluations 0.071 and 0.777 (fourth line, Table 4, columns (1)-(2)). Of the 288 prior c.d.f evaluations in these tables, the only one outside the interval $(0.01, 0.99)$ is the first order autocorrelation coefficient of S&P 500 returns in 1971-1975 (first line of Table 4, column (5)).

The second representation of the prior distribution is graphical. Figures 1 and 2 represent prior predictive distributions of selected pairs of observable functions of interest, together with the corresponding values for each of the 18 combinations of returns and time periods presented in Tables 1 and 2. In each case the axes are chosen to include all 18 observed values, with margins on all four sides that are 30% of the range of observed values. The observed values are indicated by open circles for S&P 500 returns, open squares for pound-dollar returns, and stars for bond returns. The first 1000 values from the prior predictive bivariate distribution within this range are plotted as filled red dots within these figures. For example, Figure 1(c) represents the joint prior distribution of the sample first-order autocorrelation coefficient $\widehat{\rho}_1$ of returns and the GPH estimate $\widehat{d}$ of the long-memory parameter for absolute returns. It indicates the observed values of these two parameters for S&P 500 returns in 1971-1975 is more problematic than suggested in the first line of Table 4 column (5) and Table 5 column (5). Such insights are the reason for examining bivariate as well as univariate prior distributions. The lines in Figure 2 are the points at which the ordinates are equal in absolute value.

Figure 3 presents another aspect of the prior predictive distribution, and illustrates the richness of information that can be harvested from prior predictive simulations. Section 1.1.2 noted that there is substantial variation in observed functions of interest across time periods for the same asset returns as well as across returns. Conditional on the parameter values $\boldsymbol{\theta}_A$ drawn in the prior simulation, observable returns can be simulated for as many periods as desired. By simulating 2500 successive trading days, we can learn the bivariate distribution of observable functions of interest in two successive five-year periods. Figure 3 presents the results of this exercise for four selected functions of interest. The observed values are plotted using the same symbols, with the 14 points reflecting the 14 adjacent pairs of five-year time intervals in Tables 1 and 2.

2.3.2 Implications for the challenging characteristics of asset returns

With these two representations of the prior predictive distribution in hand, return to the ten observed characteristics of asset returns discussed in Section 1.1.2. The unconditional leptokurtosis and substantial persistence in absolute moments have been most widely noted in the literature. Table 1, columns (1)-(4) (for one- and ten-day returns) and Figure 1(a) (for one-day returns) indicate that the prior predictive distribution of skewness and excess kurtosis is generous, relative to observed variations in these functions of sample moments across the asset returns and time intervals considered in this study. Table 5 indicates that the prior predictive distribution of observable measures of persistence in absolute returns accommodates the observed measures. This is confirmed, in a bivariate context, by Figures 1(b), 1(d) and 2(d).

Section 1.1.2 identified ten additional specific observed aspects of asset returns. Our prior predictive analysis provides a preliminary indication of the consistency of the complete HMNM model with these observations.

1. The model is consistent with the observed measures of skewness and excess kurtosis, and can easily account for variation in these functions across different asset returns; see Table 1, columns (1)-(4), and Figure 1(a). The prior predictive analysis indicates

that a model with unchanging parameters is consistent with the variation in these functions across most time periods for a given return series; see Figures 3(a) and (b). The relative outliers in these figures are the result of the high excess kurtosis coefficient and low skewness coefficient for S&P 500 returns 1986-1990, and the high skewness of the dollar-pound 10-day returns in 1992-1996.

2. Figure 2(a) indicates the joint prior distribution of the sample skewness coefficient of one- and ten-day returns, over the same 1250-day (five-year) period used for all of the observable functions of interest. The area above and below the superimposed crossed lines in that figure corresponds to sample skewness coefficients larger in absolute value for ten-day returns than for one-day returns. The prior probability of this event is 0.34. Thus the prior predictive distribution is consistent with this ordering, commonly observed in the asset returns studied here and noted elsewhere (Harvey and Siddique (1999)) as well.

3. The large first-order autocorrelation coefficients observed for some returns and time periods are well into the right tail of the prior predictive distribution, beyond the 99th percentile in one case. (See Table 4, column (5) and Figure 1(c); the largest value is for the S&P 500 returns in the 1970s.) The prior predictive distribution of the first-order autocorrelation coefficient in HMNM models with no serial correlation ($\boldsymbol{\phi} = \mathbf{0}$), not documented here, is almost the same. The reason is that for configurations of parameters in the HMNM model with substantial prior probability the sampling fluctuation in the first-order autocorrelation coefficient is substantially larger than that in the canonical i.i.d. Gaussian model. The other characteristics of the sample autocorrelation function of returns in Table 4, columns (6)-(8), pose no such problems. Figure 2(b) shows the joint prior predictive distribution for the first-order autocorreleation coefficients of daily and non-overlapping ten-day returns. It is more probable than not that the latter will exceed the former (area above and below the cross in this panel), and this is due to the greater sampling variability in the latter for which are constructed from only one-tenth the number of observations as the former.

4. Figure 2(c) provides the joint prior distribution of the first-order autocorrelation coefficient for returns and for absolute returns. The prior distribution provides much more much support for substantial positive than for substantial negative first-order autocorrelation in absolute returns. Absolute returns are more likely to be persistent than are returns, by this measure, but the probability that the absolute value of the first-order autocorrelation coefficient for returns exceeds that for absolute returns is nearly 0.30. Yet for several returns and periods, the observed combinations lie in a region of the joint prior distribution that is very sparse. For example, the observed autocorrelation coefficient of 0.26 for S&P 500 returns in 1971-75, which is extreme relative to the prior distribution (row 1 of Table 4, column (5)) also suggests that absolute returns should have been more persistent (by this measure) than was in fact the case.

5. The prior distribution of first- and ninth-order autocorrelation in one-day absolute returns supports a wide range of behavior, including the values observed in the 18

return samples. This is indicated in first two columns of Table 5, and in the bivariate prior distribution of these functions in Figure 2(d). The prior probability that the ninth-order exceeds the first-order autocorrelation coefficient is 0.34. This is all consistent with observed functions in the return samples.

6. The observed variance ratios of ten-day to one-day returns documented in column (3) of Table 2 are consistent with the prior distribution as indicated by the corresponding c.d.f. values in Table 5, column (3). A centered 98% prior credible set for this ratio is (4.1, 17.1), which well contains the observed function values.

7. Table 4, column (8), shows that the sums of the absolute values of the first 200 autocorrelation coefficients observed for returns are well within the support of the prior predictive density. Moreover, these sums are scattered over a wide range of the prior predictive even though the corresponding data appear closely concentrated in Table 1, column (8). Column (6) of Table 5 shows that the same is true for absolute returns, although in this case the corresponding data appear widely dispersed in the column (6) of Table 2. The corresponding bivariate prior distribution (not shown here) reveals no anomalies. In the case of absolute returns, the observed large fluctuations between time periods for the same return series (Table 2, column (6)) do not pose a problem for the prior predictive distribution as shown in Figure 3(d).

8. The prior predictive distribution of the ratio of the sum of the first 200 autocorrelation coefficients of absolute returns, to the sum of the absolute values of these autocorrelation coefficients, is a mixed continuous-discrete distribution. It places mass on the value 1.0, corresponding to the event that all 200 autocorrelation coefficients are non-negative, and the probability of this event is 0.048. This is the number recorded with an asterisk in column (8) of Table 5, corresponding to those observations in which the ratio was 1.0 (Table 2, column (8)). Figure 1(d) conveys some further insight into this distribution show that the sum of the first 200 autocorrelation coefficients of absolute returns can well exceed 20, and that in this case it is more probable than not that all coefficients will be positive. Column (8) of Table 5 shows that the observed values of this ratio are well within the support of the prior predictive distribution.

9. The HMNM model has no long-memory characteristics, by virtue of Theorems 2 and 3. Yet standard long memory models fit to absolute returns for the series and time periods under consideration suggest long memory, strongly so in some cases (Table 2, column (5)). As indicated in column (5) of Table 5 the prior predictive distribution of $\widehat{d}$ places all of these estimates above the median, yet only in two cases (the earliest S&P 500 return series) are the observed values in the upper decile of the prior distribution. Figures 1(b), 1(c), and 3(c) also document the inclusiveness of the prior predictive distribution for $\widehat{d}$. Thus the widely noted long-memory properties of absolute returns (Ding et al. (1993); Bollerslev and Mikkelsen (1996)) do not present any substantial obstacle to the HMNM model, in which all population moments display geometric decay. In view of the findings of Diebold and Inoue (2001) on long memory and Markov switching this is not surprising. Figure 3(c) shows that the apparent volatility of $\widehat{d}$ in Table 2 is also consistent with the complete HMNM model. In combination

with other functions of interest, however, the picture is more sobering. The S&P 500 returns in the 1970's display the largest first-order autocorrelation coefficients as well as the largest values of $\widehat{d}$. As revealed in Figure 1(c), the prior distribution renders this joint event implausible.

10. The sample correlation between $|y_t|$ and y_{t+1}, often termed the "leverage effect", is less than -0.1 for three of the seven five-year periods of S&P 500 returns studied (Table 2, column (4)). The point -0.1, in turn, corresponds to roughly the fourth percentile of the prior distribution of this function. In the prior distribution leverage appears to be essentially independent of the other functions of interest studied here; Figure 1(b) presents one example. Thus the observed low values of leverage for some samples do not interact to produce the same sort of stronger implausibility noted for $\widehat{d}$ and the first-order autocorrelation coefficient.

2.4 The posterior simulator

The posterior simulator for the HMNM model is, globally, a Gibbs sampling algorithm with seven blocks. The conditional posterior distributions and their derivation are provided in the on-line technical appendix.

The first block of the Gibbs sampling algorithm is the $T \times 2$ matrix of state assignments. The states are drawn by first removing $\mathbf{s}^2$ by analytic marginalization, drawing $\mathbf{s}^1$ conditional on all parameters (but not $\mathbf{s}^2$), and finally drawing $\mathbf{s}^2$ conditional on $\mathbf{s}^1$ and all of the parameters. The draw of $\mathbf{s}^1$ uses the algorithm of Chib (1996) for first-order hidden Markov processes. The T transitory states $\mathbf{s}^2$ are then conditionally independent.

The second block of the algorithm is the precision parameter h, whose conditional distribution is gamma. The third block, the vector $\mathbf{h}$, consists of m_1 conditionally independent draws from gamma distributions. The fourth block, the matrix $\mathbf{H}$, consists of $m_1 m_2$ conditionally independent draws from gamma distributions.

The fifth block of the algorithm is the transition matrix $\mathbf{P}$, broken down into separate sub-blocks for each row. The rows are not conditionally independent, because the vector of stationary probabilities $\boldsymbol{\pi}$ is a function of all parameters in $\mathbf{P}$, as is the orthonormal complement $\mathbf{C}_0$ of $\boldsymbol{\pi}$. These two relationships complicate what would otherwise be a multivariate beta distribution for each row. These complications are handled using a Metropolis-within-Gibbs algorithm using the multivariate beta distribution component as the source density.

The sixth block of the algorithm is the matrix $\mathbf{R}$ of transitory state probabilities, broken down into separate sub-blocks for each row $\mathbf{r}_j$. The conditional distribution of each row $\mathbf{r}_j$ would be multivariate beta distribution but for the dependence of the orthonormal complement $\mathbf{C}_j$ on $\mathbf{r}_j$. This dependence is again handled using a Metropolis-within-Gibbs algorithm using the multivariate beta distribution component as the source density.

The final block of the algorithm consists of $\boldsymbol{\beta}$, $\widetilde{\boldsymbol{\phi}}$ and $\widetilde{\boldsymbol{\psi}}$. These conditional distribution of these parameters is Gaussian.

As discussed in Section 2.1 the orthonormal complements $\mathbf{C}_j$ $(j = 0, \ldots, m_1)$ are not unique, but the non-uniqueness has no bearing on the posterior distribution. In the posterior simulation algorithm, however, it is important that $\mathbf{C}_0$ be a smooth function of $\boldsymbol{\pi}$ and that $\mathbf{C}_j$ be a smooth function of $\mathbf{r}_j$ $(j = 1, \ldots, m_1)$. Violation of this condition can lead

to unacceptably low rates of acceptance in the Metropolis-within-Gibbs algorithms. The technical appendix provides an algorithm that leads to sufficient smoothness and avoids this problem.

The posterior simulator produces an ergodic sequence of parameter draws. To see that this is so, let C be any subset of the parameter space with positive posterior probability. Then the probability that the algorithm will produce a draw in C in the next step is positive, regardless of its current position. Corollary 4.5.1 of Geweke (2005) then establishes ergodicity. The HMNM model shares with other mixture models the property that, because the posterior distribution is invariant to permutation (sometimes termed relabeling) of the latent states, there are many reflections of the posterior distributions in the parameters space – $m_1!m_2!$, in the case of the HMNM model. Geweke (2007a) shows that this condition leads to no special problems so long as there is to be no substantive interpretation of the states, as is the case here. The posterior simulator may visit only a few reflections of the posterior distribution – perhaps only one – in any reasonable number of iterations, but the information provided by the simulator about functions of interest is just as valid as if all the reflections were included in the simulation.

The posterior conditional distributions used in the algorithm, and the computer code that implements the Gibbs sampling algorithm, were verified to be correct using the procedures described in Geweke (2004) and Geweke (2005), Section 8.1.2.

Execution time for the algorithm is roughly proportional to $T\left(m_1^2 + m_2^2\right)$, as indicated in Table 6. The most time-consuming parts of the algorithm are sampling from the conditional distribution for $\mathbf{s}$, and sampling from the conditional distribution for $\boldsymbol{\beta}$, $\widetilde{\boldsymbol{\phi}}$ and $\widetilde{\boldsymbol{\psi}}$.

As is the case with almost all Markov chain Monte Carlo algorithms, there is serial correlation in the parameter draws generated by the algorithm and in functions of these parameters. Table 7 provides some information about relative numerical efficiency (RNE, Geweke (1989) or (2005), Section 4.2.2) for a key function of parameters, c.d.f. of the next day's return evaluated against the one-step-ahead predictive distribution. The distribution of RNE is with respect to 7323 executions of the posterior simulator using rolling samples of 1250 days, for the S&P 500 return, ending between December 15, 1976 and December 16, 2005 inclusive. The RNE is computed using 1000 iterations of the Gibbs sampler, and there are 4 skips between each iteration used. Thus the interpretation of RNE is that 5/RNE iterations of the sampler are required to yield the same information as one iteration of a hypothetical sampler making i.i.d. draws from the posterior distribution.

All code is written in compiled Fortran running under linux, and executed using a 2.6Ghz Hewlitt-Packard opteron processor with 4G memory. Taken together the tables indicate that use of the model on a daily basis is quite practical. For example, if every fifth iteration is used, then to obtain the same information as from 1000 draws from a hypothetical i.i.d. posterior simulator would require about 1.7×10^5 iterations. The evidence in Section 4 indicates that $m_1 = m_2 = 4$ is adequate for many financial asset return data considered in this study. For 1.7×10^5 iterations execution time for a five-year sample is under 10 minutes; a 10-year sample with $m_1 = m_2 = 6$ would require less than 40 minutes.

3 Model comparison and validation

There are many models of asset returns that compete with HMNM models. For example there are many varieties of ARCH models, including ARCH, GARCH, EGARCH, and others, and for each of these there are alternative parameterizations such as GARCH(p, q), where p and q are specified by the investigator. Similarly there are several varieties of SV models.

In principle a formal Bayesian approach using this rich variety of competing models is straightforward. Denoting the set of alternative models $A = \{A_1, \ldots, A_J\}$,

$$p\left(\boldsymbol{\omega} \mid \mathbf{y}_T^o, A\right) = \sum_{j=1}^{J} p\left(\boldsymbol{\omega} \mid \mathbf{y}_T^o, A_j\right) p\left(A_j \mid \mathbf{y}_T^o, A\right). \tag{27}$$

Given the posterior model probabilities $p\left(A_j \mid \mathbf{y}_T^o, A\right)$, all of the methods described in Section 2.1 can be applied. Section 3.1 outlines this process, describes some of the practical difficulties that arise in its implementation, and indicates how this study handles those difficulties using predictive likelihoods. Section 3.2 uses predictive likelihoods to assign probabilities to various choices of $m_1 = m_2$ for the asset return series studied. The result is that $m_1 = m_2 = 4$ and $m_1 = m_2 = 5$ are assigned the highest probability, the outcome depending on the particular asset and time period. That section also compares all of the HMNM models with competing ARCH and SV specifications. Depending on the asset return studied, the comparison varies from close (for bond returns), to modestly but clearly favoring the HMNM models (for the S&P 500 returns), to overwhelmingly favoring the HMNM models (for the dollar-pound returns). These comparisons also provide evidence suggesting that the HMNM models are more stable than the competing models examined, although these results, too, are specific to the particular asset returns under consideration.

Formal model comparison provides no evidence on the ability of any of the competing models to capture reliably the behavior of asset returns. Failure of all models in $A = \{A_1, \ldots, A_J\}$ in these dimensions is important in financial decision making. For example, if negative returns of more than 10 sample standard deviations are assigned probability 10^{-10} for the S&P 500 returns, then the decision maker would regard events like October 19, 1987, as effectively nonrecurring. In such a situation there would be a compelling need to understand as well as possible the dimensions of model failure both in order to appreciate the limitations of the models and to provide the basis for the development of models that do not have these failings. This study undertakes two kinds of model validation exercises designed to address this need. Section 3.3 reviews some of the characteristics of the predictive distributions of any well-specified model of economic time series, examines the predictive distributions of asset returns in the HMNM model for departures from these norms, and finds systematic evidence of failure. Section 3.4 describes an alternative model validation exercise using posterior predictive distributions of the 16 functions of interest of the data identified in Section 1.1.1. The HMNM model performs well for most of these functions, but there are a few in which it falls short. The concluding section of the article draws implications of these findings for future research.

3.1 Predictive likelihoods

The critical terms combining models in (27) are the posterior model probabilities

$$p\left(A_j \mid \mathbf{y}_T^o, A_j\right) \propto p\left(A_j \mid A\right) \cdot p\left(\mathbf{y}_T^o \mid A_j\right) = p\left(A_j \mid A\right) \cdot \int_{\boldsymbol{\Theta}} p\left(\boldsymbol{\theta}_{A_j} \mid A_j\right) p\left(\mathbf{y}^o \mid \boldsymbol{\theta}_{A_j}, A_j\right) d\boldsymbol{\theta}_{A_j},$$

the term $p\left(\mathbf{y}_T^o \mid A_j\right)$ often being called the marginal likelihood of model A_j. Approximating the integral that comprises $p\left(\mathbf{y}_T^o \mid A_j\right)$ using simulation (or any other) methods is considerably more difficult than approximating the posterior distribution of a vector of interest $\boldsymbol{\omega}$ in a single model. (Geweke (2005), Section 8.2, discusses several approaches and reviews the literature.) Apart from computational problems, the choice of the prior distribution $p\left(\boldsymbol{\theta}_{A_j} \mid A_j\right)$ presents problems for the marginal likelihood that do not arise for any model-specific posterior density $p\left(\boldsymbol{\omega} \mid \mathbf{y}_T^o, A_j\right)$. One is that for a sequence of increasingly diffuse prior distribution whose limit is an improper prior, the limiting marginal likelihood is zero regardless of the success of the posterior distribution in describing the properties of $\mathbf{y}_T^o$; this is the well-known Bartlett (1957) paradox (Geweke (2005), Section 2.6.2). A secondary but still important difficulty is that proper prior densities $p\left(\boldsymbol{\theta}_{A_j} \mid A_j\right)$ are always chosen with reference to the data $\mathbf{y}_T^o$, as was the case here in the prior predictive analysis of Section 2.3.1. The model-specific posterior density $p\left(\boldsymbol{\omega} \mid \mathbf{y}_T^o, A_j\right)$ tends not to be very sensitive to these choices, except for very small sample sizes T, whereas marginal likelihoods can be extremely sensitive.

