

Risk, Return and Volatility Feedback: A Bayesian Nonparametric Analysis*

Mark J. Jensen[†] John M. Maheu[‡]

November 2013

Abstract

The relationship between risk and return is one of the most studied topics in finance. The majority of the literature is based on a linear, parametric relationship between expected returns and conditional volatility. However, there is no theoretical justification for the relationship to be linear. This paper models the contemporaneous relationship between market excess returns and realized variances nonparametrically with an infinite mixture representation of their joint distribution. With this nonparametric representation, the conditional distribution of excess returns given realized variance will also have a infinite mixture representation but with probabilities and arguments depending on the value of realized variance. Our nonparametric approach allows for deviation from Gaussianity by allowing for higher order non-zero moments. It also allows for a smooth nonlinear relationship between the conditional mean of excess returns and realized variance. Parsimony of our nonparametric approach is guaranteed by the almost surely discrete Dirichlet process prior used for the mixture weights and arguments. We find strong robust evidence of volatility feedback in monthly data. Once volatility feedback is accounted for, there is an unambiguous positive relationship between expected excess returns and expected log-realized variance. This relationship is nonlinear. Volatility feedback impacts the whole distribution and not just the conditional mean.

*We are grateful for helpful comments from conference participants at CFE'12, NBER-NSF SBIES 2013 and the Bayesian RCEA workshop 2013 and seminar participants at McMaster University. A previous version of this work was titled "A Bayesian Nonparametric Analysis of the Relationship between Returns and Realized Variance." We are grateful to Tom McCurdy who supplied the data. The views expressed here are ours and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Maheu is grateful to the SSHRC for financial support.

[†]Federal Reserve Bank of Atlanta, Mark.Jensen@atl.frb.org

[‡]DeGroote School of Business, McMaster University, University of Toronto, Canada, and RCEA, Italy, maheujm@mcmaster.ca

1 Introduction

This paper investigates the relationship between risk, return and volatility feedback by using a Bayesian nonparametric joint model for market excess returns and realized variance. A contemporaneous model for excess returns and realized variance is used to study volatility feedback effects which simultaneously impact prices.

The early literature found conflicting results on the sign and significance of the conditional variance from GARCH models in the conditional mean of market excess returns. A good summary of these results is found in Lettau & Ludvigson (2010). The recent literature has helped to resolve some of the issues. Scruggs (1998) and Gui & Whitelaw (2006) show that additional priced factors can affect the sign and significance of risk. Lundblad (2007) argues that longer samples are necessary to find a significant relationship between the market risk premium and expected volatility with GARCH specifications. Bandi & Perron (2008) document a long-run relationship between expected excess market returns and past market variance while Maheu & McCurdy (2007) find the long-run component of realized variance is priced in annual data. Ghysels et al. (2013) find a positive risk and return relationship over sample periods that exclude financial crises.¹ Using daily data Maheu et al. (2013) find the conditional variance and conditional skewness due to jumps is significantly priced. Ignoring higher order moments may confound the evidence for a positive risk and return relation.

Despite this little has been done to study the contemporaneous relationship between returns and realized variance while making no assumptions about their distribution and using it to investigate the role of volatility feedback. Our approach is related to Brandt & Kang (2004) and Harvey (2001) in that we jointly model returns and log-volatility contemporaneously, except we follow the advice of Harvey (2001) and dispense with parametric assumptions for conditional expectations and model the conditional expectation of returns nonparametrically, and like Ludvigson & Ng (2007), replace conditional volatilities with observed realized variances. This provides additional flexibility in modelling the joint distribution and provides a better signal on the variance by using daily data to estimate monthly ex post variance.

Harrison & Zhang (1999) use a seminonparametric approach (Gallant & Tauchen 1989) based on a Hermite polynomial expansion to estimate the conditional distribution. The leading term in the expansion is a Gaussian ARCH model for excess returns. They find a positive risk and return relation at long holding intervals of one and two years.

¹This paper updates the results in Ghysels et al. (2005) which had a coding error.

In contrast, we use an infinite mixture of distributions with a flexible Dirichlet process prior to jointly model excess returns and log-realized variance. From this the nonparametric conditional distribution of returns given realized variance consists of an infinite mixture representation whose unknown probabilities and arguments depend on the value of realized variance. This allows for general types of dependence in which the conditional mean of excess returns is a function of the contemporaneous realized variance. A Dirichlet process prior is assumed for the unknown distribution governing the mixture parameters; i.e., the distribution of mixture locations and probabilities are unknown and the Dirichlet process is the prior for this unknown distribution (see Ferguson (1973) and Lo (1984)). Our work extends Muller et al. (1996) and Taddy & Kottas (2010) to slice sampling (Walker 2007) methods to accommodate non-Gaussian data densities and nonconjugate priors. The Dirichlet process mixture (DPM) is the standard approach to Bayesian nonparametrics and performs well in practise.² A Markov chain Monte Carlo (MCMC) procedure delivers posterior samples from which estimates are obtained that account for model and distributional uncertainty.

Volatility feedback refers to a causal relationship between the variance and price changes. If volatility is priced and a positive volatility shock arrives, then all things equal, the required rate of return increases which discounts all future cash flows at a higher rate. This results in a simultaneous drop in the current price so as to deliver a higher future return consistent with the increase in risk. The importance of volatility feedback and its relationship to risk and return is discussed in French et al. (1987) and Campbell & Hentschel (1992).

Campbell & Hentschel (1992) show that volatility feedback is significant and important in finding a positive risk and return relationship. Ignoring it will tend to obscure any risk and return relationship. Moreover, volatility feedback can be an important source of return asymmetry. For instance, when good (bad) news arrives volatility increases and volatility feedback implies a drop in current prices which mutes (amplifies) the price increase (decrease). Therefore, volatility feedback amplifies the effect of bad news on prices and dampens positive news. This is why it is a leading explanation for asymmetric volatility.³ Therefore, price increases from good news will be less than what would occur without volatility feedback while price decrease from bad news will

²For example, see Burda et al. (2008), Conley et al. (2008), Delatola & Griffin (2013), Griffin & Steel (2004) and Chib & Greenberg (2010), Jensen & Maheu (2010, 2013a, 2013b) for recent applications of the DPM model.

³Papers by Bekaert & Wu (2000) and Wu (2001) find that volatility feedback is an important determinant of asymmetric volatility.

be steeper.

Campbell & Hentschel (1992) derive their model by imposing economic restrictions that linearly relate log-returns to log-prices and log-dividends.⁴ From this the impact of different sources of return shocks can be derived. Additional papers that build on this approach and find empirical support for volatility feedback include Turner et al. (1989), Kim et al. (2004), Kim et al. (2005), Bollerslev et al. (2006) and Calvet & Fisher (2007).

Our paper differs in several important ways from the existing literature. First, while almost all the literature has studied volatility feedback from a tightly parametrized model we use a flexible approach with no economic restrictions. Second, we use realized variance which is an accurate ex post measure of the variance of returns and permits the joint modelling of returns and variance. Third, we nonparametrically model the relationship between contemporaneous excess returns and log-realized variance. Volatility feedback implies an instantaneous causal relationship between volatility innovations and price levels or returns and our contemporaneous model is designed to investigate this relationship directly. Fourth, our nonparametric approach allows for conditioning on predetermined conditioning variables.

Using a long calendar span of monthly data we find strong robust evidence of volatility feedback in monthly data. Expected excess returns are always positive when volatility shocks are small, however, they become negative once the volatility shock becomes larger. This relationship is very nonlinear and depends on the current level of expected volatility. Ignoring these dynamics will result in confounding evidence for risk and return. Once volatility feedback is accounted for, there is an unambiguous positive relationship between expected excess returns and expected log-realized variance. This relationship is nonlinear.

Conditional quantile and contour plots support these findings and display significant deviations from the monotonic changes in the conditional distribution of the parametric model. We show that the volatility feedback effect impacts the whole distribution and not just the conditional mean.

This paper is organized as follows. The data and construction of realized variance are discussed in the next section followed by the nonparametric model for excess market returns and realized variance in Section 3. Section 4 discusses estimation of the conditional distribution and conditional mean of excess returns given log-realized variance. Empirical results are found in Section 5 followed by the Conclusion.

⁴The approximation is based on Campbell & Shiller (1988).

2 Return and realized variance data

Using high frequency daily returns permits the construction of monthly realized variance – an ex post, observable variance that is the focus of our study. Although realized variance has been used in empirical finance for some time (French et al. 1987) there exists a strong theoretical foundation for using it as an essentially nonparametric measure of ex post volatility (for recent reviews see Andersen & Benzoni (2008) and McAleer & Medeiros (2008)).

To compute realized variance, daily price data is obtained from Bill Schwert⁵ for 1885/2-1925/12, and from CRSP for 1926/1-2011/12 on the value-weighted portfolio with distributions for the S&P500. The data is converted to continuously compounded daily returns. If $r_{t,i}$ denotes the continuously compounded return for day i in month t then we compute realized variance according to

$$RV_t^q = \gamma_0 + 2 \sum_{j=1}^q (1 - j/(q+1)) \hat{\gamma}_j, \quad \hat{\gamma}_j = \sum_{i=1}^{N_t-j} r_{t,i} r_{t,i+j} \quad (1)$$

where N_t denotes the number of daily returns in month t . This estimate of realized variance contains a bias adjustment of order q to account for market microstructure dynamics and stale prices and follows Hansen & Lunde (2006). The Bartlett weights in (1) ensure that RV_t^q is always positive. In our work we set $q = 1$ and let $RV_t \equiv RV_t^q$ for the remainder of the paper.

Monthly returns are taken from the associated monthly files from Schwert and CRSP S&P500. The risk-free rate is obtained from Amit Goyal's website for 1885/2-1925/12, and after this time period the risk-free rate equals the 1 month rate from the CRSP Treasury bill file.

Our return-risk analysis dataset, thus, consists of monthly excess returns r_t and realized variance RV_t from 1885/1-2011/12 for a total of 1519 monthly observations. Returns are scaled by 12 and RV_t by 144 in order for our findings to be interpreted in terms of annual returns. When estimating the model we reserve the first 22 observations as conditioning variables. The information set is denoted by $I_t = \{r_1, RV_1, \dots, r_t, RV_t\}$, for $t = 1, \dots, T$.

Table 1 reports various summary statistics for monthly excess returns and realized variance. Compared to squared returns, realized variance is less noisy. Returns standardized by realized variance are approximately normal with sample skewness of 0.003

⁵For details on the construction of these data see Schwert (1990).

and sample kurtosis of 2.6856. Log-realized variance is closer to being bell-shaped than the levels of RV_t . Figure 1 displays a scatter plot of market excess returns and $\log(RV_t)$ which is the basis of our time-series models.

3 Nonparametric model of market excess returns and realized variance

In this section we provide some intuition on the nonparametric model we will use to specify the joint relationship between excess returns and realized variance and the implied conditional expectation of the return given realized variance. Since no theoretical reason exists for a particular parametric relationship to hold between the conditional mean and variance (Brandt & Kang 2004), we model the joint distribution nonparametrically by assuming the following infinite mixture representation for the conditional, joint, probability density function of excess returns and log-realized variance,

$$p(r_t, \log(RV_t)|I_{t-1}, \Omega, \Theta) = \sum_{j=1}^{\infty} \omega_j f(r_t, \log(RV_t)|\theta_j, I_{t-1}), \quad (2)$$

where $\Omega = (\omega_1, \omega_2, \dots)$ and $\Theta = (\theta_1, \theta_2, \dots)$, and $\omega_j \geq 0$, for all j such that $\sum_{j=1}^{\infty} \omega_j = 1$, and $f(\cdot, \cdot|\theta_j, I_{t-1})$ is a smooth bivariate density kernel given the parameter θ_j and information set I_{t-1} . This is a nonparametric model in the sense that there is an infinite number of parameters. It can approximate any continuous bivariate distribution to arbitrary accuracy by selecting the appropriate weight ω_j and parameter θ_j for the j th cluster. To reduce the clutter from carrying around the conditional mixture arguments, Θ and Ω , we drop them from $p(r_t, \log(RV_t)|I_{t-1}, \Omega, \Theta)$, when it is clear to do so.

In the next section we will discuss how a model containing of an infinite number of unknowns can be estimated with Bayesian methods, but for the moment we will consider how to obtain a nonparametric representation of the conditional distribution of excess returns that depends on log-realized variance from Eq. (2). From the mixture model above the conditional probability density function can be derived as

$$p(r_t|\log(RV_t), I_{t-1}) = \frac{p(r_t, \log(RV_t)|I_{t-1})}{p(\log(RV_t)|I_{t-1})} = \frac{\sum_{j=1}^{\infty} \omega_j f(r_t, \log(RV_t)|\theta_j, I_{t-1})}{\sum_{j=1}^{\infty} \omega_j f(\log(RV_t)|\theta_j, I_{t-1})} \quad (3)$$

$$= \sum_{j=1}^{\infty} q_j(\log(RV_t)) f(r_t|\log(RV_t), \theta_j, I_{t-1}) \quad (4)$$

where $f(r_t|\log(RV_t), \theta_j, I_{t-1}) \equiv f(r_t, \log(RV_t)|\theta_j, I_{t-1})/f(\log(RV_t)|\theta_j, I_{t-1})$ is the conditional density kernel for the j th cluster and $f(\log(RV_t)|\theta_j, I_{t-1})$ is the associated marginal

density kernel for $\log(RV_t)$. The weights of this mixture have the particular form,

$$q_j(\log(RV_t)) = \frac{\omega_j f(\log(RV_t)|\theta_j, I_{t-1})}{\sum_{i=1}^{\infty} \omega_i f(\log(RV_t)|\theta_i, I_{t-1})} \quad (5)$$

so that they sum to 1. From (5) we see that clusters that provide a better *fit* to $\log(RV_t)$ will receive more weight in the mixture representation (4). Varying $\log(RV_t)$ will produce smooth changes in the conditional distribution.

Our object of interest is the conditional expectation of market excess returns given log-realized volatility. Since the expectation of a mixture is equivalent to a mixture of the expectation, the expectation of (4) is the desired conditional expectation

$$E[r_t | \log(RV_t), I_{t-1}] = \sum_{j=1}^{\infty} q_j(\log(RV_t)) E[r_t | \log(RV_t), \theta_j, I_{t-1}], \quad (6)$$

where $E[r_t | \log(RV_t), \theta_j, I_{t-1}]$ is the j th cluster specific conditional expectation. A plot of the conditional expectation as a function of $\log(RV_t)$ will be a smoothly changing function that weights each of the cluster specific conditional expectations according to how the weight function $q_j(\log(RV_t))$ changes as $\log(RV_t)$ changes. This is true even if each cluster's expectation, $E[r_t | \log(RV_t), \theta_j, I_{t-1}]$, is constant. In this way we can see the contemporaneous relationship of log-volatility on the conditional mean of excess returns. As mentioned above, volatility feedback occurs simultaneously and this specification is designed to shed light on it.