The marginal likelihood can be decomposed

$$p\left(\mathbf{y}_T^o \mid A_j\right) = \prod_{t=1}^{T} p\left(y_t^o \mid y_{t-1}^o, \ldots, y_1^o, A_j\right), \tag{28}$$

where each term in the product is the one-step-ahead predictive likelihood for the data point y_t^o. Based on the output $\boldsymbol{\theta}_{A_j}^{(m)}$ $(m = 1, \ldots, M)$ of a posterior simulator using a subset $\left\{y_q^o, \ldots, y_{t-1}^o\right\}$ of the data,

$$p\left(y_t^o \mid y_{t-1}^o, \ldots, y_q^o\right) \approxeq M^{-1} \sum_{m=1}^{M} p\left(y_t^o \mid y_{t-1}^o, \ldots, y_q^o, \boldsymbol{\theta}_{A_j}^{(m)}, A_j\right); \tag{29}$$

the next section discusses choice of q. In the HMNM model the term inside the sum in (29) can be evaluated efficiently using the algorithm of Chib (1996) to analytically integrate over the latent states $\mathbf{s}_q, \ldots, \mathbf{s}_{t-1}$.

The marginal likelihood (28) could be approximated in this way. The extreme sensitivity to the prior distribution is concentrated in the earlier terms of the product and declines rapidly in later terms. These predictive likelihoods provide an alternative method of model comparison that is directly related to the posterior model probabilities, but is less sensitive to the prior distribution. Due to the efficiency of the posterior simulator these terms can be approximated reliably using (29) in reasonable time.

3.2 Formal model comparison using predictive likelihoods

We present formal model comparisons using the four asset return samples described in Section 1.1. In each case we use the sample in two different ways. *Rolling samples* utilize only the most recent 1250 trading days (about five years) of asset returns, and the basis of comparison is

$$\sum_{t=1251}^{T} \log p\left(y_t^o \mid y_{t-1}^o, \ldots, y_{t-1250}^o, A\right), \tag{30}$$

where $t = 1$ denotes the start of the asset return series (e.g., the first trading day of 1972 for the S&P 500 returns). The computations utilize the approximation (29) with $q = t - 1250$ and $\boldsymbol{\theta}_A^{(m)}$ $(m = 1, \ldots, M)$ the output of the posterior simulator using the 1250 observations $y_{t-1250}^o, \ldots, y_{t-1}^o$. *Building samples* utilize the entire sample beginning with the start of the asset return series, and the basis of comparison is

$$\sum_{t=1251}^{T} \log p\left(y_t^o \mid y_{t-1}^o, \ldots, y_1^o, A\right), \tag{31}$$

utilizing the approximation (29) with $q = 1$ and $\boldsymbol{\theta}_A^{(m)}$ $(m = 1, \ldots, M)$ the output of the posterior simulator using the $t - 1$ observations $y_1^o, \ldots, y_{t-1}^o$.

We utilize both rolling and building samples because comparison of the results from the alternative sampling schemes is informative in several dimensions. First, if the data generating process implies a stable model A, then the log predictive likelihood (31) should exceed the log predictive likelihood (30) for the simple reason that the former uses more information than the latter; whereas instability for the model A would be manifest as a smaller difference and could even give rise to values of (30) that exceed (31). Thus the comparisons provide some evidence on model stability. Second, since computation time is roughly proportional to sample size for large samples, the rolling samples provide for faster execution than do the building samples, and a small difference in the alternative log predictive likelihoods would suggest that there might be little given up in using a sample truncated to recent years. Finally, there is a long tradition in the asset return literature of using rolling samples.

This comparison exercise uses five alternatives to the HMNM model: a Gaussian i.i.d. model; a Gaussian GARCH(1,1) model, a Gaussian exponential GARCH(1,1) (EGARCH(1,1); Nelson (1991)) model, a GARCH(1,1) model with i.i.d. Student t shocks and the SV model of Jacquier et al. (1994),

$$y_t = \mu_x + \exp\left(x_t/2\right)\varepsilon_t, \;\; (x_t - \mu_x) = \phi(x_{t-1} - \mu_x) + \sigma\eta_t \tag{32}$$

with ε_t and η_t uncorrelated standard normal. See also equation (4) in Durham (2005), from which the model we use differs only in allowing y_t to have a mean different from zero (μ_x). For the SV and HMNM models we use exactly the procedure just described for the computation of log predictive likelihoods with rolling and building samples. For the Gaussian i.i.d. and GARCH models we use maximum likelihood estimates $\widehat{\boldsymbol{\theta}}_A$, taking

$$p\left(y_t^o \mid y_{t-1}^o, \ldots, y_{t-1250}^o, A\right) = p\left(y_t^o \mid y_{t-1}^o, \ldots, y_{t-1250}^o, \widehat{\boldsymbol{\theta}}_A, A\right), \tag{33}$$

where $\widehat{\boldsymbol{\theta}}_A$ is the maximum likelihood estimate from the observations $y^o_{t-1250}, \ldots, y^o_{t-1}$, and similarly for

$$p\left(y^o_t \mid y^o_{t-1}, \ldots, y^o_1, A\right) = p\left(y^o_t \mid y^o_{t-1}, \ldots, y^o_1, \widehat{\boldsymbol{\theta}}_A, A\right) \tag{34}$$

where $\widehat{\boldsymbol{\theta}}_A$ is the maximum likelihood estimate from the observations $y^o_1, \ldots, y^o_{t-1}$.

Table 8 provides the results of these exercises for the asset return series, comparing the competing models just described and an array of HMNM models. The first column indicates the model, HMNM denoting the hierarchical Markov normal mixture models and MNM the conventional Markov normal mixture models. The contrast between the Gaussian i.i.d. and Gaussian GARCH(1,1) models provides a useful benchmark because the deficiencies of the former are quite well known while the latter is the most widely used alternative for asset returns.

Turning first to the S&P 500 returns and the rolling samples, the highest log predictive likelihood among the mixture models is the HMNM model with $m_1 = m_2 = 4$. The t-GARCH model provides a slight improvement over the best HMNM model. Imposition of no serial correlation leads to a slight deterioration in the log predictive likelihoods for HMNM models and a slight improvement for most MNM models, but these differences are small. All of these HMNM models have log predictive likelihoods than exceed those of the SV model, EGARCH and GARCH models.

Using building rather than rolling samples substantially improves the log predictive likelihoods of the HMNM models, moreso when serial correlatoin is permitted (about 200 log-likelihood points) than when it is excluded (about 100 points). This is consistent with stability of the mixture models for the S&P 500 return data generating process. Except for the t-GARCH model, all of the competing models show improved predictive performance with building samples, but by lower margins: no improvement exceeds 100 points. Consequently HMNM models have log predictive likelihoods exceeding those of all competing models, with the best HMNM models surpassing all of the competing models by over 200 points. HMNM models incorporating serial correlation and with building samples lead to the largest log predictive likelihoods among all the models and the two sampling methods studied. MNM model log predictive likelihoods fall short of those for HMNM, t-GARCH and SV models.

The results for the dollar-pound exchange returns differ substantially from those for the S&P 500 returns. For the rolling samples the imposition of no serial correlation in the population increases the HMNM model log predictive likelihoods, whereas those for the MNM models are scarcely affected. As with the S&P 500 returns, the HMNM models outperform the MNM models. The HMNM model with the highest log predictive likelihood ($m_1 = m_2 = 5$, no serial correlation) exceeds that of the closest competitor (t-GARCH, again) by over 300 points. The evidence very strongly favors the MNM models over competitors.

Using building rather than rolling samples makes the comparison even more striking. Except for t-GARCH the log predictive likelihoods of the competing models all deteriorate, moving from rolling to building samples. By contrast the best HMNM models improve dramatically, by over 200 points in some cases. The result is that, with building samples, the best HMNM models overwhelmingly outperform the competitors: the best HMNM

model outperforms the closest competing non-mixture model (t-GARCH) by over 500 log-likelihood points – a difference greater than that between GARCH and an i.i.d. normal model.

Comparison of the log predictive likelihoods for the bond returns, in Table 8, shows yet a different set of contrasts. The t-GARCH model is uniformly the strongest competitor, followed by the SV model. For returns of both maturities the log predictive likelihoods for the HMNM models are very close to t-GARCH, with the best HMNM models performing slightly better in the case of one-year maturity and slightly worse in the case of ten-year maturity. With only a few exceptions building samples produce somewhat higher predictive likelihoods than rolling samples for all models. The muted comparison may be due in part to the short sample size, but this alone seems unlikely to account for the similarity in results using rolling and building samples.

These comparisons use only one, simple, SV model (32), whereas there is a wide variety in the literature. Recent work by Durham (2006, 2007) enables us to project these comparisons to quite a few of these SV models for the S&P 500 return series. Durham (2007) generalizes (32) by taking $\eta_t = \rho\varepsilon_t + (1-\rho^2)^{1/2} u_t$, $u_t \sim N(0,1)$, and then specifying that the distribution of ε_t is a mixture of three normal distributions. He estimates this model by maximum likelihood using the S&P 500 returns for the period June 25, 1980 through September 20, 2002, and for comparison with some other studies, the period January 2, 1980 through December 31, 1999. Line 1 of Table 9 shows the maximized value of the log-likelihood function for this model using both data sets. We constructed log predictive likelihoods for the same periods, using both rolling and building samples, and for the HMNM models $m_1 = m_2 = 4$ and $m_1 = m_2 = 5$. These comparisons are biased against the HMNM models, because the predictive likelihood is out-of-sample, whereas the maximized likelihood is in-sample. The HMNM building samples used for inference are slightly larger than the SV samples used for estimation, with data going back to 1972 rather than 1980, but this difference should be minor in comparison. The HMNM models with building samples outperform Durham's generalization of the SV model by over 100 log-likelihood points.

Durham (2006, 2007) undertakes the same maximum likelihood computations using five variants of SV models, including both one- and two-factor models, the jump-dispersion models studied in Eraker et al. (2003), and a variant in which the distribution of ε_t in (32) is Student-t. These computations are also undertaken using for variants of affine models. These studies find that the mixture model of Durham (2007) produces the largest value of the maximized log likelihood function in both samples. In the June 25, 1980 through September 20, 2002 sample the difference between the maximized log likelihood for the mixture model and those for the SV models ranges from 28.7 to 60.6, whereas for the single affine model reported the difference is 131.1. In the 1980 - 1999 sample the difference ranges from 21.8 to 69.5 for the SV models and 48.6 to 120.5 for the affine models. We believe it would be worthwhile (though beyond the scope of this study) to compute the predictive likelihoods for these models, for reasons indicated in Section 5. There seems little doubt that these comparisons would favor HMNM models using building samples, but only such formal comparison would reveal the magnitude of the dominance and relationship of the SV and affine models to the ARCH models.

Overall, these formal comparisons suggest two working conclusions for future substan-

tiation through similar work with other asset returns. First, the predictive performance of HMNM models is generally at least as good as that of ARCH and SV models (as in the case of bond returns) but can be better (as in the case of the S&P 500 returns), sometimes substantially so (as in the case of dollar-pound returns). For S&P 500 returns, at least, existing evidence suggests that HMNM models will outperform variants of SV models not considered here, as well as affine models. Second, the predictive performance of HMNM models is improved by using time series of asset returns spanning several decades (as in the case of S&P 500 and dollar-pound returns), whereas the performance of ARCH and SV models is not similarly enhanced.

3.3 Out of sample model validation

For the true conditional c.d.f. $F\left[y_t \mid (y_{t-s}, s>0)\right]$ the random variable $u(t) = F\left[y_t \mid (y_{t-s}, s>0)\right]$ is i.i.d. uniform $(0,1)$ (Rosenblatt (1952)). For presentation of evidence about departures of $u(t)$ from this norm it is more convenient to work with $z(t) = \Phi^{-1}\left[u(t)\right]$, which under the same conditions is i.i.d. $N(0,1)$ (Berkowitz (2001)). For multiple-day returns $w_{j,t} = \sum_{s=0}^{j-1} y_{t+s}$, if F_j is the conditional c.d.f. $F_j\left[w_{j,t} \mid (y_{t-s}, s>0)\right]$, then $u_j(t) = F_j\left[w_{j,t} \mid (y_{t-s}, s>0)\right]$ is uniform $(0,1)$, and $z_j(t) = \Phi^{-1}\left[u_j(t)\right]$ is $N(0,1)$ but follows a moving average process of order $j-1$. While no model provides exactly the conditional c.d.f. (because all models are false, and because the pseudo-true parameter values of these models are unknown) the departures of a model's implications for $u(t)$ and $z_j(t)$ provide a convenient perspective on model misspecification that is also directly relevant in prediction applications of the kind studied in Section 4.

We investigated these implications for HMNM models as a by-product of the computations undertaken to compute the log predictive likelihoods discussed in the previous section. For the one-step-ahead predictive distributions this entails simply evaluating the c.d.f. as well as the pdf of $y_t^o \mid \left(y_{t-1}^o, \ldots, y_q^o, \boldsymbol{\theta}_A^{(m)}, A\right)$ at each parameter vector $\boldsymbol{\theta}_A^{(m)}$ drawn from the posterior distribution from the sample $y_q^o, \ldots, y_{t-1}^o$, where $q = t-1250$ for rolling samples and $q = 1$ for building samples. (This is facilitated by the algorithm of Chib (1996) which yields exact state probabilities at time $t-1$ conditional on the sample and $\boldsymbol{\theta}_A^{(m)}$.) We also investigated the distributional implications for two- through ten-day returns by simulating the next ten days of returns once for each parameter vector $\boldsymbol{\theta}^{(m)}$ drawn from the posterior and then using the empirical c.d.f. of the simulations to approximate $F_j\left(w_{j,t}^o \mid y_{t-1}^o, \ldots, y_q^o, \boldsymbol{\theta}_A^{(m)}, A\right)$.

Tables 10 and 11 provide the results of these out of sample model validation exercises for the four asset return series used in this study. For each return four columns show results for rolling and building samples, and for one- and ten-day returns. The first four lines of each table provide the first four moments of the corresponding random variables $z(t)$ and $z_{10}(t)$, followed by the first-order autocorrelation coefficient of $z(t)$. As a convenient benchmark, rather than a test of the false hypothesis of perfect calibration, an asterisk denotes an outcome significantly different from the i.i.d. $N(0,1)$ norm at the 0.1% level. The next 10 lines indicate the fraction of observations falling in each predictive decile, and the last line provides the difference between the largest and smallest of these fractions.

The results for the S&P 500 and dollar-pound returns in Table 10 are similar in several respects. First, the distributions over deciles in the lower part of each table appear excellent,

with only a few entries outside the range $(0.09, 0.11)$. Second, not surprisingly, the calibration of predictive distributions for one-day returns is better than for ten-day returns in all cases. Third, the random variables $z(t)$ are decidedly positively autocorrelated. Finally, the out-of-sample predictive distributions based on rolling samples appear to be better calibrated than those based on building samples. The last result is somewhat surprising given the superior predictive likelihoods for building samples documented in Table 8, as well as the other evidence indicating stability of the HMNM models. The main difference in the results for these two asset returns is that the predictive distributions for the S&P 500 ten-day returns are somewhat over-dispersed, with too few observations in the tails, whereas the opposite is the case with the dollar-pound exchange returns, both one- and ten-day.

The results for the bond returns in Table 11 are also similar, in important respects, that in turn differ from the findings with the S&P 500 and dollar-pound returns. The distributions over deciles are not as close to their theoretical norms as was the case in Tables 10 and 11. This is due to some extent to the smaller size of the samples for the bond returns. Ten-day returns are again more poorly calibrated than one-day returns, as one might expect. There is positive serial dependence in the transformed inverse predictive c.d.f. $z(t)$, but it is weaker than for the stock and foreign exchanged series studied. The contrast between the rolling and building samples for the bond series is much less striking than was the case for the S&P 500 and dollar-pound return series. There is evidence of over-dispersion of predictive intervals for both bond returns, especially evident in the right tail for ten-day returns.

Like the model comparison exercise, perhaps the most striking feature of the out-of-sample model validations is the specificity of results for each series. Overall these may be attributed to insufficient flexibility in the HMNM model. More constructively, the persistent finding of positive autocorrelation in $z(t)$ is consistent with the limitations on dynamic properties established in the theoretical work in Section 2.2. Relative to the conventional Markov normal mixture model the additional flexibility of the HMNM resides in its second layer, which provides much greater flexibility than the normal distribution, and the calibration of one-step-ahead predictive densities appears excellent. The out-of-sample model prediction exercise thus suggests that a fruitful extension of the HMNM model might be in providing dynamic flexibility in the second layer of the model.

3.4 In sample model validation

Section 2.3 used the prior predictive distribution to describe the implications of the complete HMNM model by means of observable functions z_j like those described in 1.1. The same approach can be used to understand the implications of the posterior distribution of the model parameters. The mechanics of simulating observable functions of returns from the posterior predictive distribution are exactly the same as in the case of the prior predictive distribution, except that the parameters are drawn from the posterior distribution rather than the prior distribution.

This section uses this approach for the four asset return series. For each series, the posterior simulator produces a sample of parameter vectors $\boldsymbol{\theta}_A^{(m)}$ using the MCMC algorithm described in Section 2.4. Conditional on each draw, five years (1250 observations) are

simulated and the implied functions of interest $z_j^{(m)}$ are computed. Comparison of the posterior predictive distribution with sample values can reveal limitations in the model, in the same way as the prior predictive distribution.

The posterior predictive analysis in this section uses the model $m_1 = m_2 = 4$, allowing serial correlation, for the S&P 500 returns; the model $m_1 = m_2 = 5$, excluding serial correlation, for the dollar-pound returns; and the model $m_1 = m_2 = 5$, allowing serial correlation, for the bond returns. In each case the posterior distribution conditions on the entire sample (described in Section 1.1), and observed characteristics z_j^o pertain to the same subperiods introduced in Tables 1 and 2.

These posterior predictive distributions can be studied from the same two perspectives used for the prior predictive distribution in Section 2.3. Tables 12 and 13 provide the posterior c.d.f.s evaluated at the observed values z_j^o, parallel to the prior predictive analysis in Tables 4 and 5. Figures 4 through 7 display the posterior distributions of the same combinations of functions of interest used in Figure 1. Each of these four figures corresponds to one of the panels in Figure 1, and, like that figure, indicate observed function values with open symbols and draws from the posterior predictive distribution as closed dots. Similarly Figures 8 through 11 show posterior predictive distributions corresponding to the prior predictive distributions in each of the four panels of Figure 2, and Figures 12 through 15 display posterior predictive distributions corresponding to the four panels of Figure 3.