3.1 A Bayesian model

The Dirichlet process prior has a long history, beginning with Ferguson (1973), of use in Bayesian nonparametric problems. It was used as a prior in countable infinite mixtures for density estimation in Ferguson (1983) and Lo (1984) but applications were limited until modern computational techniques. The seminal paper by Escobar & West (1995) show how to perform Bayesian nonparametric density estimation with Gibbs sampling.

Our approach is similar in that we place a Dirichlet process prior on ω_j and θ_j . In direct analogy to (2), according to Sethuraman (1994), the model can be represented as

$$p(r_t, \log(RV_t) | I_{t-1}, \Omega, \Theta) = \sum_{j=1}^{\infty} \omega_j f(r_t, \log(RV_t) | \theta_j, I_{t-1}), \quad (7)$$

where

$$\omega_1 = v_1, \quad \omega_j = v_j \prod_{i=1}^{j-1} (1 - v_i), \quad v_j \stackrel{iid}{\sim} \text{Beta}(1, \kappa) \quad (8)$$

$$\theta_j \stackrel{iid}{\sim} G_0. \quad (9)$$

The weights are generated by a stick breaking process since the unit interval is successively broken into smaller pieces by random draws from the beta distribution. Each cluster has a unique parameter θ_j independently drawn from the base distribution G_0 . The positive scalar κ controls the dispersion of the unit mass over the set Ω . A small value will put most of the weight on a few clusters while larger values will spread the weight over many clusters.

Another representation of (7)-(9) is in terms of the hierarchical model

$$r_t, \log(RV_t) | \theta_t^*, I_{t-1} \sim f(r_t, \log(RV_t) | \theta_t^*, I_{t-1}), \quad t = 1, \dots, T, \quad (10)$$

$$\theta_t^* | G \stackrel{iid}{\sim} G, \quad t = 1, \dots, T, \quad (11)$$

$$G | G_0, \kappa \sim DP(G_0, \kappa) \quad (12)$$

where the distribution of the mixture parameters, G , is unknown and modeled with the Dirichlet process distribution, $DP(G_0, \kappa)$, with precision parameter κ and base distribution G_0 .

The key quantity of interest is G – the unknown distribution of θ_t^* . Given the stick breaking definition of the Dirichlet process in (8)-(9), the prior distribution for G is

$$G(\theta_t^*) = \sum_{j=1}^{\infty} \omega_j \delta_{\theta_j}(\theta_t^*), \quad (13)$$

where $\delta_{\theta_j}(\cdot)$ denotes a point mass at θ_j and ω_j and θ_j are defined above. Hence, G will almost surely be a discrete distribution which means the θ_t^* s will contain repeats over $t = 1, \dots, T$. This clustering feature is one of the reasons the DP prior is so attractive. Several data observations can share the same mixture parameter vector. A set of θ_t^* s all having the same unique mixing parameter θ_{s_t} , where $s_t = j$ when $\theta_t^* = \theta_j$.

The DP prior is centered around G_0 in the sense that for any measurable set A , we have $E[G(A)] = G_0(A)$ while κ controls the precision $\text{Var}(G(A)) = G_0(A)(1 - G_0(A))/(1 + \kappa)$. As $\kappa \rightarrow \infty$ then $G \rightarrow G_0$. In this case, we have a mixture model with mixing measure G_0 . On the other hand, as $\kappa \rightarrow 0$ the mixture model is lost⁶ and is replaced by a parametric specification with parameter θ which has a fixed prior G_0 .

Based on theoretical considerations (Andersen et al. 2003), the empirical distribution of $\log(RV_t)$ being bell-shaped, and standardized excess returns being approximately normally distributed, we assume the joint kernel density, f , in (10) is,

$$f(r_t, \log(RV_t) | \theta, I_{t-1}) \equiv f(r_t | \log(RV_t), \theta, I_{t-1}) f(\log(RV_t) | \theta, I_{t-1}) \quad (14)$$

⁶The mixture collapses to a single cluster since $\omega_1 = 1$ and $\omega_j = 0$ for all $j > 1$.

where

$$f(r_t | \log(RV_t), \theta, I_{t-1}) \equiv f_N(r_t | \alpha_0 + \alpha_1 RV_t, \eta_1^2 RV_t) \quad (15)$$

$$\begin{aligned} f(\log(RV_t) | \theta, I_{t-1}) \equiv & f_N \left(\log(RV_t) \left| \gamma_0 + \gamma_1 \log(RV_{t-1}) + \gamma_2 \frac{1}{6} \sum_{i=1}^6 \log(RV_{t+1-i}) \right. \right. \\ & \left. \left. + \gamma_3 \frac{r_{t-1}}{\sqrt{RV_{t-1}}} + \gamma_4 \left| \frac{r_{t-1}}{\sqrt{RV_{t-1}}} \right|, \eta_2^2 \right) \end{aligned} \quad (16)$$

with $f_N(\cdot | \mu, \sigma^2)$ defined as the normal density kernel centered at μ with variance σ^2 and $\theta = \{\alpha_0, \alpha_1, \eta_1, \gamma_0, \dots, \gamma_4, \eta_2\}$.

Note, that although both excess returns and log-realized variance following conditional normal distributions their joint distribution is non-Gaussian since RV_t enters the variance of r_t . All of the elements in θ are permitted to differ over each cluster providing the maximum flexibility in modelling. For excess returns, RV_t can impact the conditional mean and the variance. Note that under certain conditions RV_t will be an unbiased estimate of the variance of returns but we allow for deviations that are captured by η_1 in the mixture model. The specification for $\log(RV_t)$ is along the lines of the models in Andersen et al. (2007), Corsi (2009) and the joint models of Maheu & McCurdy (2007, 2011), adapted to monthly data. It features a 6 month component to capture persistence beyond one month as well as asymmetric terms from lagged returns.

3.2 Posterior simulation

To sample the posterior density of this model we will exploit the mixture representation in (7) and a slice sampler based on Walker (2007), Kalli et al. (2011) and Papaspiliopoulos (2008).⁷ This Markov chain Monte Carlo (MCMC) approach introduces a random auxiliary, latent, variable, $u_t \in (0, 1)$, that slices away any mixtures clusters with a weight ω_j less than u_t . In this way the infinite mixture model is reduced to a finite mixture.

Introducing the latent variable u_t , we define the joint conditional density of the observed variables $(r_t, \log(RV_t))$ and u_t as,

$$p(r_t, \log(RV_t), u_t | \Omega, \Theta, I_{t-1}) = \sum_{j=1}^{\infty} \mathbf{1}(u_t < \omega_j) f(r_t, \log(RV_t) | \theta_j, I_{t-1}). \quad (17)$$

This infinite mixture is truncated to only include *alive* clusters with $u_t < \omega_j$ while *dead* clusters have a weight of 0 and can be ignored. If u_t has a uniform distribution then

⁷Alternative methods (Escobar & West 1995) based on the hierarchical form of the model in (10) are more difficult as our model and prior are non-conjugate.

integration of $p(r_t, \log(RV_t), u_t | \Omega, \Theta, I_{t-1})$ with respect to u_t gives back the original model $p(r_t, \log(RV_t) | \Omega, \Theta, I_{t-1})$. On the other hand, the marginal density of u_t is $\sum_{j=1}^{\infty} \mathbf{1}(u_t < \omega_j)$.