Comparing Figures 4 through 7 with Figure 1 it is evident that, in general, the posterior predictive distributions are more concentrated than the prior predictive distributions, although never strikingly so. This is because the posterior predictive distributions pertain to characteristics observable in a five-year period that would be random even in the hypothetical limiting case in which the model and parameters values were known with certainty. Comparing Tables 12 and 13 with Tables 4 and 5 this phenomenon is evident in the fact that for a given financial return series the c.d.f. evaluations in the former tables are typically more dispersed. For example, the evaluations under the prior predictive c.d.f. of all but the lowest of the seven excess skewness coefficients for the S&P 500 returns range from 0.320 to 0.702 (Table 4, column (1)) whereas under the posterior predictive c.d.f. they range from 0.199 to 0.799 (Table 12, column (1)).

The posterior predictive distributions identify difficulties with the same functions of interest and return series as did the prior predictive analysis, but the problems are somewhat more sharply defined. The three most notable cases all arise with the S&P 500 return series. First, the model does not account well for the sample autocorrelation of this time series in the 1970s (Table 12, column (5), and Figure 6(a)). Owing to the comparative tightness of the posterior distribution, these statistics are even farther in the right tail of the posterior distribution than they were in the prior distribution. Second, the leverage statistics (i.e., the sample correlation of yesterday's return and today's absolute return) for the S&P 500 series are much farther in the left tail of the posterior distribution than of the prior distribution in the periods 1986-90, 1996-2000 and 2001-05 (see Table 13, column (4), and Figure 5(a)). Finally, the GPH estimate of the long memory parameter in the early 1970s is much farther in the right tail of the posterior distribution (Table 13, column (5), and Figures 5(a) and 6(a)) than was the case with the prior distribution.

This documentation of the posterior predictive distribution bears on the consistency of

the HMNM model with the ten aspects of asset returns first discussed in Section 1.1.2.

1. The model accounts well for substantial variations in skewness and kurtosis across returns series and time periods; see Table 12, columns (1)-(4), and Figure 4. The posterior distributions of these sample statistics differ notably from one series to the next, as well as from the prior distribution (Table 4, columns (1)-(4), and Figure 1(a)). The skewness coefficient for 10-day S&P 500 returns in 1986-1990 is an extreme event. The apparent volatility in skewness and kurtosis evident in Table 1, columns (1) and (2), does not present a problem under the posterior predictive distribution, as documented in Figures 12 and 13. Comparison of these figures with their respective counterparts under the prior predictive distribution, panels (a) and (b) of Figure 3, highlights an interesting difference in the posterior and prior dynamics. Relative to the prior, the posterior eliminates much of the support for transitions between high values of skewness and excess kurtosis for successive five-year periods, placing weight instead on transitions between rather small values and extreme values.

2. Figure 8 shows that all of the posterior distributions place substantial probability on the event that, over a five-year period, the sample skewness coefficient for ten-day returns exceeds that for one-day returns (the regions above and below the crossed lines in these figures) as well as the opposite (the regions left and right of the crossed lines). Comparison of the Figure 2(a) with Figure 8 shows that at least in the context of the HMNM model the samples are not very informative about skewness that is likely to be observed in these asset returns.

3. The size of the first-order autocorrelation coefficients for S&P 500 returns in the 1970s is even more anomalous in the posterior predictive distribution than it is in the prior predictive distribution: compare column (5) of Tables 4 and 12, Figure 1(c) with Figure 6(a), and Figure 2(b) with Figure 9(a). The latter comparison shows that the posterior predictive distribution removes essentially all support for the first order auto correlation coefficient in the early 1970s, and almost all in the latter 1970s. On the other hand, it also shows that there is no mystery in the fact that the first-order autocorrelation coefficient for ten-day returns exceeds that for daily returns, in absolute value, about as often as not in these data. It is due to the greater sampling fluctuation in the former relative to the latter.

4. Figure 10 provides the joint posterior predictive distributions of the first-order autocorrelation coefficient for returns and absolute returns, complementing Figure 2(c) for the joint prior predictive distribution. The support provided by the prior predictive distribution (Figure 2(c)) to negative first-order autocorrelation of absolute returns is almost completely removed in the posterior predictive distribution of the S&P 500 and dollar - pound returns, and is substantially reduced for the bond return series. The poor handling of first-order autocorrelation of S&P 500 returns in the l970s previously noted is striking in Figure 10(a).

5. The joint posterior predictive distributions account well for the observed first- and ninth-order autocorrelation coefficients for absolute returns; see Table 13, columns (1)

and (2), and Figure 11. The joint distribution for bond returns reflects the lower persistence in volatility observed for those series. In all four posterior distributions it is more probable than not that the first-order autocorrelation coefficient of absolute returns will exceed that for returns, but the complementary event always has substantial probability.

6. Table 13, column (3), shows that the posterior distribution accounts well for the observed variance ratios of ten-day to one-day returns. In the case of the S&P 500 returns this ratio is near the high end in the 1970s and the low end in the late 1990s. In general, the observed ratios exceeding ten (Table 2 column (3)) do not pose problems for the posterior predictive distributions.

7. Comparison of Table 4, column (8), with the corresponding column of Table 12 shows that the posterior predictive distribution for the sum of absolute values of autocorrelation coefficients for returns is substantially tighter than the prior predictive distribution. The same is true for absolute returns: compare column (6) of Tables 5 and 13. Thus the posterior predictive distributions account simultaneously for the similar values of the former, across return series and time periods, and the relatively dissimilar values of the latter. The large changes over time periods in this statistic (Table 2, column (6)) challenges the posterior predictive distribution in the case of the dollar-pound returns, as exhibited Figure 15(b).

8. The posterior predictive distributions assign positive probability to the event that all of the first 200 sample autocorrelation coefficients for absolute returns will be nonnegative: this is evident in column (8) of Table 13 for the S&P 500 and dollar-pound return series, and in Figure 7 for all four series. At the same time, the posterior predictive distribution for the dollar-pound return series also accommodates the near-zero sum of these autocorrelation coefficients in the 1987-1991 subsample.

9. Relative to the prior predictive distribution, the posterior predictive distributions of the long-memory parameter estimate $\widehat{d}$ are centered and tighter. Whereas all 18 observed values were above the median of the prior predictive distribution (Table 5, column (5)), 8 observed values are below the medians of the posterior predictive distributions (Table 13, column (5)). For the S&P 500 return series in the early 1970s the posterior predictive distribution places the observed value in the far right tail, at the 99.5 percentile. This is the period presenting difficulties with respect to the first-order autocorrelation coefficient of returns, and Figure 6(a) shows that the joint event is indeed anomalous interpreted under the posterior predictive distribution. On the other hand, Figure 14 shows that the apparent instability of $\widehat{d}$ in Table 2, column (5), and noted by Ding and Granger (1996) for S&P 500 returns, is entirely plausible in this model.

10. The leverage effect – i.e., a negative sample correlation coefficient for returns in one period and absolute returns in the next – poses an even greater problem with respect to the posterior predictive distribution than the prior predictive. Table 13, column (4), indicates that values are well below the first percentile in three of seven periods for the

S&P 500 return series and in one of three periods for one-year maturity bond returns. As in the prior predictive distribution, leverage appears to be nearly independent of the other functions studied: Figure 5 provides one example.

Taken together these detailed findings bear on our understanding of asset returns and have implications for fruitful future research strategy. Section 5 takes up these conclusions.

4 Using the model in volatile times

As described in Section 1.2 it is generally straightforward to produce a random sample from a predictive distribution

$$p\left(y_{T+1},\ldots,y_{T+f} \mid y_q^o,\ldots,y_T^o,A\right) \tag{35}$$

so long as a posterior simulator is available; for a more detailed discussion see Geweke and Whiteman (2006), Section 3.3. For the models A described in this article it takes only a few minutes to generate thousands of draws from (35), and so this is a practical procedure for application on a daily basis. This section studies predictive distributions using the four return series studied, for some interesting days T and a two-week prediction horizon $f = 10$ in each case.

These exercises all use rolling samples of size 1250, $q = T - 1249$ in (35). A typical application selects an interesting interval of days $[T_1, T_2]$ for study. The exercise begins with a draw from the prior distribution, followed by 1,000 iterations of the MCMC algorithm using the return data $\left\{y_{T_1-1249}^o,\ldots,y_{T_1}^o\right\}$. It then executes an identical procedure using successively the samples $\left\{y_{T_1-1249}^o,\ldots,y_{T_1}^o\right\}$, ..., $\left\{y_{T_2-1249}^o,\ldots,y_{T_2}^o\right\}$. This procedure draws and discards 1,000 iterations beginning with the last parameter vector drawn from the previous sample (the last of the initial 1,000 iterations for the first sample). Then it draws 200,000 parameter vectors $\boldsymbol{\theta}_A^{(m)}$ using the MCMC algorithm, and for every 20'th draw it successively samples returns y_{t+j} for the next ten trading days using (2) with the current draw $\boldsymbol{\theta}_A^{(m)}$ of the parameter vector in place of $\boldsymbol{\theta}_A$. After the final parameter vector draw for the sample, the procedure assesses the convergence of the MCMC algorithm using the separated partial means test (Geweke (2005), Section 4.7) using as functions of interest the p.d.f. and c.d.f.

$$p\left(y_{T+1}^o \mid y_1^o,\ldots,y_T^o,\boldsymbol{\theta}^{(m)},A\right) \text{ and } P\left(y_{T+1}^o \mid y_1^o,\ldots,y_T^o,\boldsymbol{\theta}^{(m)},A\right),$$

both of which are easy to evaluate analytically. If the $|z|$-score of either test exceeds 3, then the entire 200,000 draws are discarded and the procedures is repeated beginning with the last draw that was dropped; this is repeated until both $|z|$-scores are less than 3.

At the completion of the exercise, for each trading day $T \in [T_1, T_2]$ there is a sample of size 10,000 drawn from the predictive density (35) for the percent log returns $y_{T+1},\ldots,y_{T+10}$. The corresponding percent log returns at daily rates are $(y_{T+s}/s,\ s = 1,\ldots,10)$. Sorting these returns then provides quantiles for the predictive distributions for the 1-day-ahead, 2-day-ahead, ... , 10-day-ahead predictive distributions.

Figure 16 provides some of these results from one such exercise using the S&P 500 returns and the four successive trading days September 25, 28, 29 and 30 of 1992. This is

an unremarkable period, chosen because it is subsequently useful in Section 4.1, and which therefore illustrates the kinds of results one might routinely obtain on a day-to-day basis. The organization of this figure is identical to those that follow in this section. The solid line indicates the actual percent log returns at daily rates: for example, the value of the line corresponding to "3" on the horizontal axis in the upper left panel is the percent log return at a daily rate for the S&P 500 index from the close of trading on September 25 to the close of trading three trading days later on September 30. The other symbols in the graph indicate quantiles of the predictive c.d.f.s over each of the 10 horizons. For each horizon indicated on the horizontal axis the solid dot provides the median, the $\times$ symbols provide the 0.25 and 0.75 quantiles, the plus symbols the 0.10 and 0.90 quantiles, the circles the 0.05 and 0.95 quantiles, and the asterisks the 0.01 and 0.99 quantiles.

Figure 16 immediately provides the return at risk for various horizons in each trading day. For example, 5% percent log return at risk for a one-day horizon (the left-most lower circle in each panel) is about 1.25 on all four days, while the 5% percent log return at risk for a ten-day horizon is about 0.40 on September 25, declining to about 0.37 on September 30. (Since the panels show returns at daily rates, the latter figures should be multiplied by 10 for total return at risk over the 10-day horizon.) These very slow movements are typical of most days in the asset return samples, consistent with the information in realized returns leading to almost no up-dating of the subjective distribution of holding period returns. In turn, this is consistent with little updating of the posterior distribution as one day leaves the sample and another enters it, moving through the successive panels in Figure 16, as well as with no substantial changes in state probabilities conditional on parameters and observed returns.

4.1 Stock returns

On Monday, October 19, 1987, the S&P 500 index daily percent log return was -20.47, the largest absolute return since the index began in 1885. On the preceding trading day, Friday, October 16, the daily percent log return was -5.16, the largest one-day decline since 1962. The largest absolute daily percent log return in the 1250 trading days (the size of the rolling sample) preceding October 19 was -4.81 (September 11, 1986).

Figure 17 portrays the quantiles of the predictive distributions of percent log returns from the close of trade on Thursday, October 15 (panel (a)) through the close of trade on Tuesday, October 20 (panel (d)). The quantiles on October 15 are similar to those in Figure 16: the 5% percent log return at risk for a one-day horizon is 1.34, although the appearance is different owing to the necessarily dissimilar vertical scales in Figures 16 and 17.

The October 16 decline is well beyond the one percent return at risk at the close of trade on October 15 (panel (a), left vertical axis). Friday's percent log return of -5.16 has a dramatic impact on the predictive distributions (panel (b)). The dispersion increases, with the interquartile range increasing from 1.16 to 3.03, a factor of 2.61. On the other hand, the centered 90% range increases from 3.60 to 12.45, a factor of 3.45, indicating increasing tail thickness. The median also declines, from 0.09 to -0.77. At the close of trading on October 16 the 0.01 quantile for the one-day percent log return is slightly less than -50 and the 0.99 quantile is slightly more than +50, both well off the scale of Figure 17(b).

Close comparison of panels (a) and (b) also reveals changes in the term structure of the predictive distributions between October 15 and October 16: the interquartile range shrinks with increasing horizon on both dates, but whereas the centered 90% range displays this behavior on October 15 it is nearly absent on October 16.

The October 19 daily percent log return of -20.47 is of course much larger in absolute value than the corresponding October 16 return. But in the context of the October 16 return (and those of preceding days) it is less surprising than was the October 16 return in the context of returns preceding that date. As indicated in panel (b) the actual loss on October 19 was less than the one-percent return at risk as of the close of trading on October 16. (The October 16 one-day percent log return predictive c.d.f, evaluated at -20.47, is in fact about 0.0185.)

The events of October 19 further increase the spread of the predictive distribution, immediately evident in panel (c). The interquartile range of the one-day predictive distribution increases from 3.03 to 6.77, a factor of 2.23, and the centered 90% range from 12.45 to 57.23, a factor of 4.60. By the former measure the proportionate increase in dispersion following October 19 is less than corresponding increase on October 16; by the latter measure it is greater. By implication, tail thickness of the predictive distribution continues to increase after October 19 just as it did after October 18.

The percent log return of 5.33 on October 20 is extremely high by historical standards, but there is a modest decrease in the dispersion of the predictive distribution updated at the close of trading on October 20, evident in panel (d). The following day's percent log return, the highest since 1940 at 9.10, is in the 86th percentile of this distribution. Both tail thickness and the tendency of tail thickness to increase with prediction persist in the October 20 predictive distributions.

In the following weeks predictive distributions slowly return to the more typical pattern, as illustrated in Figure 18. This figure displays quantiles of predictive distributions at 25-trading-day intervals following October 19, using the same vertical scale employed in Figure 16. Inspection of these figures shows that both dispersion and tail thickness steadily decrease. One hundred trading days after October 19, panel (d), the predictive distributions are similar to those in Figure 16.

The returns on the four successive trading days of October 16, 19, 20 and 21 – the first two negative and the last two positive – were all extraordinary. The first two, especially, had dramatic impacts on the predictive distribution. There are two principal channels through which this change could occur. At one extreme, the changes in the predictive distributions could be due entirely to changes in the posterior distribution of the parameters driven by the new and extreme data. At the other extreme, the returns could have negligible impact on the posterior distribution of the parameters but produce substantial changes in the state probabilities conditional on parameters and the history of returns. If the first channel were important, then there should be noticeable changes in the predictive distribution 1250 trading days in the future as the 1987 dates drop out of the rolling sample of size 1250. The October 19, 1987 return is included in the samples used on September 25 and 28, 1992 (Figure 16 panels (a) and (b)) but not in the samples used on September 29 and 30, 1992 (panels (c) and (d)). Since there is almost no change in the predictive distributions across the four panels of Figure 16, changes in the posterior distribution of parameters cannot

account in any significant way for the changes observed in the four panels of Figure 17. This is consistent with evidence previously examined that suggests the HMNM model is stable for these asset return time series, both absolutely and in comparison with competing models.

4.2 Dollar-pound returns

In September, 1992, the UK ended its participation in the European Exchange Rate mechanism. Largely as the result of intense adverse speculation about the decision of the UK monetary authorities, the dollar-pound exchange rate increased from 0.4993 on Tuesday, September 8 to 0.5839 on Monday, September 21. (We employ the usual convention that the dollar-pound exchange rate is quoted in dollars per pound, unlike all other exchange rates that are quoted in units of foreign currency per US dollar. Thus when the pound falls relative to the dollar, returns are positive rather than negative.)

The combination of speculation and intervention by the monetary authorities resulted in large but not volatile changes over the September 8 through 21 period. The daily percent log return was positive each day except September 10 when it was -0.04, never exceeded 3.30, and averaged 1.74 over the nine-day period. Such episodes are not unprecedented in foreign currency markets: there was a similar episode for the pound in March 1985 when the average percent log depreciation was 1.46 over a nine-day period, and similar multi-day appreciations and depreciations may be found for other foreign currencies. Similar events can be found within the 1250-day rolling sample used in our exercise for September 8, 1992, but the changes are all smaller: the largest 10-day moving average of percent log returns is 0.60 (January, 1988) and the smallest is -0.78 (October, 1987).

Figure 19 indicates quantiles of the predictive distributions using the same format as Figure 16, at intervals of three trading days beginning September 8, 1992. The average one-day percent log return of 1.74 lies between the 95th and 99th percentile of the predictive distribution for most of the period. Returns for periods of three or more holding days that end on or before September 21, however, all exceed the 99th percentile. This is especially evident in panel (a) for September 8, where 4- to 10-day returns are all above the 99.9th percentile.

The predictive distributions respond steadily to the events of the period. The interquartile range for the next day's percent log return increases from 0.835 on September 8, to 1.071 on September 11, to 1.119 on September 16, to 1.231 on September 21. The length of the centered 90% interval increases from 2.400 on September 8 to 3.695 on September 21. Thus there is at most a very modest increase in the tail thickness of the predictive density over the period. The medians of the predictive distributions are always very close to zero, but close inspection of Figure 19 reveals modest positive skewness in the predictive distributions that increases through the nine-day period.

4.3 Bond returns

On January 3, 2001, the Federal Open Market Committee convened an unscheduled meeting in which it initiated a long series of reductions in the Federal Funds rate following the collapse of the tech stock market bubble in 2000. Immediately preceding this meeting there

were expectations of an imminent change in the direction of monetary policy, but there was substantial uncertainty about the magnitude of the change. Changes in these expectations were reflected in movements in bond prices, with returns being greater for shorter-maturity issues.

Returns to one-year maturities were large and positive in the five trading days from Friday, December 29, 2000 through Friday, January 5, 2001, the average daily percent log return being 0.0272. Changes in one-year maturity bond prices were greater during this period than any other five-day period in the sample except for the period beginning with the re-opening of markets on Thursday, September 13, 2001. The closest precedents for changes of this magnitude in the 1250-day sample used in forming the predictive distributions are the five-day period beginning October 29, 1998 (-0.0138) and the five-day period beginning January 5, 1998 (0.0126).