Let $s_t = j$ assign observation $(r_t, \log(RV_t))$ to the data density with parameter θ_j . We will augment the parameter space to include estimation of $S = (s_1, \dots, s_T)$. Let $U = (u_1, \dots, u_T)$, $\Omega_K = (\omega_1, \dots, \omega_K)$ and $\Theta_K = (\theta_1, \dots, \theta_K)$, then the full likelihood is

$$\prod_{t=1}^T p(r_t, \log(RV_t), u_t, s_t | \Omega_K, \Theta_K, I_{t-1}) = \prod_{t=1}^T \mathbf{1}(u_t < w_{s_t}) f(r_t, \log(RV_t) | \theta_{s_t}, I_{t-1}) \quad (18)$$

and the joint posterior is

$$p(\Omega_K) \left[\prod_{i=1}^K p(\theta_i) \right] \prod_{t=1}^T \mathbf{1}(u_t < w_{s_t}) f(r_t, \log(RV_t) | \theta_{s_t}, I_{t-1}), \quad (19)$$

where K is the smallest natural number that satisfies the condition $\sum_{j=1}^K \omega_j > 1 - \min\{U\}$. This value of K ensures that there are no $\omega_k > u_t$ for $k > K$. In other words, we have the set of all clusters that are alive, $\{j : u_t < w_j\}$.

Posterior simulation consists of sampling from the following densities:

1. $\pi(\theta_j | I_T, S) \propto g_0(\theta_j) \prod_{\{t: s_t = j\}} f(r_t, \log(RV_t) | \theta_j, I_{t-1})$, $j = 1, \dots, K$.
2. $\pi(v_j | S) \propto \text{Beta}(v_j | a_j, b_j)$, $j = 1, \dots, K$, with $a_j = 1 + \sum_{t=1}^T \mathbf{1}(s_t = j)$, $b_j = \kappa + \sum_{t=1}^T \mathbf{1}(s_t > j)$.
3. $\pi(u_t | \Omega_K, S) \propto \mathbf{1}(0 < u_t < w_{s_t})$, $t = 1, \dots, T$.
4. Find the smallest K such that $\sum_{j=1}^K w_j > 1 - \min\{U\}$.
5. $P(s_t = j | I_T, \Theta_K, U, \Omega_K) \propto \sum_{j=1}^K \mathbf{1}(u_t < \omega_j) f(r_t, \log(RV_t) | \theta_{s_t}, I_{t-1})$.

The first step depends on the model and the density $g_0(\cdot)$ to the DP priors base measure, G_0 . For the kernel densities in (15)-(16), specifying a normal prior for the regression coefficients and an independent inverse gamma prior for the variance, in other words, defining $G_0 \equiv N(\underline{b}, \underline{V}) \times G(\underline{v}/2, \underline{s}/2)$, we can employ standard Gibbs sampling techniques in Step 1 (see Greenberg (2013) for details on the exact form of these conditional distributions). Step 2 results from the conjugacy of the generalized Dirichlet distribution and multinomial sampling (Ishwaran & James 2001). Given Ω_K and S each u_t is uniformly distributed on $(0, \omega_{s_t})$. The next step updates the truncation parameter K . If K is incremented, Step 4 will also involve drawing additional ω_j and θ_j from the

DP prior. The final step is a multinomial draw of the cluster assignment variable s_t based on a mixture with equal weights.

Repeating all these steps forms one iteration of the sampler. The MCMC sampler yields the following set of variables at each iteration i ,

$$\{(\theta_{i,j}, v_{i,j}), j = 1, 2, \dots, K_i; (s_{i,t}, u_{i,t}), t = 1, \dots, T\}. \quad (20)$$

Note that $v_{i,j}, j = 1, 2, \dots, K_i$ implies $\omega_{i,j}, j = 1, 2, \dots, K_i$ through (8). After dropping the burnin phase from the above sampler, we collect $i = 1, \dots, N$ samples. Each iteration of the algorithm produces a draw of the unknown mixing distribution G from its posterior $[G|I_T]$ as

$$G_i = \sum_{j=1}^{K_i} \omega_{i,j} \delta_{\theta_{i,j}} + \left(1 - \sum_{j=1}^{K_i} \omega_{i,j}\right) G_0(\theta). \quad (21)$$

We will make use of this to form the predictive density and conditional expectations based on this.

4 Nonparametric conditional density estimation

To flexibly estimate the conditional density $p(r_t | \log(RV_t), I_{t-1})$ found in (4), or the conditional mean in (6), we use the method of Muller et al. (1996). This is an elegant approach to nonparametric estimation that allows the conditional density and expectation of excess returns to depend on covariates, in this case $\log(RV_t)$. The method requires the joint modelling of the predictor variable and its covariates and uses well know estimation methods for Dirichlet process mixture models. We extend Muller et al. (1996) to the slice sampler to accommodate the non-Gaussian data densities and non-conjugate priors found in our nonparametric model of market excess returns and realized variances.⁸

Based on the previous section, and given the history I_{t-1} , and G_i , the i th realization from the posterior of the joint predictive density for $(r_t, \log(RV_t))$ is

$$p(r_t, \log(RV_t) | G_i, I_{t-1}) = \int f(r_t, \log(RV_t) | \theta, I_{t-1}) G_i(d\theta). \quad (22)$$

⁸Additional papers that also build on Muller et al. (1996) are Rodriguez et al. (2009), Shahbaba & Neal (2009) and Taddy & Kottas (2010).

Substituting in the stick breaking representation for G_i found in (21), the posterior draw of the predictive density has the equivalent representation

$$\begin{aligned} p(r_t, \log(RV_t)|G_i, I_{t-1}) &= \sum_{j=1}^{K_i} w_{i,j} f(r_t, \log(RV_t)|\theta_{i,j}, I_{t-1}) \\ &\quad + \left(1 - \sum_{j=1}^{K_i} w_{i,j}\right) p(r_t, \log(RV_t)|G_0, I_{t-1}) \end{aligned} \quad (23)$$

where $p(r_t, \log(RV_t)|G_0, I_{t-1}) = \int f(r_t, \log(RV_t)|\theta, I_{t-1}) G_0(d\theta)$ is the expectation of (10). To integrate out the uncertainty associated with G , one averages (23) over the posterior realizations, $G_i \sim [G|I_T]$, $i = 1, \dots, N$, to obtain the posterior predictive density

$$p(r_t, \log(RV_t)|I_T) \approx \frac{1}{N} \sum_{i=1}^N p(r_t, \log(RV_t)|G_i, I_{t-1}). \quad (24)$$