Figure 20 indicates the quantiles of the predictive distributions using the same format as Figure 16, at intervals of three trading days beginning December 26, 2000. The only one-day return above the 99th percentile of the predictive distribution for any of these days (including the ones not shown in Figure 20) is that on January 2, shown in panel (b). The most extreme multiple-day returns, relative to their predictive distributions, are those for the period ending with January 4. This includes the 7-day return in panel (a) and the 4-day return panel (b). Together with the 5-day return for December 28 (not shown) the latter is the most extreme event, just above the 99th percentile.

Unlike predictive distributions for foreign exchange returns, bond return predictive distributions are centered about a positive return, and this is perhaps most clearly evident in Figure 20 by comparing the 5th and 95th percentiles of the distributions (the circles). There is little indication of skewness in the distributions. The predictive distributions respond steadily to the events of the period, with the interquartile range for the next day's percent lot return increasing from 0.0599 on December 28 to 0.0821 on January 10 (a factor of 1.37), after which it slowly decays in much the same fashion as did the dispersion for the S&P 500 returns portrayed in Figure 18. The centered 90% interval increases from 0.1808 on December 28 to 0.3161 on January 10 (a factor of 1.75) indicating an increase in tail thickness similar to that observed in Sections 4.1 and 4.2.

The response in 10-year maturity returns, Figure 21, was quite different. Percent log returns in the four-day period January 2 through January 5 were 1.637, -1.917, 0.813 and 0.773. These are all large in magnitude by historical standards, but not extreme, and there are several days in the 1250-day sample used in forming the predictive distributions with larger absolute returns. The January 5 return is the second largest in the sample and the January 6 return is the sixth smallest. Returns over the five-day period from December 29 through January 5 are unremarkable either historically or relative to this sample.

The predictive distributions for 10-year maturity returns displayed in Figure 21 are consistent with this history. The positive return on January 2 is above the 99th percentile (panel (b)), congruent with the history of returns. Returns over longer periods are nearly all within the centered 90% range of the predictive distribution, also consistent with history. These features are, again, consistent with the evidence that the model is well calibrated, discussed in Section 3.3.

5 Conclusions and future research

This study leads to a number of substantive conclusions about the modeling of asset returns with HMNM models, and some likely avenues for improving these models. It also illustrates some important implications of the BMCMC approach to inference for applied econometrics.

5.1 Substantive conclusions

Model comparison. For the four asset return series studied the HMNM model is strongly competitive with the most widely used tightly parameterized alternatives. These comparisons are based on the log predictive likelihoods of the models, $\sum_{t=T_1}^{T_2} \log p\left(y_t^o \mid \Phi_{t-1}\right)$, where the conditioning set Φ_{t-1} consists of the model specification and returns $y_{t-1250}^o, \ldots, y_{t-1}^o$ for rolling samples and the returns $y_1^o, \ldots, y_{t-1}^o$ for building samples. For any two models and samples with respective conditioning sets Φ_{t-1}^A and Φ_{t-1}^B,

$$\exp\left\{\sum_{t=T_1}^{T_2}\left[\log p\left(y_t^o \mid \Phi_{t-1}^{(A)}\right) - \log p\left(y_t^o \mid \Phi_{t-1}^{(B)}\right)\right] / \left(T_2 - T_1 + 1\right)\right\} - 1 \tag{36}$$

provides the geometric mean of the proportionate increase in the conditional probability density of the next day's asset return evaluated using model and sampling convention A, relative to the conditional probability density of the next day's asset return evaluated using model and sampling convention B. For example, the value 0.01 would indicate that typically model and sampling convention A produce a conditional probability density that is 1% higher, when evaluated at the realized return y_t^o, than does model and sampling convention B.

Take as the benchmark model B the normal GARCH(1,1) model with building samples, perhaps the single most widely used model of asset returns. Then this exercise produces the comparison of models exhibited in Table 14. In every case, t-GARCH models perform better than SV models, which in turn outperform GARCH and EGARCH models. The performance of the best HMNM model is tabulated in each case, but from Table 8 it is clear that in every case there is a fairly wide range of HMNM models with characteristics similar to the one selected for the table. For the S&P 500 and dollar - pound exchange returns the best HMNM model with building samples improves very substantially on t-GARCH, more than doubling the improvement of t-GARCH over GARCH in both cases. In the case of bond returns the HMNM model performance is similar to t-GARCH, being slightly better for one-year maturity bonds; for ten-year maturity bonds differences are within simulation approximation error. Results reported in Durham (2006, 2007) for S&P 500 returns indicate that were this comparison extended to include the most popular affine and SV models, their predictive likelihoods would also fall short of those of HMNM models.

Stability. Both absolutely and relative to its competitors, the HMNM model is stable despite the very different characteristics in different time period for the same return series documented in Tables 1 and 2. The strongest evidence for stability is in Table 14, which shows that the HMNM models with building samples have higher predictive likelihoods than those with rolling samples for all four return series. For a stable model this should always be the case, simply because larger samples have more information. For a model that is

misspecified, the pseudo-true parameter vector (Gourieroux et al. (1984) or Geweke (2005) Section 3.4) may vary from one time period to the next, leading to improved predictive performance with rolling samples. By contrast the record for the competing models studied is mixed, performance being better with building samples for some return series and with rolling samples for others. The prior and posterior predictive evidence, presented in Figure 3 and Figures 12 through 15 support this conclusion: the changes in characteristics of asset return series from one period to the next are consistent with the dynamics of the HMNM model. Moreover, even the unprecedented S&P 500 returns of October, 1987, appear not to affect the posterior distribution much as this period of great volatility leaves a 1250-day rolling sample. Against this evidence, predictive distributions using rolling samples appear to be better calibrated than those based on building samples. This is true for all four return series, and we do not understand the apparent discrepancy.

Practicality. The computational requirements of the HMNM model are greater than those of ARCH models with inference by maximum likelihood. They are similar to those of SV models and well within the reach of home desktop computers, to say nothing of the computing resources available to central banks and investment banks.

Distribution calibration. The HMNM model is quite flexible in its ability to describe faithfully the unconditional distributions of asset returns. This would be expected, given that the models used amount to a mixture of 16 normal distributions (when $m_1 = m_2 = 4$) or 25 normal distributions (when $m_1 = m_2 = 5$). This conclusion is supported by the evidence in the prior and posterior predictive distributions for skewness and kurtosis coefficients for both one- and ten-day returns. Conditional distributions over short horizons are also well calibrated. This conclusion is supported by the evidence from the out-of-sample predictive exercises reported in Section 3.3, which find no systematic departure from a uniform distribution on the unit interval for one-step-ahead predictive distributions evaluated at realized returns, and modest departures from this distribution for returns over horizons of several days.

Dynamic calibration. With respect to dynamics, the HMNM model is less well calibrated. The theoretical results in Section 2.2 document restrictions the model places on dynamics. Subsequent evidence from the return series suggests these limitations may, indeed, compromise the conditional predictive distributions that are central to the application of models of asset returns. The out-of-sample validation study in Section 3.3 found positive autocorrelation in the one-step-ahead predictive c.d.f.s evaluated at realized returns. Consistent, but much more suggestive, evidence emerged in the study of predictive distributions in Section 4, in particular the persistence over many weeks of extremely large tails in predictive distributions of the S&P 500 returns following the market crash of October, 1987 (Section 4.1), and the apparent failure to accommodate strong positive or negative returns persisting over several days in conjunction with a major currency realignment (Section 4.2).

Model shortcomings. The prior predictive analyses in Section 2.3 and posterior predictive analyses in Section 3.4 identified three likely specific deficiencies in model dynamics, All are important aspects of asset return dynamics, and in each case the problems arise mainly in the S&P 500 return series.

- The leverage effect, a negative sample correlation between one day's return y_t and the next day's absolute return $|y_{t+1}|$, is below -0.1 in three of the seven periods of S&P

500 returns studied. These sample correlations are roughly in the left 3% tail of the prior predictive distribution and well into the left 1% tail of the posterior predictive distributions.

- Long-memory models applied to absolute returns often suggested evidence of persistence incompatible with the geometric decay in the HMNM model documented in Section 2.2. For five-year periods this emerges as a difficulty for the HMNM model when the GPH estimate $\widehat{d}$, exceeds about 0.75, as it does for S&P 500 returns in the 1971-1975 period; Table 1 in Granger and Ding (1996) indicates that $\widehat{d}$ is substantially lower in earlier periods, back to 1928.

- The third deficiency is that while the HMNM model limits autocorrelation in returns, with the 99th percentile being about 0.20 for the first-order autocorrelation coefficient computed from five years of daily returns. Autocorrelation of this magnitude occurred in S&P 500 returns in the 1970s and presents even greater difficulties in the posterior predictive distribution than in the prior predictive.

Model improvement. These findings all suggest that further improvement in the HMNM model may lie in increasing the flexibility of the model's dynamics. Consistent with the hierarchical structure of the model and its interpretation as an artificial neural network with two hidden layers, an obvious generalization is to permit the latent states s_{t2} of the second layer to follow a first-order Markov model with transition probabilities specific to the value of the corresponding latent state s_{t1} of the first layer. This leads to a substantial but reasonable increase in the number of parameters in the model, for example from 90 to 170 for $m_1 = m_2 = 5$.

5.2 BMCMC applied econometrics

Modeling. Posterior simulation methods in general, and BMCMC in particular, have been critical in facilitating Bayesian analysis of models which previously were accessible only using non-Bayesian methods. While this has surely advanced the state of the art in Bayesian applied econometrics, we believe that a more significant consequence of these methods can be found in rethinking the entire econometric approach to a given applied problem bearing in mind their capabilities. This article provides one such example. The tractability of latent variable models and hierarchical parameter structures is well established in Bayesian statistics. This article demonstrates that this approach can be used in real time to provide a solution to an important practical problem that is superior to competing models and methods.

We believe that returns to further investments in this approach are likely to be high, both because the methods work and because they have been lightly explored relative to existing alternatives, such as the tightly parameterized models that are widespread in applied econometric studies of asset returns. A critical condition for successful outcomes is a good understanding of the capabilities of BMCMC methods and familiarity with the wide array of promising latent variable models, such as Markov mixtures and artificial neural networks.

Prior predictive analysis. This approach emphasizes serious thinking about the properties of models and prior distributions. It also provides an important and much underappreciated tool for this process, in the form of the prior predictive distribution. This distribution is useful for several reasons. First, it is easy to construct – typically an order of magnitude (or several) simpler than posterior simulators. Second, it can be used in conjunction with any approach to inference, including non-Bayesian extremum estimators, and it is typically an order of magnitude (or several) simpler than these procedures as well. Third, the prior predictive distribution enables the modeler to understand the properties of the model directly in terms of observables, which seems to us to be the only way a model is ever understood. Finally, the prior predictive distribution can be used to identify deficiencies in a model, in the form of observed data characteristics or observable stylized facts that turn out to be points or regions at which the prior predictive density is very small, or even zero.

The latter use of the prior predictive density can streamline research. It enables the applied econometrician to organize and construct a thorough specification analysis of a model before turning to inference, a reversal of the ordering that is more convenient, if indeed not required, using other inference methods. The hard work of formal inference begins only when the prior predictive analysis indicates it is sufficiently promising to be worth the effort. It should be self-evident that prior predictive distributions are not substitutes for posterior distributions, as comparisons of the posterior and prior predictive distributions in this study make clear. The HMNM model based on prior rather than posterior distributions would fare poorly in the competition whose outcome is displayed in Table 14.

Model validation. Without knowing whether the finding will generalize, we find it noteworthy that the prior and posterior predictive distributions provided similar information about model limitations in this study. One of us (Geweke (2007b)) has pointed out elsewhere that the posterior predictive distribution does not have a true Bayesian interpretation – in short, because it conditions on that which it is "predicting" – whereas the prior predictive distribution is not subject to this problem. Geweke (2007b) conjectured that deficiencies revealed post-inference, as in the posterior predictive distribution, would also emerge in the prior predictive distribution, and demonstrated this feature in an illustrative simple example. That this conjecture is confirmed in this study suggests that the Bayesian approach to model evaluation in Geweke (2007b) based on prior predictive distributions may, in fact, be useful in competitive research models.

Model comparison. The predictive likelihood is a compelling measure of model performance. We have given it a fully Bayesian expression here, for example in (36). But we have used it to compare models for which only maximum likelihood estimates are available as was the case for all of the ARCH models in this study. The important criterion is which method of inference will be used in forming predictive distributions – if it is maximum likelihood, then this is the appropriate method of comparison, and if it is Bayesian then the full posterior predictive distribution should be used, as was the case for the SV models in this study. Systematic reporting of predictive likelihoods for benchmark asset return series and time intervals would do much to advance communication among applied econometricians and the state of the art generally.

Acknowledgements

We acknowledge useful comments on this and earlier versions of the paper, from seminar participants in the JAE Conference on Changing Structures in International and Financial Markets, NBER-NSF Time Series workshop, the Research Triangle Econometrics workshop, Reserve Bank of Australia, Queensland University of Technology, University of New South Wales, University of British Columbia, University of Chicago, University of Iowa, University of Maryland, University of Pennsylvania, University of Technology - Sydney, and University of Victoria. Without imputing responsibility or endorsement we acknowledge in particular productive discussions with T. Bollerslev, F.X. Diebold, G. Durham, R.F. Engle, L. Hansen, and G. Tauchen. The first author acknowledges partial financial support from the U.S. National Science Foundation through grants SBR-9819444 and SBR-0720547.

References

Bank for International Settlements 2005. Zero-coupon yield curves: technical documentation. BIS Papers n. 25, October 2005.

Bartlett JS. 1957. A comment on D.V. Lindley's statistical paradox. *Biometrika* **44**: 533-534.

Berkowitz J. 2001. The accuracy of density forecasts in risk management. *Journal of Business and Economic Statistics* **19**: 465-474.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307-327.

Bollerslev T. 1996. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* **73**: 151-184.

Box, GEP. 1980. Sampling and Bayes' inference in scientific modeling (with discussion and rejoinder). *Journal of the Royal Statistical Society Series* A **143**: 383-430.

Chib, S. 1996. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* **75**: 79-97.

Diebold FX, Inoue A. 2001. Long memory and regime switching. *Journal of Econometrics* **105**: 131-159.

Ding Z, Granger CWJ, Engle RF. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1**: 83-106.

Durham GB. 2006. Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models. *Journal of Econometrics* **133**: 273-305.

Durham GB. 2007. Mixture models with application to S&P 500 index returns. *Journal of Financial Economics* (forthcoming).

Engle, RF. 1983. Estimates of the variance of United States inflation based upon the ARCH model. *Journal of Money, Credit and Banking* **15**: 286-301.

Eraker B, Johannes M, Polson NG. 2003. The impact of jumps in volatility and returns. *Journal of Finance* **58**: 1269-1300.

Gallant, AR, Tauchen T. 1989. Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica* **57**: 1091-1120.

Geweke J. 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica **57**: 1317-1340.

Geweke J. 2004. Getting it right: Joint distribution tests of posterior simulators. *Journal of the American Statistical Association* **99**: 799-804.

Geweke J. 2005. *Contemporary Bayesian Econometrics and Statistics.* John Wiley: Hoboken, NJ.

Geweke J. 2007a. Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics and Data Analysis* **51**: 3529-3550.

Geweke J. 2007b. Bayesian model comparison and validation. *American Economic Review* (forthcoming).

Geweke J, Porter-Hudak S. 1984. The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**: 221-238.

Geweke J, Whiteman C. 2006. Bayesian forecasting. Granger CWJ, Timmerman A (eds.) *Handbook of Economic Forecasting.* Elsevier: Amsterdam, The Netherlands.

Gourieroux C, Monfort A, Trognon A. 1984. Pseudo maximum liklelihod methods: Theory. *Econometrica* **52**: 681-800.

Granger CWJ, Ding Z. 1996. Varieties of long memory models. Journal of Econometrics **73**: 61-77.

Harvey CR, Siddique A. 1999. Autoregressive conditional skewness. *The Journal of Financial and Quantitative Analysis* **34**: 465-487.

Jacquier E, Poslon NG, Rossi PE. 1994. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* **12**: 371-389.

Kuan CM, White H. 1992. Artificial neural networks: An econometric perspective. *Econometric Reviews* **13**: 1-19.

Lancaster T. 2004. *An Introduction to Modern Bayesian Econometrics.* Blackwell: Malden, MA.

Lindgren, G. 1978. Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics* **5**: 81-91.

Lo AW. 1988. Maximum likelihood estimation of generalized Ito processes with discretely sampled data. *Econometric Theory* **4**: 231-247.

McLachlan D, Peel G. 2000. *Finite Mixture Models.* John Wiley: Hoboken, NJ.

Nelson DB. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**: 347-370.

Nelson CR, Siegel AF. 1987. Parsimonious modeling of yield curves. *Journal of Business* **60**: 473-89.

Rao CR. 1965, *Linear Statistical Inference and Its Applications.* Wiley: New York, NY.

Rosenblatt J. 1952. Remarks on a multivariate transformation. *Annals of Mathematical Statistics* **23**: 470-472.

Rydén T, Teräsvirta T, Åsbrink S. 1998. Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics* **13**: 217-244.

Svensson LEO. 1994. Estimating and interpreting forward interest rates: Sweden 1992-4. *NBER Working Paper Series No. 4871.*

Taylor, S. 1986. *Modelling Financial Time Series.* John Wiley: New York, NY.

Tyssedal, JS, Tjøstheim, D. 1988. An autoregressive model with suddenly changing parameters and an application to stock market prices. *Applied Statistics* **37**: 353-369.