Now, the conditional predictive density of r_t given $\log(RV_t)$ can be estimated as well. For each G_i draw, we have

$$\begin{aligned} p(r_t|\log(RV_t), G_i, I_{t-1}) &= \frac{p(r_t, \log(RV_t)|G_i, I_{t-1})}{p(\log(RV_t)|G_i, I_{t-1})} \\ &= \frac{p(r_t, \log(RV_t)|G_i, I_{t-1})}{\sum_{j=1}^{K_i} w_{i,j} f(\log(RV_t)|\theta_{i,j}, I_{t-1}) + (1 - \sum_{j=1}^{K_i} w_{i,j}) f(\log(RV_t)|G_0, I_{t-1})} \\ &= \sum_{j=1}^{K_i} q_{i,j}(\log(RV_t)) f(r_t|\log(RV_t), \theta_{i,j}, I_{t-1}) \\ &\quad + \left(1 - \sum_{j=1}^{K_i} q_{i,j}(\log(RV_t))\right) f(r_t|\log(RV_t), G_0, I_{t-1}) \end{aligned} \quad (25)$$

where $f(r_t|\log(RV_t), \theta_{i,j}, I_{t-1})$ is the conditional density of (15), $f(\log(RV_t)|\theta_{i,j}, I_{t-1})$ is the marginal density of (16) and

$$\begin{aligned} q_{i,j}(\log(RV_t)) &= w_{i,j} f(\log(RV_t)|\theta_{i,j}, I_{t-1}) \bigg/ \left[\sum_{l=1}^{K_i} w_{i,l} f(\log(RV_t)|\theta_{i,l}, I_{t-1}) \right. \\ &\quad \left. + \left(1 - \sum_{l=1}^{K_i} w_{i,l}\right) f(\log(RV_t)|G_0, I_{t-1}) \right]. \end{aligned} \quad (26)$$

The denominator of $q_{i,j}(\log(RV_t))$ is the marginal of (23) obtained by integrating out r_t . $f(\log(RV_t)|\theta_{i,j}, I_{t-1})$ is the marginal data density of $\log(RV_t)$ for the j th cluster with the marginal cluster parameter θ_j and $f(\log(RV_t)|G_0, I_{t-1})$ is the marginal data density

with mixing over the base measure. The terms in (25) and (26) involving G_0 are defined as follows

$$f(r_t | \log(RV_t), G_0, I_{t-1}) = \frac{\int f(r_t, \log(RV_t) | \theta, I_{t-1}) G_0(d\theta)}{\int f(\log(RV_t) | \theta, I_{t-1}) G_0(d\theta)} \quad (27)$$

$$f(\log(RV_t) | G_0, I_{t-1}) = \int f(\log(RV_t) | \theta, I_{t-1}) G_0(d\theta). \quad (28)$$

Assuming the marginal data density $f(\log(RV_t) | \theta, I_{t-1})$ is available in analytic form both of these expressions can be approximated by the usual MCMC methods. For instance, $f(\log(RV_t) | G_0, I_{t-1}) \approx N^{-1} \sum_{i=1}^N f(\log(RV_t) | \theta^{(i)}, I_{t-1})$, $\theta^{(i)} \sim G_0$, with a similar expression for the numerator of (27).

The predictive conditional density is estimated by averaging (25) over the posterior simulations of G_i as

$$p(r_t | \log(RV_t), I_T) \approx \frac{1}{N} \sum_{i=1}^N p(r_t | \log(RV_t), G_i, I_{t-1}). \quad (29)$$

Using this result, features of the conditional distribution such as conditional quantiles can be derived.

4.1 Nonparametric conditional mean estimation

Our focus will be on the conditional expectation which can be estimated from these results. First, the conditional expectation of r_t given $\log(RV_t)$ and G_i is

$$\begin{aligned} E[r_t | \log(RV_t), G_i, I_{t-1}] &= \sum_{j=1}^{K_i} q_{i,j}(\log(RV_t)) E[r_t | \log(RV_t), \theta_{i,j}, I_{t-1}] \\ &\quad + \left(1 - \sum_{j=1}^{K_i} q_{i,j}(\log(RV_t)) \right) E[r_t | \log(RV_t), G_0, I_{t-1}] \end{aligned} \quad (30)$$

where $E[r_t | \log(RV_t), G_0, I_{t-1}]$ is taken with respect to (27). Note that this final term is only a function of G_0 and can be computed once, at the start of estimation, for a grid of values of $\log(RV_t)$. It is estimated as⁹

$$\begin{aligned} E[r_t | \log(RV_t), G_0, I_{t-1}] &= \frac{\int E[r_t | \log(RV_t), \theta, I_{t-1}] f(\log(RV_t) | \theta, I_{t-1}) G_0(d\theta)}{\int f(\log(RV_t) | \theta, I_{t-1}) G_0(d\theta)} \quad (31) \\ &\approx \frac{M^{-1} \sum_{i=1}^M E[r_t | \log(RV_t), \theta^{(i)}, I_{t-1}] f(\log(RV_t) | \theta^{(i)}, I_{t-1})}{M^{-1} \sum_{i=1}^M f(\log(RV_t) | \theta^{(i)}, I_{t-1})} \quad (32) \end{aligned}$$

⁹This result makes use of expressing the numerator as $\int xp(x, y | \theta) p(\theta) d\theta dx = \int xp(x | y, \theta) p(y | \theta) p(\theta) d\theta dx = \int E[x | y, \theta] p(y | \theta) p(\theta) d\theta$.

for $\theta^{(i)} \sim G_0$, $i = 1, \dots, M$.

Given G_i , equation (30) shows the conditional expectation of r_t is a convex combination of cluster specific conditional expectations $E[r_t | \log(RV_t), \theta_j, I_{t-1}]$, $j = 1, \dots, K_i$, along with the expectation taken with respect to the base measure G_0 . The weighting function changes with the conditioning variable $\log(RV_t)$, which in turn changes for each I_{t-1} .

Finally, with this we can obtain the predictive conditional mean estimate by averaging over (30) as follows

$$E[r_t | \log(RV_t), I_T] \approx \frac{1}{N} \sum_{i=1}^N E[r_t | \log(RV_t), G_i, I_{t-1}], \quad (33)$$

to integrate out uncertainty concerning G .¹⁰ Point-wise density intervals of the conditional mean can be estimated from the quantiles of $E[r_t | \log(RV_t), G_i, I_{t-1}]$.

We evaluate the predictive conditional mean for a grid of values over $\log(RV_t)$. This will produce a smooth curve for date t and we will have a unique curve for every observation in our sample $t = 1, \dots, T$.

5 Results

For our analysis we specify the following priors. The base measure G_0 contains priors for each regression parameter in (15) and (16) as independent $N(0, 1)$ while $\eta_1^{-2} \sim G(5/2, 5/2)$ and $\eta_2^{-2} \sim G(6/2, 3/2)$ where $G(a, b)$ denotes a gamma distribution with mean a/b . Note that we expect η_1^2 to be close to 1 and the prior reflects this with $E[\eta_1^{-2}] = 1$ but allows for deviations from this. The precision parameter of the Dirichlet process is estimated and has a prior $G(2, 10)$. Each cluster contains 9 parameters in θ_j .

We use 5000 initial iterations of the posterior sampler for burn-in and then collect the following 20000 for posterior inference. The Markov chain mixes well and the posterior mean (0.95 density interval) for κ is 0.2046, (0.0439, 0.4831) and the posterior mean (0.95 density interval) for the number of alive clusters is 2.6, (2, 4). In other words, about 2.6 components are used to fit the joint model of r_t and $\log(RV_t)$.

Before we turn to the nonparametric estimates a parametric version of the model is reported in Table 2. This is a one state model. The coefficient α_1 on RV_t in the excess return equation is significantly negative. η_1^2 is close to 1 and indicates no systematic

¹⁰Note that the quantity $E[r_t | \log(RV_t), I_{t-1}]$ in (6) assumes parameters are known. In our case they need to be estimated by the posterior density using the full sample of data I_T . Therefore our estimate includes the conditioning argument I_T in $E[r_t | \log(RV_t), I_T]$.

bias in RV_t . The estimates of γ_1 and γ_2 indicate persistence in $\log(RV_t)$. The lagged standardized excess return terms entering the log-volatility equation show asymmetry. A negative return shock results in a larger conditional mean for log-volatility next period compared to a positive shock.

The conditional expectation of excess returns given log-realized variance is computed for a grid of 100 values over -4.0 to 2.0. We interpolate with a straight line between each of the grid values to approximate the smooth curve $E[r_t | \log(RV_t), I_T]$ as a function of $\log(RV_t)$. This is done for every observation t , in our sample.

Figure 2 displays the relationship between expected excess returns and $\log(RV_t)$ for the parametric model. Although the model specifies a linear relation between excess returns and RV_t , this yields a nonlinear relation with $\log(RV_t)$. Note that this curve holds for every time t in our sample and is not affected by low or high volatility periods.

In contrast, Figure 3 displays the conditional expectation of excess returns as a function of log-realized variance for every information set in our dataset for the non-parametric model. Overall there is a general increase in the conditional mean of excess returns as log-realized variance increases from low levels to a point in which expected returns become negative. This is a general pattern that is found in all the plots. However, the point of decrease in the conditional expectation differs. It is clear that if one averaged over these expectations you could obtain a positive value for expected excess returns or a negative value.¹¹ To really see what is happening we need to consider the conditional expectation and the innovation of log-volatility as well.

To better understand what is happening we isolate three typical periods of low, average and high volatility periods from our sample and report the conditional expectations in Figure 4-6. Each figure contains the conditional expectation of market excess returns given a range of log-realized variance values as well as the conditional expectation of log-realized variance (blue) and the resulting realized value of log-realized variance (black). Point wise 0.9 probability density intervals are included for the expected excess return. Recall the discussion of volatility feedback in Section 1. Only if the log-variance happens to occur on the expected value of log-realized variance is the volatility feedback effect 0. Values of log-realized variance above (below) the expected value are positive (negative) shocks to volatility. As was discussed, this will have a simultaneous impact on current prices, if volatility risk is priced, and result in a decrease (increase) in prices.

This is exactly what these figures show for a positive volatility shock. For instance,

¹¹In fact averaging the curves from the nonparametric model would give something close to the parametric model in Figure 2.

consider Figure 4 which conditions on a low period of volatility, 1964:10.¹² Expected log-realized variance is -3.158 . The expected excess return is positive before and briefly after this value but eventually becomes negative. Before -3.158 there is a gentle increase in the expectation of r_t but after it there is a strong decrease to negative values. In other words, if the volatility shock is positive and sufficiently large we expect a contemporaneous decrease in prices from volatility feedback.

Figure 5 displays a similar pattern for an average value of volatility from 1996:2. Here expected $\log(RV_t)$ is -2.117 . As before, the expected excess return is positive before this and to the right remains positive but eventually become negative. If the log-volatility shock is sufficiently large (about $+0.68$) then the expected excess return is negative and decreases as the shock increases. For this time period the realized $\log(RV_t)$ was -1.43 . Finally, notice that the whole $E[r_t | \log(RV_t), I_T]$ curve has shifted rightward as the expected $\log(RV_t)$ has increased from Figure 4 to 5 (low to average $\log(RV_t)$). This suggest an increase in compensation for the higher perceived risk.

A high volatility period corresponding to 2008:12 is found in Figure 6. Just as before, $E[r_t | \log(RV_t), I_T]$ is essentially linear and flat as $\log(RV_t)$ is increased prior to $E[\log(RV_t) | I_T]$ but after this point the expectation of excess returns become negative.¹³ This is consistent with a volatility feedback effect. Note that in each of the three figures the effect of volatility feedback on returns appears to be stronger with an increased slope in moving from expected low to average to high volatility periods.

Figure 7 plots $E[r_t | \log(RV_t), I_T]$ for each of the 3 episodes of expected $\log(RV_t)$. As $E[\log(RV_t) | I_T]$ increases the conditional expectation of excess returns shifts rightward and up. This is consistent with a positive and increasing reward for baring higher levels of risk.

In summary, we find a robust volatility feedback effect which is most notable for positive shocks to volatility. Expected excess returns are positive below $E[\log(RV_t) | I_T]$ but after this value eventually become negative. This suggests that risk is priced and the previous figure was consistent with this.

¹²From Table 1, average $\log(RV_t)$ is -1.5602 with a minimum of -4.4595 and maximum of 2.4245 .

¹³ $E[\log(RV_t) | I_T]$ denotes the in-sample Bayesian estimate of the expectation of $\log(RV_t)$ given I_{t-1} . This conditions on regressors in the information set $t - 1$ but uses the full posterior density based on $t = 1, \dots, T$ for the model parameters to integrate out parameter uncertainty.

5.1 Risk and return

To focus on risk and return we need to account for the volatility feedback effect. In each of our figures the point in the $E[r_t | \log(RV_t), I_T]$ line which corresponds to $\log(RV_t) = E[\log(RV_t) | I_T]$ is exactly the point with no volatility feedback. This point is where the investor receives exactly the reward for risk with no adjustment for volatility feedback because the volatility shock is 0. This will be at a different place in each of our curves of $E[r_t | \log(RV_t), I_T]$. Using interpolation between each of the grid values we can estimate the value of $E[r_t | \log(RV_t), I_T]$ at $\log(RV_t) = E[\log(RV_t) | I_T]$ for each time period t . This represents a *pure* risk and return relationship which nets out volatility feedback.

Figure 8 displays the pure risk and return relationship. It shows the expected excess return as a function of expected log-realized variance according to our model estimates when volatility feedback is removed. Each dot represents the point of $E[r_t | \log(RV_t), I_T]$ in which volatility feedback is zero for date t . The relationship is unambiguously positive and increasing in $\log(RV_t)$ which accords with theory. The relationship is nonlinear. It is approximately linear for small value of log-volatility but increases sharply as expected log-volatility surpasses 0.

In contrast to Campbell & Hentschel (1992) and the subsequent literature on volatility feedback, we find evidence of a positive risk and return relationship and a volatility feedback effect without imposing any economic restrictions. The key is flexibly modelling the contemporaneous distribution of market excess returns and log-realized variance and accounting for the volatility shock.

5.2 Conditional quantiles and contour plots

Figures 9 - 12 display conditional quantile plots of the distribution of excess returns given different levels of $\log(RV_t)$ for the parametric model and several cases of the nonparametric model. In each figure the green line is the conditional mean that was discussed above.

For the parametric model, as before, the conditional quantiles do not change for different information sets. However, the estimated weights and component densities in the mixture model of (4) are sensitive to the information set and result in very different conditional distributions. Each of the conditional quantile plots show a highly nonlinear distribution that is at odds with the parametric model.