Table 1: Functions of the observed return series

	Return moments				Return dynamics			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Skew y^1	Kurt y^1	Skew y^{10}	Kurt y^{10}	$\widehat{\rho}_1\ y^1$	$\widehat{\rho}_1\ y^{10}$	$\widehat{\rho}_1\ y^5$	$\sum_{i=1}^{200} \lvert\widehat{\rho}_i\rvert\ y^1$
S&P 500 returns								
71-75	0.244	1.739	-0.053	0.974	0.266	0.068	0.191	5.573
76-80	-0.002	1.289	-0.397	0.730	0.143	-0.049	-0.347	4.363
81-85	0.403	1.747	0.602	1.950	0.086	-0.064	-0.222	4.800
86-90	-4.859	84.693	-3.028	21.600	0.036	-0.034	-0.203	4.438
91-95	0.048	2.697	0.529	2.496	0.017	-0.033	-0.064	4.834
96-00	-0.340	3.547	-0.283	0.589	0.000	-0.137	-0.286	4.695
01-05	0.166	2.388	-0.488	3.165	-0.034	0.091	-0.035	4.839
Dollar-pound returns								
72-76	1.042	10.064	0.671	2.096	0.053	0.164	0.057	4.126
77-81	0.523	6.279	0.364	0.724	0.031	0.016	0.120	4.498
82-86	-0.490	3.403	-1.053	3.026	0.074	-0.082	0.246	4.933
87-91	0.246	1.776	0.326	-0.007	0.092	0.128	0.117	4.522
92-96	0.504	3.090	2.165	10.505	0.043	0.137	-0.001	5.104
One-year bonds								
90-94	-0.054	10.142	-0.154	0.156	0.096	0.222	0.158	5.443
95-99	0.490	6.091	-0.022	0.907	0.091	0.112	0.137	4.443
00-04	0.663	8.526	0.665	3.335	0.052	0.002	0.237	4.994
Ten-year bonds								
90-94	0.052	0.080	0.170	4.455	0.052	0.080	0.170	4.455
95-99	0.080	0.010	0.214	4.075	0.080	0.010	0.214	4.075
00-04	0.036	-0.058	-0.251	4.844	0.036	-0.058	-0.251	4.844

Table 2: Functions of the observed return series

	Absolute return short-run dynamics				Absolute return long-run dynamics $\lvert y^1\rvert$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\widehat{\rho}_1$ $\lvert y^1\rvert$	$\widehat{\rho}_9$ $\lvert y^1\rvert$	$\frac{var\left(y^{10}\right)}{var(y^1)}$	Leverage	GPH $\widehat{d}$	$\sum_{i=1}^{200}\lvert\widehat{\rho}_i\rvert$	$\frac{\sum_{i=1}^{100}\lvert\widehat{\rho}_i\rvert}{\sum_{i=1}^{200}\lvert\widehat{\rho}_i\rvert}$	$\frac{\sum_{i=1}^{200}\widehat{\rho}_i}{\sum_{i=1}^{200}\lvert\widehat{\rho}_i\rvert}$
S&P 500 returns								
71-75	0.215	0.222	13.296	-0.037	0.838	24.950	0.596	1.000
76-80	0.095	0.123	12.444	-0.011	0.670	7.697	0.628	0.884
81-85	0.025	0.071	11.435	0.015	0.417	7.339	0.699	0.821
86-90	0.242	0.141	8.911	-0.137	0.331	6.924	0.770	0.769
91-95	0.031	0.057	8.812	-0.053	0.542	6.846	0.614	0.893
96-00	0.111	0.081	7.640	-0.130	0.452	7.092	0.639	0.869
01-05	0.184	0.203	8.821	-0.114	0.475	23.185	0.655	1.000
Dollar-pound returns								
72-76	0.313	0.091	12.389	0.107	0.353	7.188	0.668	0.636
77-81	0.236	0.181	10.514	0.065	0.534	16.676	0.765	0.990
82-86	0.121	0.094	11.698	0.007	0.611	11.325	0.719	0.938
87-91	0.100	0.075	11.169	0.072	0.244	4.949	0.592	0.011
92-96	0.219	0.184	11.826	0.080	0.400	18.586	0.637	1.000
One-year bonds								
90-94	0.109	0.009	13.785	-0.095	0.188	4.290	0.533	0.309
95-99	0.096	0.038	10.536	0.054	0.295	5.305	0.623	0.417
00-04	0.109	0.151	10.031	0.013	0.203	8.778	0.638	0.926
Ten-year bonds								
90-94	0.018	0.044	8.660	-0.026	0.445	5.198	0.571	0.243
95-99	0.013	0.026	9.540	-0.046	0.317	4.430	0.509	0.238
00-04	0.017	0.103	10.033	-0.056	0.320	5.492	0.600	0.505

Table 3: Values of prior distribution hyperparameters

Model parameters	Corresponding prior hyperparameters
β	$\underline{\beta} = 0,\ \underline{H}_\beta = 100$
$\mathbf{P}$	$\underline{p}_0 = 10\,,\ \underline{p}_1 = 0.1$
$\mathbf{R}$	$\underline{r} = 1$
h	$\underline{s}^2 = 0.001,\ \underline{\nu} = 2$
h_i	$\underline{\nu}_1 = 0.5$
h_{ij}	$\underline{\nu}_2 = 0.5$
$\widetilde{\boldsymbol{\phi}}$	$\underline{h}_\phi = 1$
$\widetilde{\boldsymbol{\psi}}_j$	$\underline{h}_\psi = 1$

Table 4: Prior predictive c.d.f. at observed values, HMNM model, m1=m2=4

	Return moments				Return dynamics			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Skew y^1	Kurt y^1	Skew y^{10}	Kurt y^{10}	$\widehat{\rho}_1$ y^1	$\widehat{\rho}_1$ y^{10}	$\widehat{\rho}_1$ y^{50}	$\sum_{i=1}^{200} \lvert\widehat{\rho}_i\rvert$ y
S&P 500 returns								
71-75	0.640	0.067	0.451	0.271	**0.995**	0.751	0.866	0.937
76-80	0.494	0.046	0.254	0.232	0.979	0.348	0.066	0.596
81-85	0.702	0.067	0.795	0.384	0.949	0.295	0.178	0.833
86-90	0.071	0.777	0.072	0.807	0.822	0.396	0.202	0.651
91-95	0.529	0.110	0.778	0.427	0.695	0.401	0.433	0.843
96-00	0.320	0.148	0.299	0.204	0.500	0.136	0.108	0.798
01-05	0.600	0.097	0.228	0.473	0.194	0.803	0.508	0.845
Dollar-pound returns								
72-76	0.806	0.351	0.809	0.397	0.908	0.725	0.926	0.967
77-81	0.728	0.252	0.729	0.231	0.844	0.871	0.688	0.922
82-86	0.276	0.142	0.153	0.464	0.688	0.733	0.880	0.587
87-91	0.641	0.068	0.711	0.055	0.645	0.688	0.818	0.934
92-96	0.724	0.127	0.911	0.689	0.825	0.875	0.892	0.945
One-year bonds								
90-94	0.460	0.353	0.377	0.094	0.959	0.948	0.830	0.928
95-99	0.722	0.245	0.475	0.261	0.954	0.844	0.806	0.654
00-04	0.756	0.314	0.807	0.483	0.887	0.534	0.902	0.877
Ten-year bonds								
90-94	0.344	0.114	0.322	0.057	0.888	0.778	0.843	0.664
95-99	0.308	0.083	0.353	0.012	0.943	0.566	0.886	0.364
00-04	0.290	0.036	0.175	0.254	0.821	0.316	0.143	0.846

Table 5: Prior predictive c.d.f. at observed values, HMNM model, m1=m2=4

	Absolute return short-run dynamics				Absolute return long-run dynamics $\|y^1\|$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\widehat{\rho}_1\ \|y^1\|$	$\widehat{\rho}_9\ \|y^1\|$	$\frac{var(y^{10})}{var(y^1)}$	Leverage	GPH $\widehat{d}$	$\sum_{i=1}^{200}\|\widehat{\rho}_i\|$	$\frac{\sum_{i=1}^{100}\|\widehat{\rho}_i\|}{\sum_{i=1}^{200}\|\widehat{\rho}_i\|}$	$\frac{\sum_{i=1}^{200}\widehat{\rho}_i}{\sum_{i=1}^{200}\|\widehat{\rho}_i\|}$
S&P 500 returns								
71-75	0.821	0.910	0.954	0.169	0.987	0.975	0.592	0.048*
76-80	0.635	0.789	0.929	0.355	0.951	0.791	0.662	0.902
81-85	0.398	0.678	0.853	0.668	0.781	0.777	0.787	0.881
86-90	0.851	0.820	0.233	0.023	0.704	0.758	0.878	0.863
91-95	0.427	0.634	0.214	0.107	0.884	0.754	0.632	0.905
96-00	0.667	0.702	0.080	0.025	0.816	0.766	0.684	0.896
01-05	0.785	0.894	0.215	0.032	0.833	0.969	0.714	0.048*
Dollar-pound returns								
72-76	0.908	0.725	0.926	0.967	0.727	0.771	0.736	0.817
77-81	0.844	0.871	0.688	0.922	0.879	0.936	0.872	0.957
82-86	0.688	0.733	0.880	0.587	0.922	0.878	0.814	0.924
87-91	0.645	0.688	0.818	0.934	0.612	0.607	0.582	0.360
92-96	0.825	0.875	0.892	0.945	0.769	0.948	0.681	0.048*
One-year bonds								
90-94	0.663	0.363	0.963	0.042	0.539	0.340	0.362	0.645
95-99	0.636	0.552	0.694	0.895	0.668	0.653	0.650	0.707
00-04	0.661	0.835	0.554	0.652	0.557	0.826	0.683	0.920
Ten-year bonds								
90-94	0.360	0.582	0.187	0.232	0.809	0.642	0.527	0.603
95-99	0.326	0.485	0.399	0.129	0.693	0.426	0.211	0.599
00-04	0.348	0.750	0.554	0.099	0.695	0.669	0.600	0.757

*Observed value is at the mass point 1.0, with indicated probability.

Table 6: Total execution time per iteration and allocation by parameter group

$T =$	1250	1250	5000	5000
$m_1 = m_2 =$	4	6	4	6
Total time (secs.)	3.05×10^{-3}	6.68×10^{-3}	12.0×10^{-3}	26.6×10^{-3}
$\mathbf{s}^1$	60.5%	53.6%	65.1%	60.5%
$\mathbf{s}^2$	7.6%	4.0%	7.8%	3.8%
$h, \mathbf{h}, \mathbf{H}$	0.9%	0.9%	0.8%	0.4%
$\mathbf{P}$	7.7%	8.1%	4.4%	3.8%
$\mathbf{R}$	1.9%	1.1%	0.9%	0.6%
β, $\boldsymbol{\phi}$, $\boldsymbol{\psi}$	21.3%	32.4%	21.1%	31.0%

Table 7: Distribution of RNE over 7,323 posterior distributions

	Quantile		
Model	0.10	0.50	0.90
$m_1 = m_2 = 2$	.172	.450	.859
$m_1 = m_2 = 3$	.118	.328	.632
$m_1 = m_2 = 4$	.126	.314	.606
$m_1 = m_2 = 5$	.133	.333	.626

Table 8: Log predictive likelihoods

	S&P 500		Dollar-pound		One-year bonds		Ten-year bonds	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Rolling	Building	Rolling	Building	Rolling	Building	Rolling	Building
Gaussian i.i.d.	-10570.8	-10477.0	-5370.4	-5570.8	5799.7	5868.8	-2891.2	-2872.8
GARCH	-9574.4	-9523.0	-5097.7	-5133.8	6032.4	6054.5	-2789.2	-2780.2
EGARCH	-9549.4	-9476.9	-5046.1	-5061.9	6025.1	6056.1	-2801.0	-2796.6
t-GARCH	-9317.5	-9327.3	-4576.7	-4675.3	6314.3	6319.0	-2721.3	-2716.8
Stochastic vol.	-9462.2	-9382.8	-4650.3	-4965.9	6256.3	6283.4	-2749.3	-2770.9
$HMNM$, serial correlation								
$m_1 = m_2 = 3$	-9335.5	-9185.9	-4379.4	-4246.5	6315.8	6305.2	-2734.3	-2721.9
$m_1 = m_2 = 4$	-9323.7	-9113.1	-4330.0	-4256.5	6322.2	6317.6	-2733.5	-2722.8
$m_1 = m_2 = 5$	-9332.8	-9094.3	-4298.9	-4043.5	6323.9	6320.4	-2736.2	-2725.0
$m_1 = m_2 = 6$	-9346.4	-9123.6	-4323.4	-4043.2	6322.3	6320.6	-2738.9	-2725.6
$HMNM$, no serial correlation								
$m_1 = m_2 = 3$	-9335.4	-9295.3	-4361.4	-4288.5	6318.0	6319.0	-2731.8	-2722.2
$m_1 = m_2 = 4$	-9327.4	-9229.8	-4287.7	-4164.6	6319.4	6324.9	-2732.4	-2722.8
$m_1 = m_2 = 5$	-9334.4		-4273.8	-4059.3	6319.5	6325.2	-2733.5	-2722.5
$m_1 = m_2 = 6$	-9352.2		-4282.0	-4027.2	6317.9	6325.5	-2739.0	–2725.6
MNM, serial correlation								
$m_1 = m_2 = 3$	-9398.5	-9482.9	-4692.4	-4793.7	6249.0	6250.5	-2752.5	-2743.2
$m_1 = m_2 = 4$	-9372.6		-4614.3	-4696.8	6284.6	6290.2	-2748.9	-2732.0
$m_1 = m_2 = 5$	-9369.5		-4517.4	-4419.8	6291.0	6301.2	-2747.6	-2725.7
$m_1 = m_2 = 6$	-9372.8		-4533.3	-4425.6	6298.8	6306.5	-2745.6	-2727.3
MNM, no serial correlation								
$m_1 = m_2 = 3$	-9398.7		-4697.9	-4788.8	6254.5	6270.3	-2752.1	-2746.6
$m_1 = m_2 = 4$	-9371.1		-4611.0	-4695.9	6295.3	6311.6	-2749.8	-2728.4
$m_1 = m_2 = 5$	-9365.3		-4518.6	-4416.6	6298.2	6317.5	-2747.8	-2726.9
$m_1 = m_2 = 6$	-9365.0		-4513.2		6301.3	6318.0	-2748.3	-2725.1

Table 9: Some further comparisons with stochastic volatility models

Model	Sample	6/25/80-9/20/02	1/2/80-12/13/99
Durham SV mixture	All	-7258.07	-6269.21
HMNM, $m_1 = m_2 = 4$	Rolling	-7335.66	-6232.90
HMNM, $m_1 = m_2 = 4$	Building	-7188.41	-6195.24
HMNM, $m_1 = m_2 = 5$	Rolling	-7342.17	-6330.71
HMNM, $m_1 = m_2 = 5$	Building	-7150.75	-6157.05

Table 10: Out of sample predictive calibration

Series	S&P 500				Dollar - Pound			
Model	HMNM ($m_1 = m_2 = 4$)				HMNM* ($m_1 = m_2 = 5$)			
Samples	Rolling		Building		Rolling		Building	
Returns	1-day	10-day	1-day	10-day	1-day	10-day	1-day	10-day
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean	-0.0030	-0.1327	0.0126	0.0337	-0.0039	0.0028	-0.0121	-0.0231
Variance	0.9775	0.9650	1.0125	0.9754	1.0092	1.0864*	1.0667*	1.1919*
Skewness	-0.0471	0.2159*	0.1620	0.2308*	-0.0478	-0.2321*	-0.2026	-0.2480*
Ex. Kurtosis	-0.0752	-0.1631*	-0.1364*	-0.2262*	0.0254	0.0868	0.0216	0.1154*
$\widehat{\rho}_1$	0.0642*		0.0523*		0.0630*		0.0581*	
Decile 1	0.0994	0.0963	0.0987	0.0941	0.1006	0.1142	0.1102	0.1300
Decile 2	0.0995	0.0993	0.0918	0.0928	0.1030	0.0973	0.1057	0.1026
Decile 3	0.0987	0.0949	0.0903	0.0871	0.1044	0.0933	0.1068	0.0977
Decile 4	0.0978	0.0976	0.1051	0.0915	0.1006	0.0990	0.0984	0.0986
Decile 5	0.1017	0.1002	0.0974	0.0972	0.0930	0.1062	0.0836	0.0957
Decile 6	0.0998	0.1062	0.1040	0.1025	0.0924	0.0977	0.0837	0.0870
Decile 7	0.1036	0.1102	0.1099	0.1174	0.1031	0.0895	0.1028	0.0785
Decile 8	0.1006	0.1120	0.1025	0.1147	0.1017	0.0901	0.0997	0.0883
Decile 9	0.1017	0.0990	0.0976	0.1125	0.1015	0.0950	0.1053	0.0955
Decile 10	0.0971	0.0844	0.1027	0.0901	0.0997	0.1176	0.1039	0.1262
Decile range	0.0065	0.0276	0.0196	0.0303	0.0114	0.0275	0.0266	0.0515

Table 11: Out of sample predictive calibration

Series	One-year bonds				Ten-year bonds			
Model	HMNM*($m_1 = m_2 = 5$)				HMNM($m_1 = m_2 = 3$)			
Samples	Rolling		Building		Rolling		Building	
Returns	1-day	10-day	1-day	10-day	1-day	10-day	1-day	10-day
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean	-0.0177	-0.0458	-0.0205	-0.0496	-0.0026	-0.0181	-0.0052	-0.0308
Variance	-.9551	0.8581*	0.9554	0.7721*	0.9560	0.9139*	-.9716	0.9199
Skewness	0.0218	0.0351	-0.0945	0.3657*	0.0087	-0.1294	-0.1850	-0.1831
Ex. Kurtosis	0.0159	0.0270	0.0245	0.1017	-0.0721	-0.2612*	-0.0804	-0.2471*
$\widehat{\rho}_1$	0.0395		0.0036		0.0498		0.0494	
Decile 1	0.0904	0.0990	0.0956	0.0813	0.0958	0.1015	0.0987	0.1069
Decile 2	0.1095	0.1049	0.1075	0.1101	0.0958	0.1027	0.1035	0.0987
Decile 3	0.1081	0.0964	0.1067	0.0973	0.1067	0.0907	0.1032	0.0930
Decile 4	0.1024	0.0956	0.1089	0.1067	0.0970	0.0870	0.0956	0.0890
Decile 5	0.1004	0.1141	0.0987	0.1186	0.1015	0.0970	0.0976	0.0953
Decile 6	0.1001	0.1081	0.0953	0.1183	0.1055	0.1104	0.0973	0.1072
Decile 7	0.0990	0.1069	0.0936	0.1143	0.0927	0.1106	0.0933	0.1155
Decile 8	0.1018	0.1163	0.1013	0.1129	0.1112	0.1041	0.1061	0.1024
Decile 9	0.0950	0.0913	0.0961	0.0850	0.0998	0.1254	0.1069	0.1172
Decile 10	0.0933	0.0674	0.0964	0.0555	0.0939	0.0705	0.0978	0.0748
Decile range	0.0191	0.0467	0.0153	0.0631	0.0173	0.0549	0.0136	0.0424

Table 12: Posteror predictive c.d.f. at observed vlaues

	Return moments				Return dynamics			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Skew y^1	Kurt y^1	Skew y^{10}	Kurt y^{10}	$\widehat{\rho}_1$ y^1	$\widehat{\rho}_1$ y^{10}	$\widehat{\rho}_1$ y^{50}	$\sum_{i=1}^{200} \lvert\widehat{\rho}_i\rvert$ y
S&P 500 returns (HMNM $(m_1 = m_2 = 4)$*)*								
71-75	0.713	0.211	0.473	0.331	**1.000**	0.747	0.855	0.974
76-80	0.467	0.122	0.219	0.252	**0.999**	0.344	0.057	0.300
81-85	0.799	0.212	0.882	0.555	0.985	0.291	0.174	0.711
86-90	0.025	0.947	0.015	0.972	0.834	0.394	0.199	0.711
91-95	0.531	0.382	0.863	0.645	0.676	0.402	0.439	0.735
96-00	0.199	0.484	0.279	0.204	0.496	0.109	0.103	0.625
01-05	0.655	0.332	0.180	0.718	0.177	0.808	0.497	0.737
*Dollar-pound returns (HMNM** $(m_1 = m_2 = 5)$*)*								
72-76	0.881	0.703	0.864	0.623	0.896	0.929	0.692	0.216
77-81	0.827	0.657	0.776	0.301	0.786	0.579	0.798	0.595
82-86	0.152	0.534	0.091	0.701	0.944	0.240	0.929	0.882
87-91	0.696	0.245	0.757	0.032	0.965	0.882	0.795	0.618
92-96	0.823	0.499	0.937	0.824	0.856	0.895	0.581	0.929
One-year bonds (HMNM $(m_1 = m_2 = 5)$*)*								
90-94	0.385	0.796	0.236	0.121	0.943	0.911	0.682	0.981
95-99	0.789	0.569	0.360	0.519	0.929	0.643	0.648	0.416
00-04	0.850	0.733	0.903	0.925	0.697	0.263	0.796	0.890
Ten-year bonds (HMNM $(m_1 = m_2 = 5)$*)*								
90-94	0.350	0.726	0.268	0.131	0.948	0.790	0.832	0.579
95-99	0.196	0.526	0.330	0.017	0.989	0.537	0.879	0.146
00-04	0.143	0.090	0.034	0.713	0.866	0.268	0.132	0.910

Table 13: Posterior predictive c.d.f. at observed values

	Absolute return long-run dynamics $\|y^1\|$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\widehat{\rho}_1\ \|y^1\|$	$\widehat{\rho}_9\ \|y^1\|$	$\frac{var(y^{10})}{var(y^1)}$	Leverage	GPH $\widehat{d}$	$\sum_{i=1}^{200}\|\widehat{\rho}_i\|$	$\frac{\sum_{i=1}^{100}\|\widehat{\rho}_i\|}{\sum_{i=1}^{200}\|\widehat{\rho}_i\|}$	$\frac{\sum_{i=1}^{200}\widehat{\rho}_i}{\sum_{i=1}^{200}\|\widehat{\rho}_i\|}$
S&P 500 returns (HMNM $(m_1 = m_2 = 4)$*)*								
71-75	0.886	0.955	0.987	0.181	**0.995**	0.990	0.177	0.048*
76-80	0.307	0.589	0.961	0.404	0.923	0.401	0.280	0.642
81-85	0.047	0.268	0.873	0.676	0.389	0.360	0.581	0.559
86-90	0.942	0.694	0.201	**0.002**	0.202	0.314	0.849	0.499
91-95	0.060	0.191	0.177	0.098	0.710	0.305	0.232	0.654
96-00	0.387	0.323	0.022	**0.002**	0.478	0.322	0.328	0.620
01-05	0.779	0.919	0.179	**0.005**	0.539	0.983	0.396	0.048*
*Dollar-pound returns (HMNM** $(m_1 = m_2 = 5)$*)*								
72-76	0.944	0.546	0.939	0.967	0.483	0.639	0.613	0.732
77-81	0.879	0.906	0.698	0.920	0.828	0.977	0.882	0.980
82-86	0.517	0.564	0.896	0.574	0.920	0.896	0.766	0.943
87-91	0.405	0.446	0.829	0.933	0.263	0.218	0.321	0.096
92-96	0.853	0.910	0.906	0.945	0.584	0.983	0.506	0.257*
One-year bonds (HMNM $(m_1 = m_2 = 5)$*)*								
90-94	0.815	0.166	0.895	**0.006**	0.239	0.106	0.202	0.402
95-99	0.728	0.400	0.230	0.930	0.459	0.492	0.684	0.508
00-04	0.810	0.985	0.133	0.643	0.263	0.914	0.745	0.929
Ten-year bonds (HMNM $(m_1 = m_2 = 5)$*)*								
90-94	0.232	0.555	0.104	0.225	0.847	0.566	0.516	0.409
95-99	0.190	0.366	0.341	0.086	0.568	0.207	0.122	0.404
00-04	0.216	0.919	0.525	0.050	0.577	0.648	0.663	0.666

*Observed value is at the mass point 1.0, with indicated probability.

Table 14: Geometric mean of predictive likelihoods relative to GARCH with building samples

		GARCH	EGARCH	t-GARCH	SV	HMNM (best)
S&P500 returns	Rolling	-0.70%	-0.36%	2.85%	0.83%	2.76%
	Building	0.00%	0.63%	2.71%	1.93%	6.03%
Dollar-pound returns	Rolling	0.66%	1.60%	11.06%	9.16%	11.68%
	Building	0.00%	1.31%	8.67%	3.09%	22.21%
One-year bonds	Rolling	-0.63%	-0.83%	7.67%	5.91%	7.96%
	Building	0.00%	0.05%	7.81%	6.72%	8.01%
Ten-year bonds	Rolling	-0.25%	-0.59%	1.69%	0.88%	1.39%
	Building	0.00%	-0.47%	1.81%	0.26%	1.67%

Figure 1: Prior predictive distribution of asset return characteristics, HMNM model, $m_1 = m_2 = 4$, serial correlation permitted.

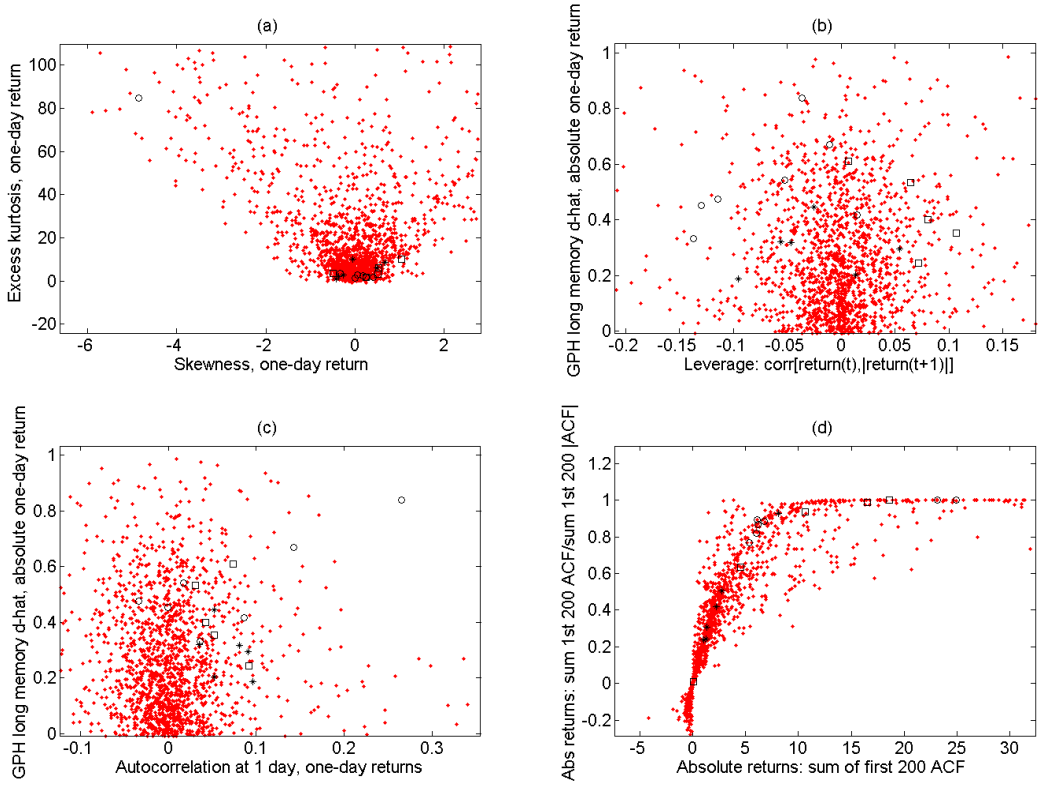

Figure 2: Prior predictive distribution of asset return characteristics, HMNM model, $m_1 = m_2 = 4$, serial correlation permitted.

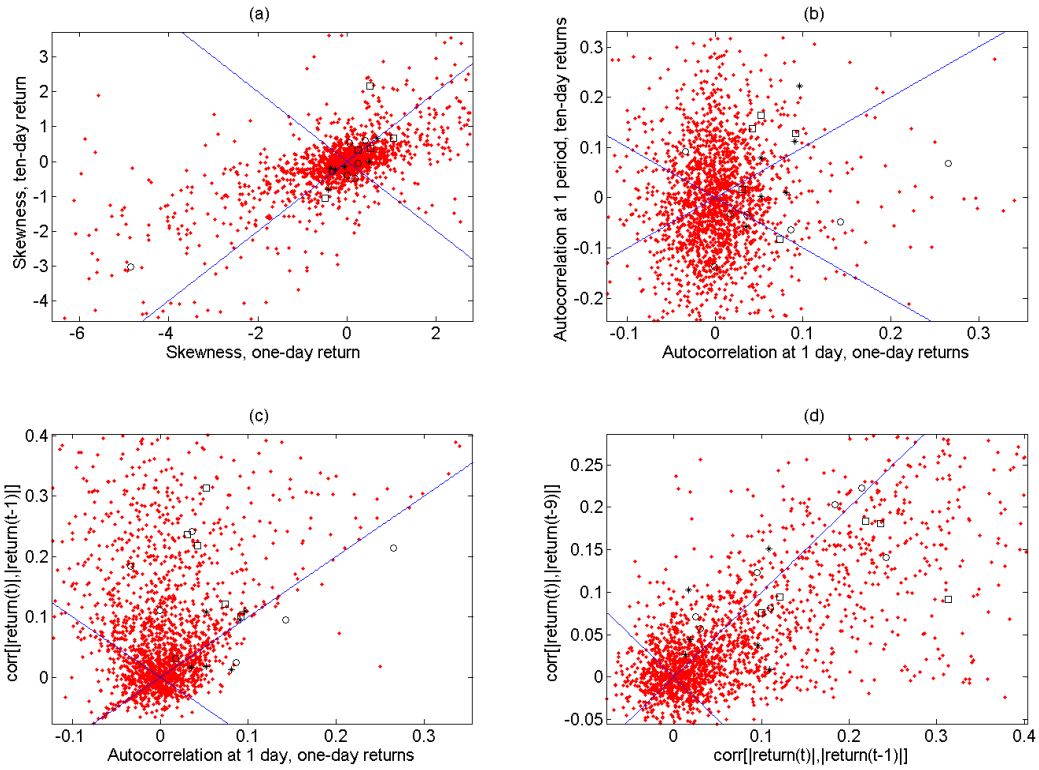

Figure 3: Prior predictive distribution of asset return characteristics in two successive five-year time periods, HMNM model, $m_1 = m_2 = 4$, serial correlation permitted.

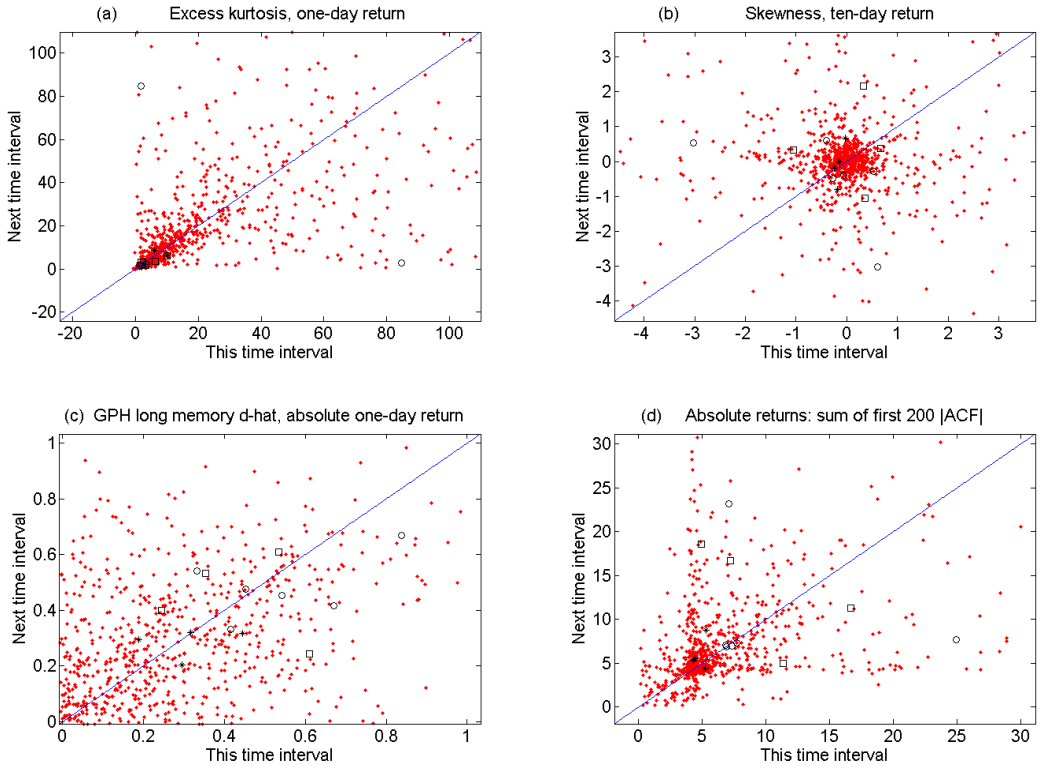

Figure 4: Full-sample posterior predictive distributions of skewness (horizontal axis) and excess kurtosis (vertical axis) in a five-year sample

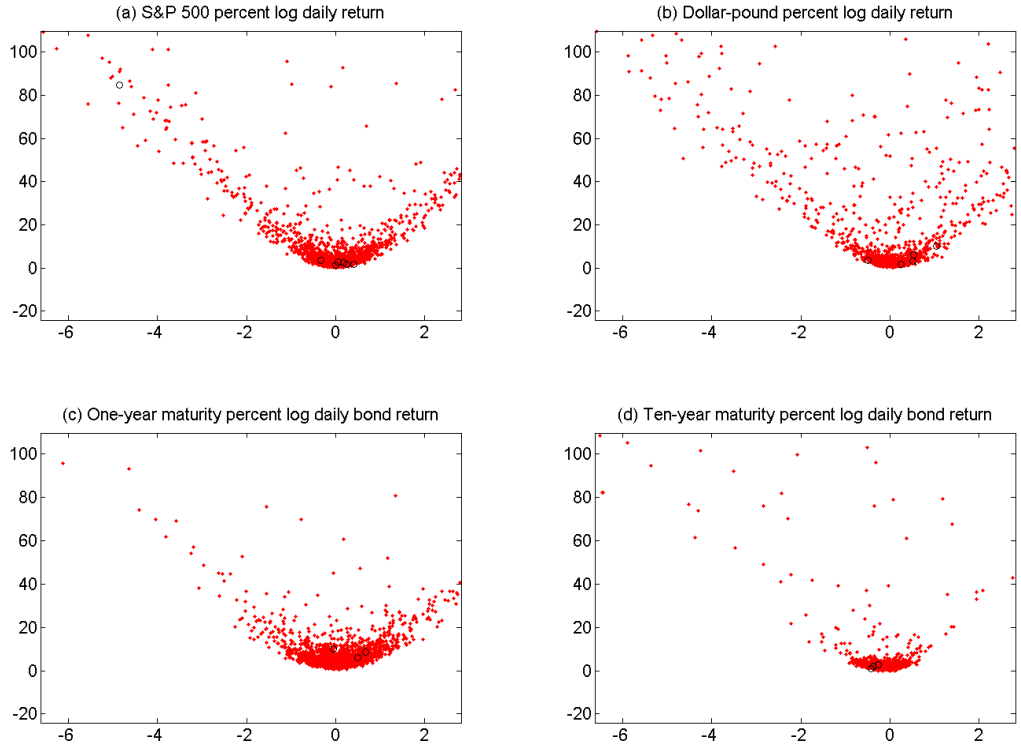

Figure 5: Full-sample posterior predictive distributions of levereage (horizontal axis) and one-day absolute return long memory $\widehat{d}$ (vertical axis) in a five-year sample

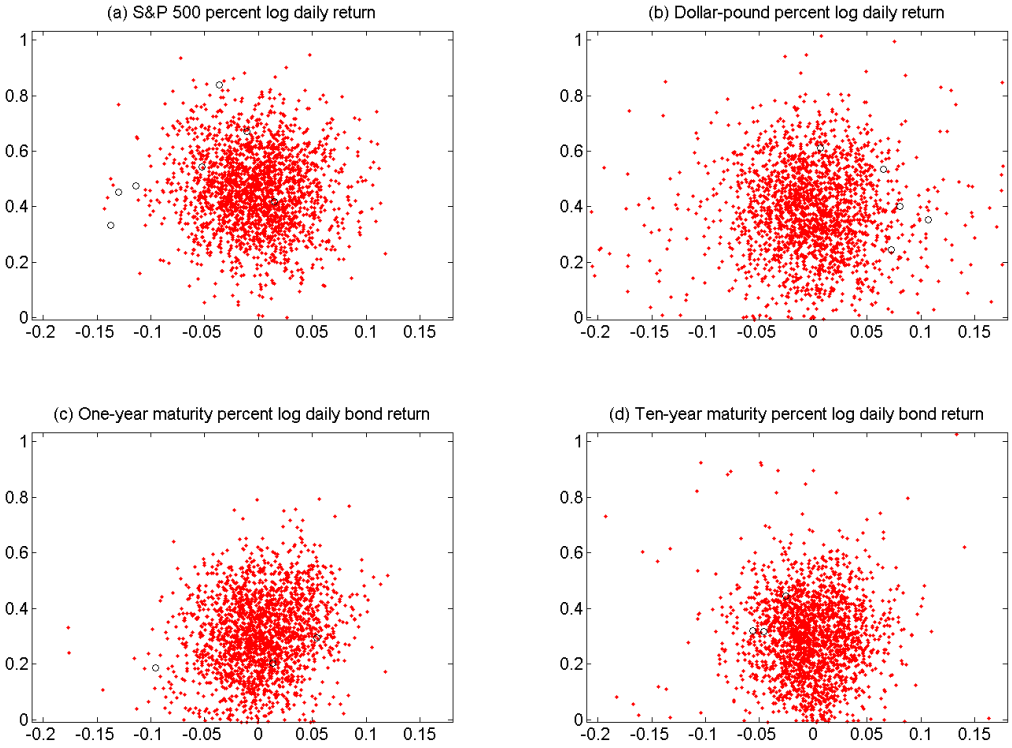

Figure 6: Full-sample posterior predictive distributions of one-day return autocorrelation $\widehat{\rho}_1$ (horizontal axis) and one-day absolute return long memory $\widehat{d}$ (vertical axis) in a five-year sample

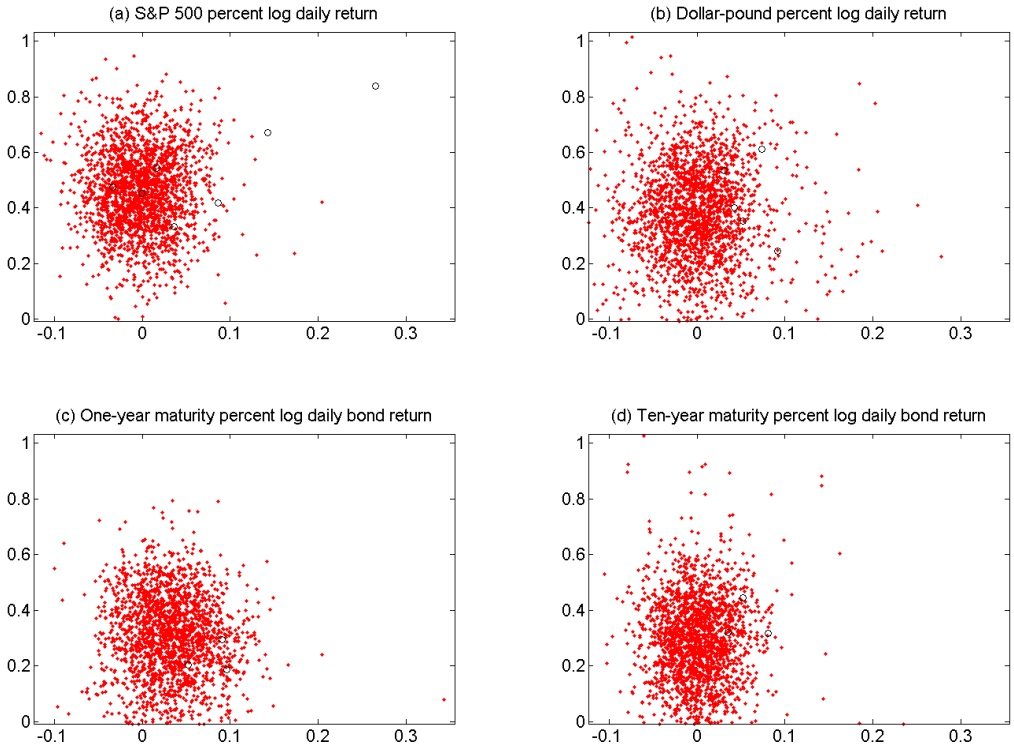

Figure 7: Full-sample posterior predictive distributions of the sum of the first 200 autocorrelations (horizontal axis) and the ratio of the sum to the first 200 absolute autocorrelations (vertical axis), for one-day absolute returns