Recall from the previous discussion that the conditional expectations of the low, average and high levels of $\log(RV_t)$ were -3.158 , -2.117 and 0.509 , respectively. At

each of these points in Figures 10 - 12 the bulk of the distribution is above 0. At these points the investor is most likely to receive a positive excess return from the market. As $\log(RV_t)$ increases and the volatility shock becomes larger, most of the mass in each conditional density is over a negative range of excess returns. Here the investor is likely to have a loss from investing in the market. Volatility feedback has an impact on the whole distribution and not just the conditional mean. The changes in the density, as $\log(RV_t)$ increases, are non-monotonic. In Figures 10 and 11 there is an increase in the spread of the density followed by a decrease and final increase. The point of these changes in the conditional density is to the right of the conditional mean of $\log(RV_t)$. The parametric quantile plot is inconsistent with these features.

Contour plots of the predictive density for $(r_t, \log(RV_t))$, over the selected time periods discussed above, are found in Figure 13-15. Each of the figures are consistent with deviations from Gaussian behaviour in the conditional bivariate distribution. It is clear that the conditional distribution changes a great deal over time and is not a result of changes in location and/or scale. There is a thick tail for small values of r_t and larger values of $\log(RV_t)$ in each figure but the shape of this tail is very different depending on t . These important changes in the conditional density are the features that our non-parametric model are designed to capture. Conventional parametric approaches cannot accommodate these features.

5.3 Parameter estimates and robustness

Figures 16 and 17 display the posterior mean of each of the model parameters as a function of the drawn state s_t over the sample period. A one state model would be a straight line. We see considerable switching between clusters in all plots and the changes between states is often large. This shows that multiple states in the mixture is a significant feature of the data. Compared to the parametric model in Table 2, α_1 the coefficient on RV_t , is negative and positive over different time periods. The variability of the parameters in the figures is well beyond the 0.95 density intervals for the parametric model reported in Table 2. Although the parametric model estimate of η_1^2 is about 1 the nonparametric estimate varies between 0.4 to 0.85. This is due to the significantly improved fit that the nonparametric model offers in the conditional mean which contributes to a lower innovation variance.

Our results are robust to changes in the priors and the model for the data density. For instance, we obtain the same qualitative results for $E[r_t | \log(RV_t), I_T]$ if we omit RV_t

($\alpha_1 = 0$) in (15) or drop the lagged return terms in $\log(RV_t)$, ($\gamma_3 = \gamma_4 = 0$). Although our priors are quite diffuse and provide a wide range of empirically realistic parameter values, making them more diffuse produces similar results but the density intervals for $E[r_t | \log(RV_t), I_T]$ are generally larger.

6 Conclusion

This paper nonparametrically models the contemporaneous relationship between market excess returns and realized variances. An infinite mixture of distributions is given a flexible Dirichlet process prior. From this the nonparametric conditional distribution of returns given realized variance consists of an infinite mixture representation whose probabilities and arguments depend on the value of realized variance. This allows for a smooth nonlinear relationship between the conditional mean of market excess returns and realized variance. The model is estimated with MCMC techniques based on slice sampling methods that extends the posterior sampling methods in the literature.

Applied to a long span of monthly data we find strong robust evidence of volatility feedback. Once volatility feedback is accounted for, there is an unambiguous positive relationship between expected excess returns and expected log-realized variance. In contrast to the existing literature, we find evidence of a positive risk and return relationship and a volatility feedback effect without imposing any economic restrictions. We show that the volatility feedback impacts the whole distribution and not just the conditional mean.

References

- Andersen, T., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), ‘Modeling and forecasting realized volatility’, *Econometrica* **71**, 529–626.
- Andersen, T. G. & Benzoni, L. (2008), Realized volatility. FRB of Chicago Working Paper No. 2008-14, <http://ssrn.com/abstract=1092203>.
- Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2007), ‘Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility’, *The Review of Economics and Statistics* **89**(4), 701–720.
- Bandi, F. M. & Perron, B. (2008), ‘Long-run risk-return trade-offs’, *Journal of Econometrics* **143**(2), 349 – 374.
- Bekaert, G. & Wu, G. (2000), ‘Asymmetric volatility and risk in equity markets’, *The Review of Financial Studies* **13**(1), 1–42.
- Bollerslev, T., Litvinova, J. & Tauchen, G. (2006), ‘Leverage and volatility feedback effects in high-frequency data’, *Journal of Financial Econometrics* **4**(3), 353–384.
- Brandt, M. W. & Kang, Q. (2004), ‘On the relationship between the conditional mean and volatility of stock returns: A latent var approach’, *Journal of Financial Economics* **72**(2), 217–257.
- Burda, M., Harding, M. & Hausman, J. (2008), ‘A bayesian mixed logitprobit model for multinomial choice’, *Journal of Econometrics* **147**(2), 232 – 246.
- Calvet, L. E. & Fisher, A. J. (2007), ‘Multifrequency news and stock returns’, *Journal of Financial Economics* **86**(1), 178–212.
- Campbell, J. Y. & Hentschel, L. (1992), ‘No news is good news: An asymmetric model of changing volatility in stock returns’, *Journal of Financial Economics* **31**, 281–318.
- Campbell, J. Y. & Shiller, R. J. (1988), ‘The dividend-price ratio and expectations of future dividends and discount factors’, *Review of Financial Studies* **1**(3), 195–228.
- Chib, S. & Greenberg, E. (2010), ‘Additive cubic spline regression with dirichlet process mixture errors’, *Journal of Econometrics* **156**(2), 322 – 336.

- Conley, T. G., Hansen, C. B., McCulloch, R. E. & Rossi, P. E. (2008), ‘A semi-parametric bayesian approach to the instrumental variable problem’, *Journal of Econometrics* **144**(1), 276 – 305.
- Corsi, F. (2009), ‘A simple approximate long-memory model of realized volatility’, *Journal of Financial Econometrics* **7**(2), 174–196.
- Delatola, E.-I. & Griffin, J. (2013), ‘A bayesian semiparametric model for volatility with a leverage effect’, *Computational Statistics & Data Analysis* **60**(0), 97 – 110.
- Escobar, M. D. & West, M. (1995), ‘Bayesian density estimation and inference using mixtures’, *Journal of the American Statistical Association* **90**(430), 577–588.
- Ferguson, T. (1973), ‘A Bayesian analysis of some nonparametric problems’, *The Annals of Statistics* **1**(2), 209–230.
- Ferguson, T. S. (1983), *Recent Advances in Statistics*, Academic Press Inc., Eds H. Rizvi and J. Rustagi, New York, chapter Bayesian Density Estimation by Mixtures of Normal Distribution, pp. 287–302.
- French, K. R., Schwert, G. W. & Stambaugh, R. F. (1987), ‘Expected stock returns and volatility’, *Journal of Financial Economics* **19**, 3–29.
- Gallant, R. A. & Tauchen, G. (1989), ‘Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications’, *Econometrica* **57**, 1091–1120.
- Ghysels, E., Plazzi, A. & Valkanov, R. (2013), The risk-return relationship and financial crises. working paper, University of North Carolina at Chapel Hill, Department of Economics.
- Ghysels, E., Santa-Clara, P. & Valkanov, R. (2005), ‘There is a risk-return tradeoff after all’, *Journal of Financial Economics* **76**(3), 509–548.
- Greenberg, E. (2013), *Introduction to Bayesian Econometrics*, 2 edn, Cambridge University Press, New York USA.
- Griffin, J. & Steel, M. (2004), ‘Semiparametric bayesian inference for stochastic frontier models’, *Journal of Econometrics* **123**(1), 121 – 152.