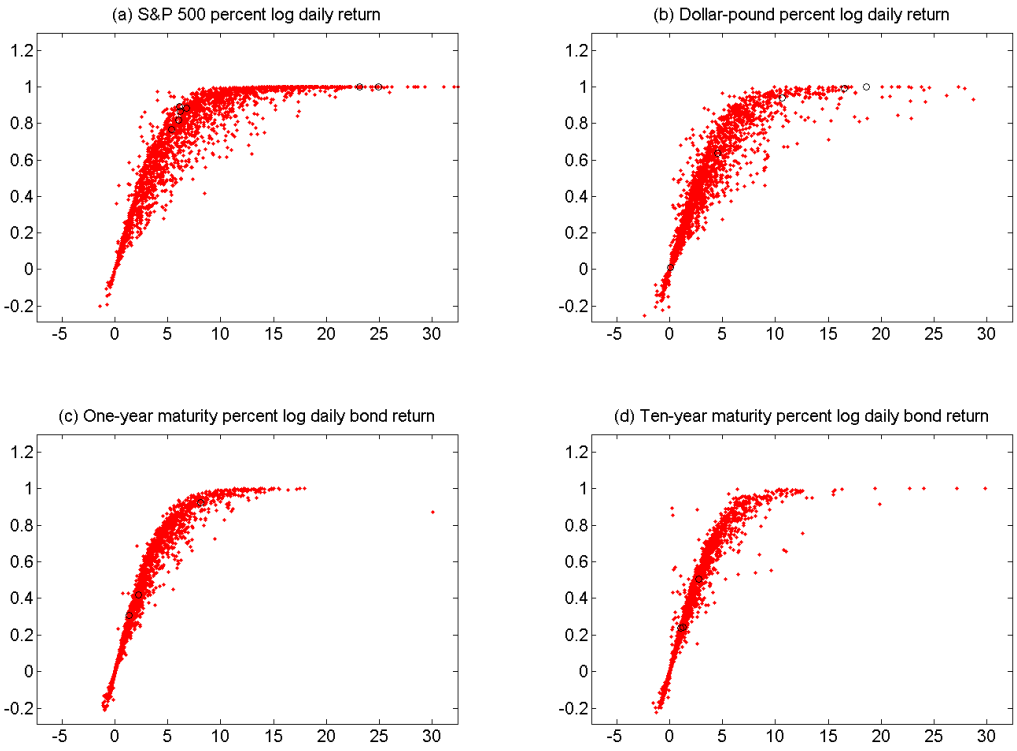

Figure 8: Full-sample posterior predictive distributions of skewness of one-day returns (horizontal axis) and skewness of ten-day returns (vertical axis) in a five-year sample

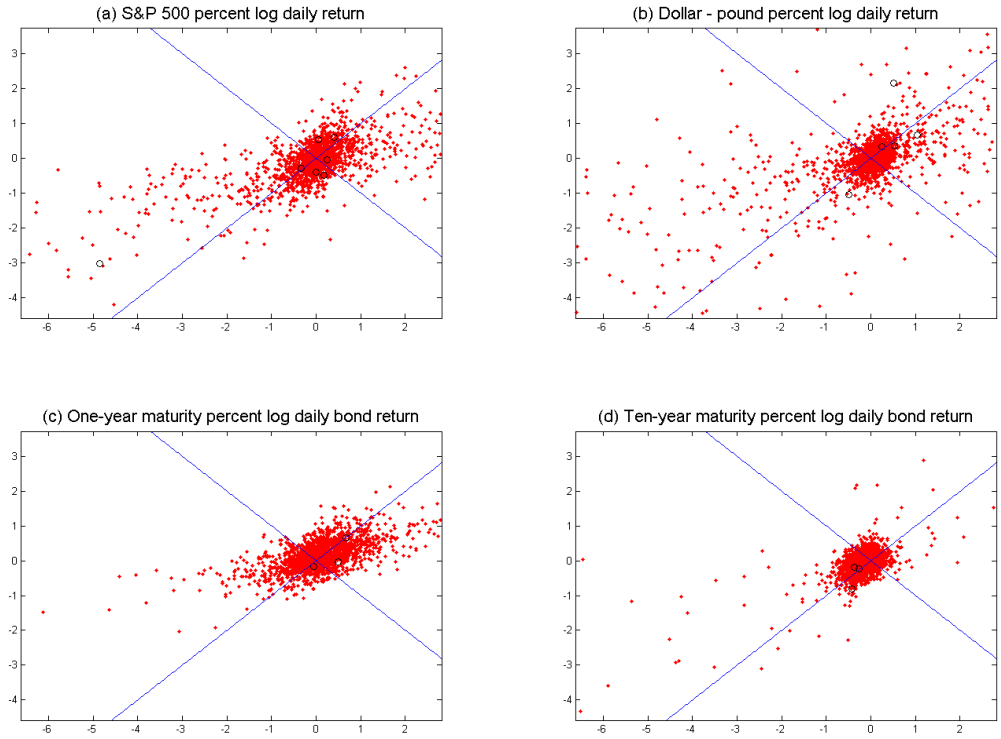

Figure 9: Full-sample posterior predictive distributions of autocorrelation $\widehat{\rho}_1$ of one-day returns (horizontal axis) and autocorrelation $\widehat{\rho}_1$ of nonoverlapping ten-day returns (vertical axis) in a five-year sample

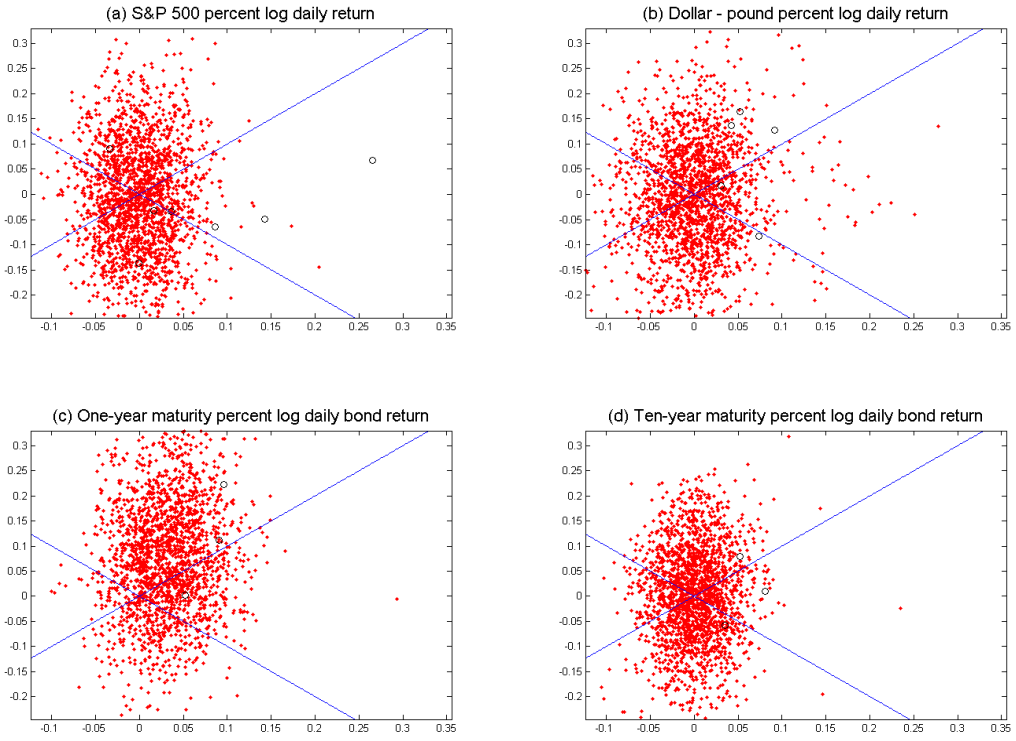

Figure 10: Full-sample posterior predictive distributions of autocorrelation $\widehat{\rho}_1$ of one-day returns (horizontal axis) and autocorrelation $\widehat{\rho}_1$ of one-day absolute returns (vertical axis) in a five-year sample

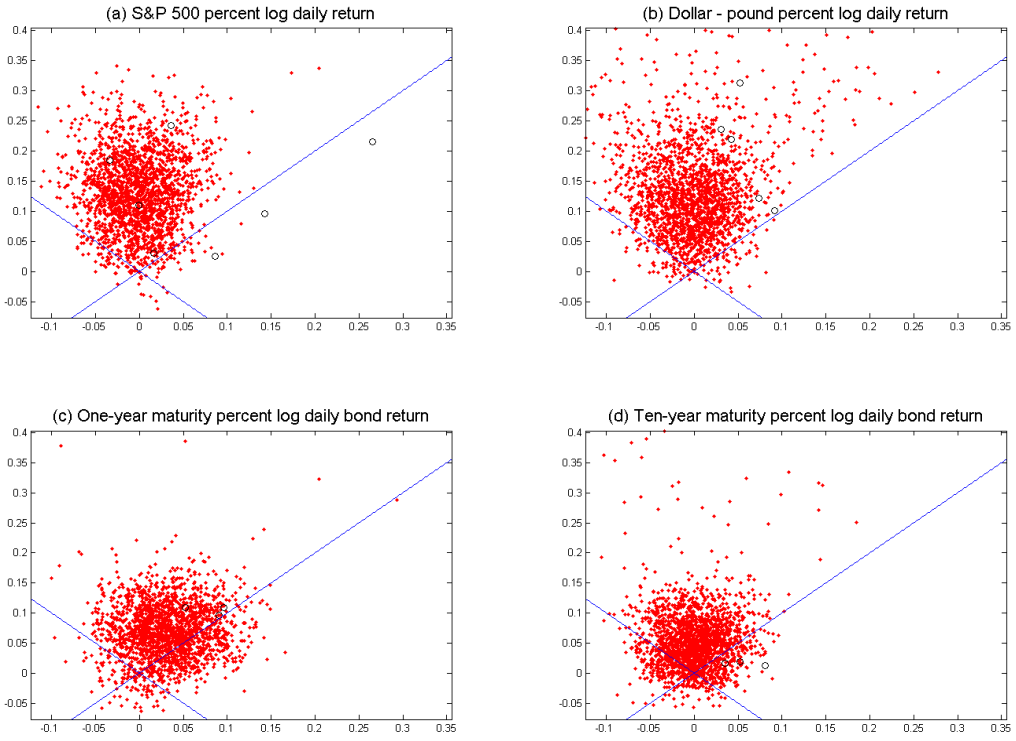

Figure 11: Full-sample posterior predictive distributions of autocorrelation $\widehat{\rho}_1$ of one-day absolute returns (horizontal axis) and autocorrelation $\widehat{\rho}_9$ of one-day absolute returns (vertical axis) in a five-year sample

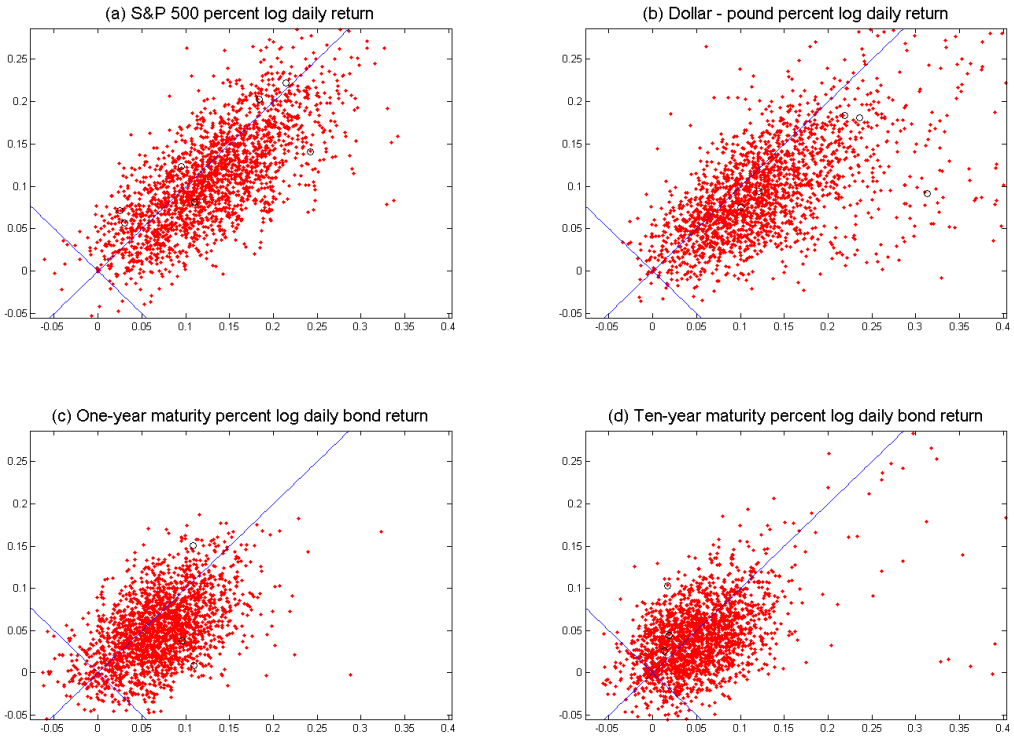

Figure 12: Full-sample posterior predictive distributions of the excess kurtosis coefficients in two successive five-year samples of one-day returns

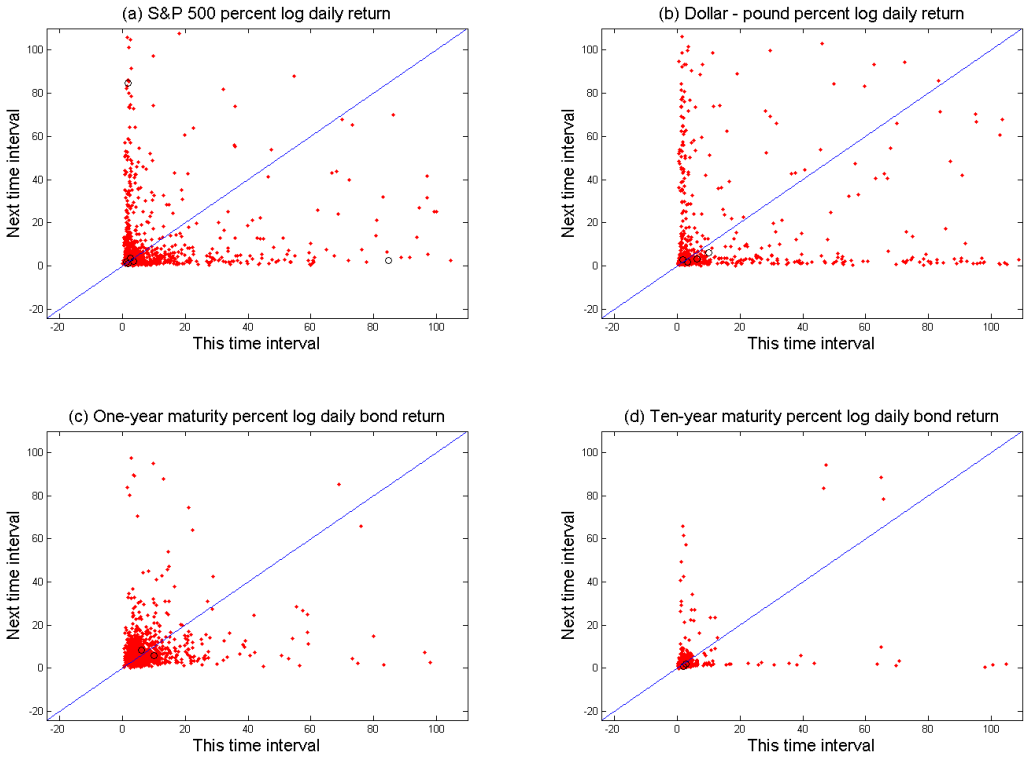

Figure 13: Full-sample posterior predictive distributions of skewness coefficients in two successive five-year samples of ten-day returns

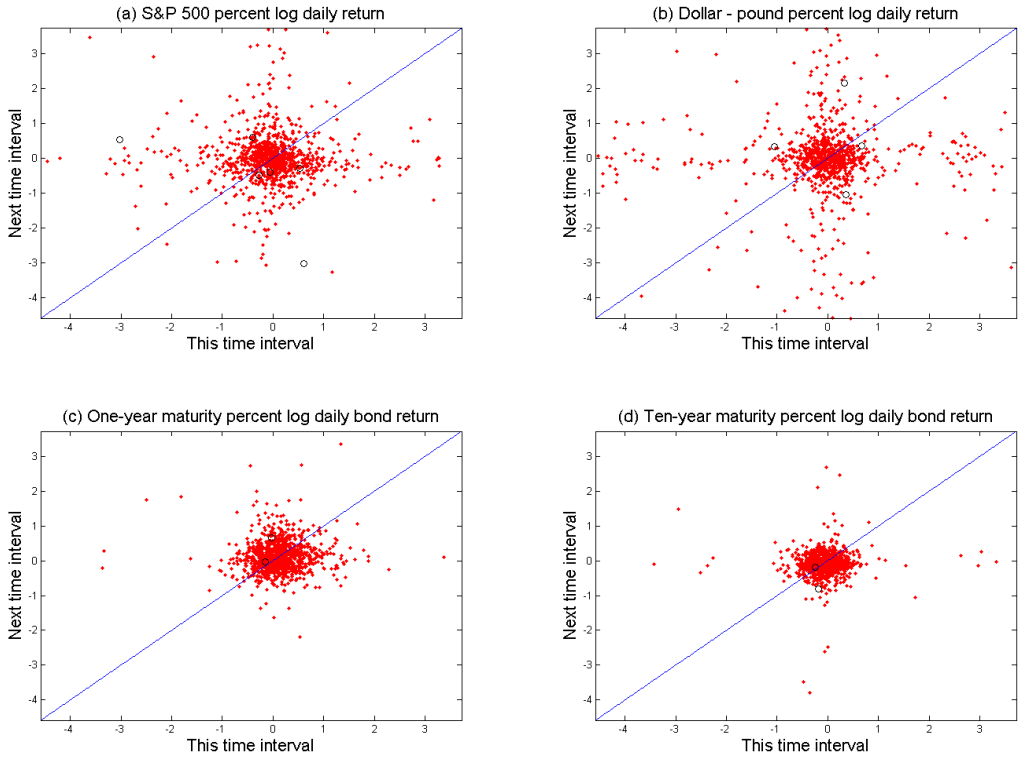

Figure 14: Full-sample posterior predictive distributions of long memory $\widehat{d}$ in two successive five-year samples

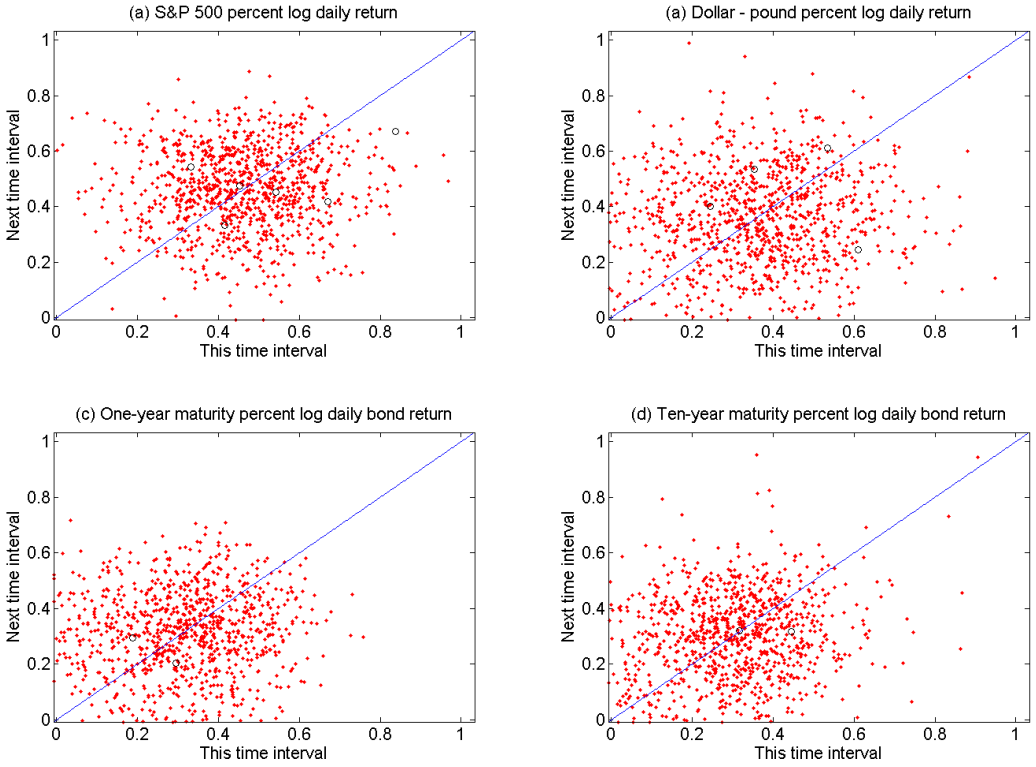

Figure 15: Full-sample posterior predictive distributions of the sum of the first 200 autocorrelation coefficients of absolute one-day returns in two successive five-year samples

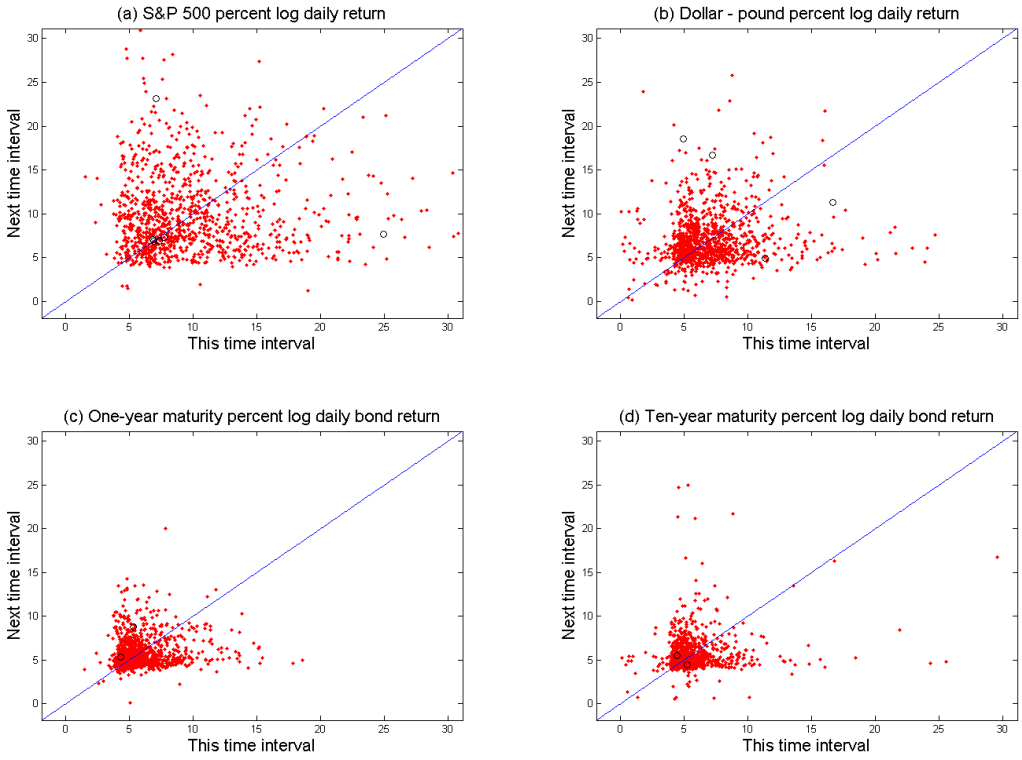

Figure 16: Quantiles of the predictive distribution of the term structure of S&P 500 returns, September 1992

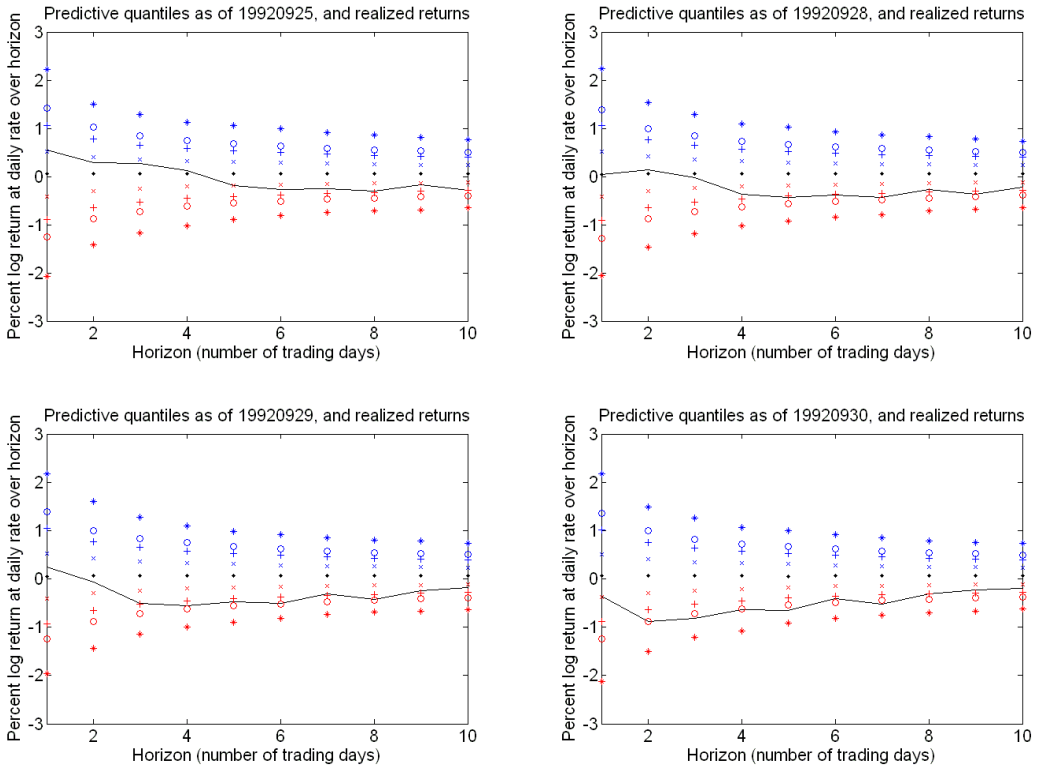

Figure 17: Quantiles of the predictive distribution of the term structure of S&P 500 returns, October 1987

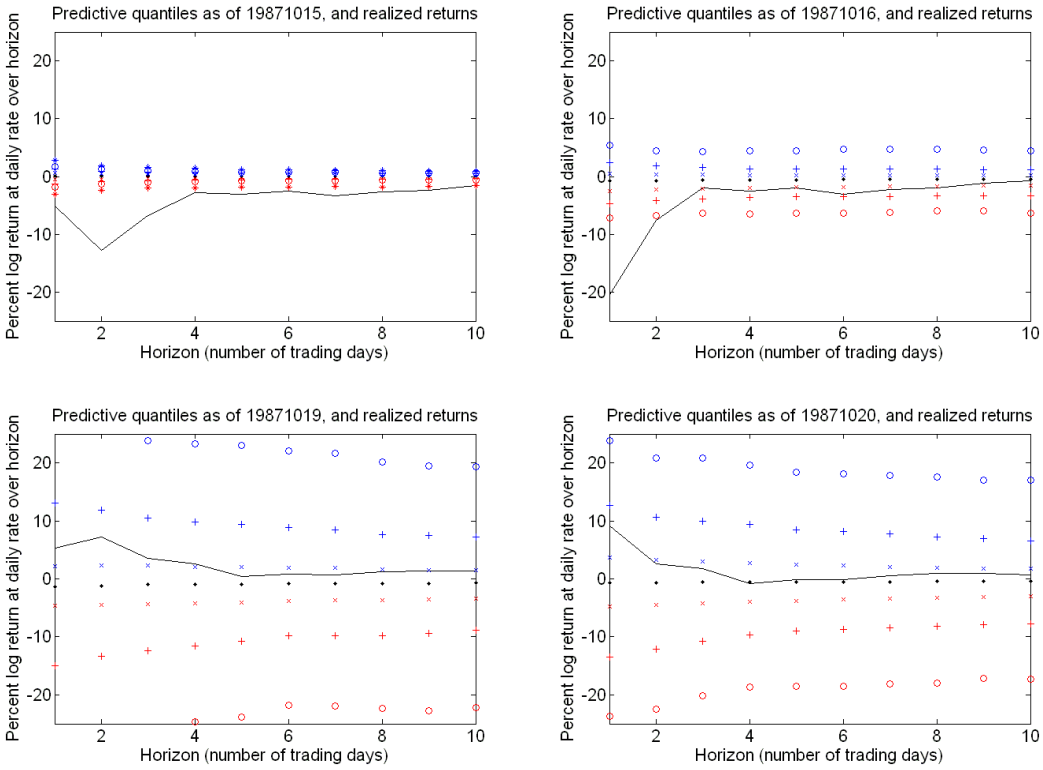

Figure 18: Quantiles of the predictive distribution of the term structure of S&P 500 returns, 1987-1988

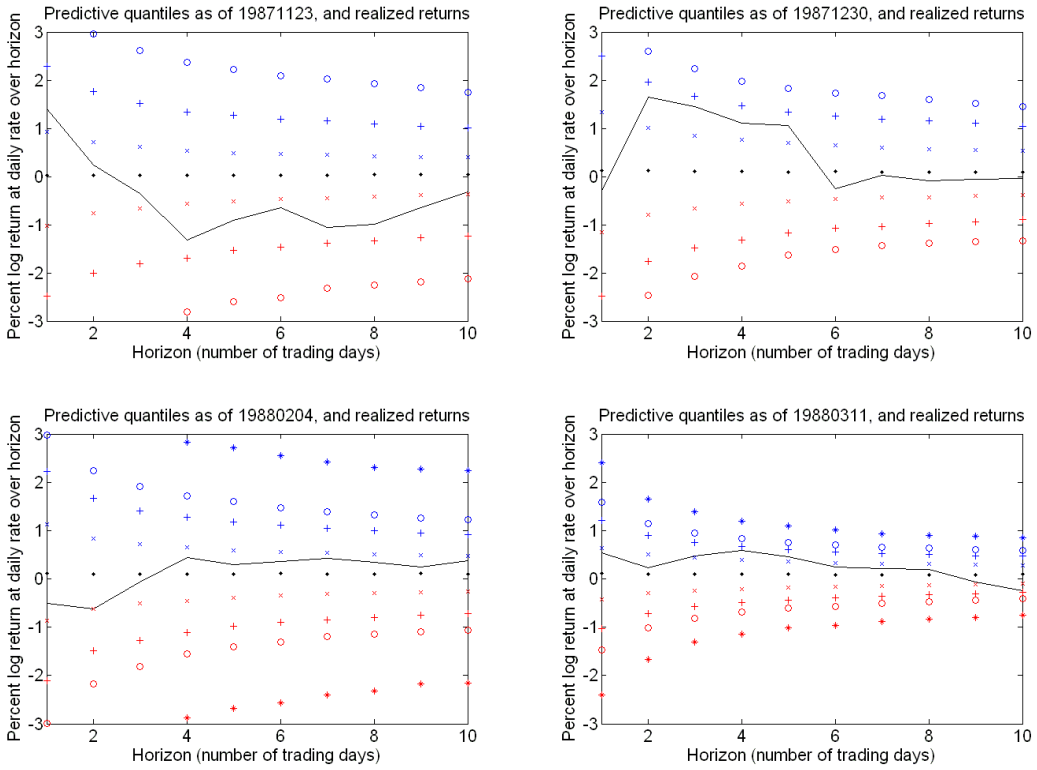

Figure 19: Quantiles of the predictive distribution of the term structure of dollar-pound returns, September 1992

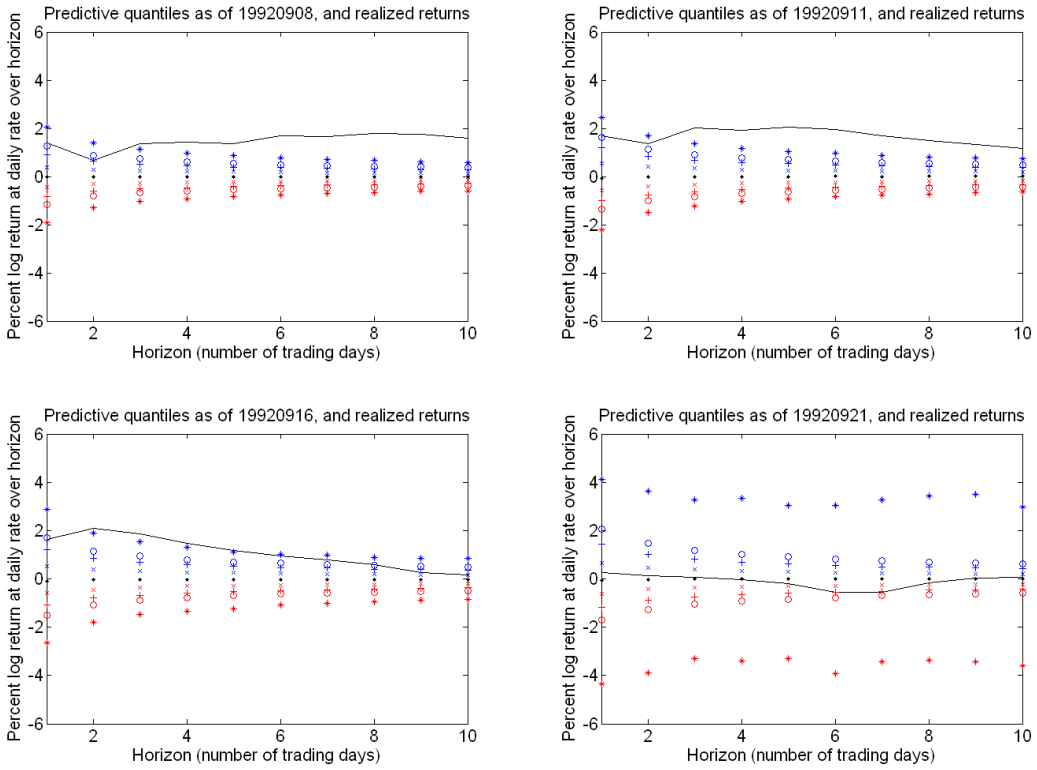

Figure 20: Quantiles of the predictive distribution of the term structure of one-year maturity bonds, December 2000 - January 2001

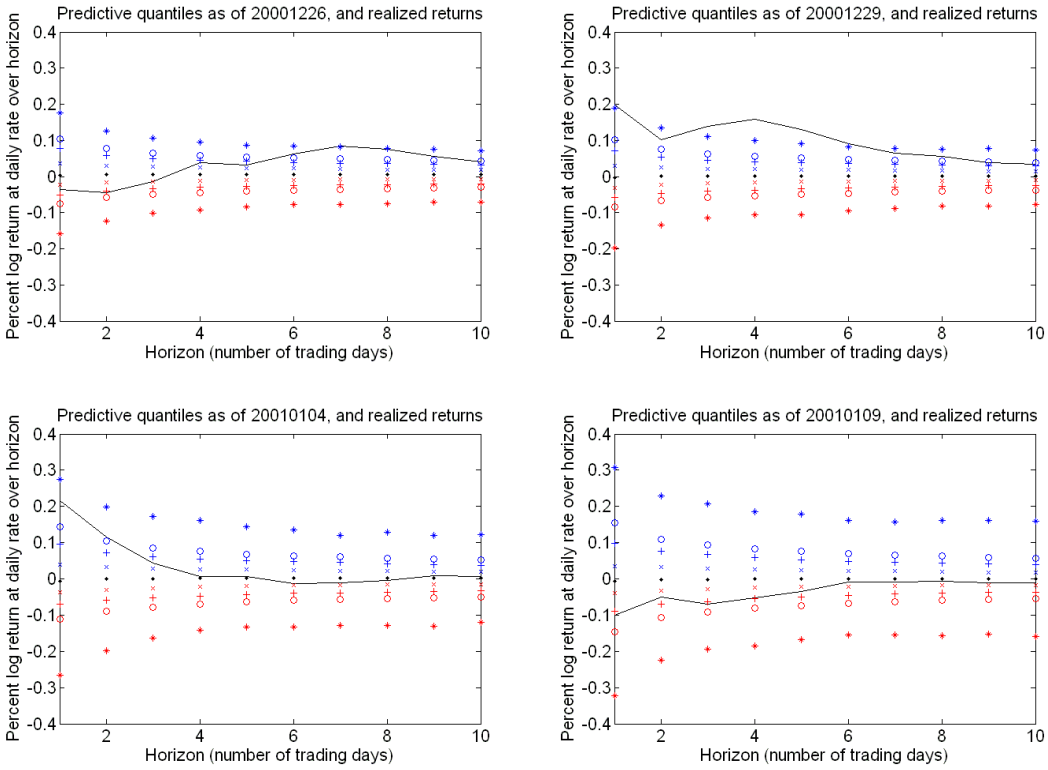

Figure 21: Quantiles of the predictive distribution of the term structure of ten-year maturity bonds, December 2000 - January 2001