- Gui, H. & Whitelaw, R. F. (2006), ‘Uncovering the risk–return relation in the stock market’, *The Journal of Finance* **61**(3), 1433–1463.
- Hansen, P. R. & Lunde, A. (2006), ‘Realized variance and market microstructure noise’, *Journal of Business & Economic Statistics* **24**(2), 127–161.
- Harrison, P. & Zhang, H. H. (1999), ‘An investigation of the risk and return relation at long horizons’, *The Review of Economics and Statistics* **81**(3), 399–408.
- Harvey, C. R. (2001), ‘The specification of conditional expectations’, *Journal of Empirical Finance* **8**(5), 573 – 637.
- Ishwaran, H. & James, L. F. (2001), ‘Gibbs sampling methods for the stick breaking priors’, *Journal of the American Statistical Association* **96**(453), 161–173.
- Jensen, M. J. & Maheu, J. M. (2010), ‘Bayesian semiparametric stochastic volatility modeling’, *Journal of Econometrics* **157**(2), 306 – 316.
- Jensen, M. J. & Maheu, J. M. (2013a), ‘Bayesian semiparametric multivariate GARCH modeling’, *Journal of Econometrics* **176**(1), 3 – 17.
- Jensen, M. J. & Maheu, J. M. (2013b), ‘Estimating a semiparametric asymmetric stochastic volatility model with a dirichlet process mixture’, *forthcoming Journal of Econometrics* .
- Kalli, M., Griffin, J. & Walker, S. (2011), ‘Slice sampling mixture models’, *Statistics and Computing* **21**, 93–105.
- Kim, C.-J., Morley, J. C. & Nelson, C. R. (2004), ‘Is there a positive relationship between stock market volatility and the equity premium?’, *Journal of Money, Credit, and Banking* **36**(3), 339–360.
- Kim, C.-J., Morley, J. C. & Nelson, C. R. (2005), ‘The structural break in the equity premium’, *Journal of Business & Economic Statistics* **23**(2), 181–191.
- Lettau, M. & Ludvigson, S. (2010), *Handbook of Financial Econometrics*, Elsevier, chapter Measuring and Modeling Variation in the Risk-Return Tradeoff. Eds Yacine Ait-Shalia and Lars-Peter Hansen.
- Lo, A. Y. (1984), ‘On a class of Bayesian nonparametric estimates. I. density estimates’, *The Annals of Statistics* **12**, 351–357.

- Ludvigson, S. C. & Ng, S. (2007), ‘The empirical riskreturn relation: A factor analysis approach’, *Journal of Financial Economics* **83**(1), 171 – 222.
- Lundblad, C. (2007), ‘The risk return tradeoff in the long run: 1836-2003’, *Journal of Financial Economics* **85**(1), 123 – 150.
- Maheu, J. M. & McCurdy, T. H. (2007), ‘Components of market risk and return’, *Journal of Financial Econometrics* **5**(4), 560–590.
- Maheu, J. M. & McCurdy, T. H. (2011), ‘Do high-frequency measures of volatility improve forecasts of return distributions?’, *Journal of Econometrics* **160**(1), 69 – 76.
- Maheu, J. M., McCurdy, T. H. & Zhao, X. (2013), ‘Do jumps contribute to the dynamics of the equity premium?’, *Journal of Financial Economics* **110**(2), 457 – 477.
- McAleer, M. & Medeiros, M. C. (2008), ‘Realized volatility: A review’, *Econometric Reviews* **27**(1-3), 1045.
- Muller, P., Erkanli, A. & West, M. (1996), ‘Bayesian curve fitting using multivariate normal mixtures’, *Biometrika* **83**(1), 67–79.
- Papaspiliopoulos, O. (2008), A note on posterior sampling from dirichlet mixture models. manuscript, Department of Economics, Universitat Pompeu Fabra.
- Rodriguez, A., Dunson, D. B. & Gelfand, A. E. (2009), ‘Bayesian nonparametric functional data analysis through density estimation’, *Biometrika* **96**(1), 149–162.
- Schwert, G. W. (1990), ‘Indexes of u.s. stock prices from 1802 to 1987’, *Journal of Business* **63**(3), 399–426.
- Scruggs, J. T. (1998), ‘Resolving the puzzling intertemporal relation between the market risk premium and conditional market variance: A two-factor approach’, *Journal of Finance* **53**, 575–603.
- Sethuraman, J. (1994), ‘A constructive definition of Dirichlet priors’, *Statistica Sinica* **4**, 639–650.
- Shahbaba, B. & Neal, R. (2009), ‘Nonlinear models using dirichlet process mixtures’, *Journal of Machine Learning Research* **10**, 1829–1850.

- Taddy, M. A. & Kottas, A. (2010), ‘A Bayesian nonparametric approach to inference for quantile regression’, *Journal of Business & Economic Statistics* **28**(3), 357–369.
- Turner, C., Startz, R. & Nelson, C. (1989), ‘A markov model of heteroskedasticity, risk, and learning in the stock market’, *Journal of Financial Economics* **25**, 3–22.
- Walker, S. G. (2007), ‘Sampling the dirichlet mixture model with slices’, *Communications in Statistics – Simulation and Computation* **36**, 45–54.
- Wu, G. (2001), ‘The determinants of asymmetric volatility’, *The Review of Financial Studies* **14**(3), 837–859.

	mean	variance	skewness	kurtosis	min	max
r_t	0.0514	0.3884	-0.4047	10.0461	-4.0710	4.1630
r_t^2	0.3907	1.3474	9.7037	119.5948	0.0000	17.3300
RV_t	0.3790	0.5611	7.0305	69.4529	0.0116	11.3000
$\log(RV_t)$	-1.5602	0.8846	0.8051	4.2910	-4.4595	2.4245
$z = r_t/\sqrt{RV_t}$	0.2296	1.0789	0.0030	2.6856	-2.4080	2.8580

Table 1: Summary statistics

This table reports summary statistics for the monthly data on excess returns r_t and monthly realized volatility RV_t . Data is from 1885/1-2011/12 giving 1519 observations.

	mean	0.95 density interval
α_0	0.1922	(0.1672, 0.2171)
α_1	-0.2801	(-0.3895,-0.1748)
η_1^2	1.0177	(0.9460, 1.0962)
γ_0	-0.3319	(-0.4151,-0.2470)
γ_1	0.3766	(0.3179, 0.4329)
γ_2	0.4505	(0.3817, 0.5180)
γ_3	-0.1518	(-0.1842,-0.1170)
γ_4	0.1258	(0.0680, 0.1861)
η_2^2	0.3981	(0.3702, 0.4278)

Table 2: Parametric Model Estimates

This table reports posterior summary statistics for the parametric model:

$$\begin{aligned}
r_t &= \alpha_0 + \alpha_1 RV_t + \eta_1 \sqrt{RV_t} z_t, \quad z_t \sim NID(0, 1) \\
\log(RV_t) &= \gamma_0 + \gamma_1 \log(RV_{t-1}) + \gamma_2 \frac{1}{6} \sum_{i=1}^6 \log(RV_{t+1-i}) \\
&\quad + \gamma_3 \frac{r_{t-1}}{\sqrt{RV_{t-1}}} + \gamma_4 \left| \frac{r_{t-1}}{\sqrt{RV_{t-1}}} \right| + \eta_2 v_t, \quad v_t \sim NID(0, 1).
\end{aligned}$$

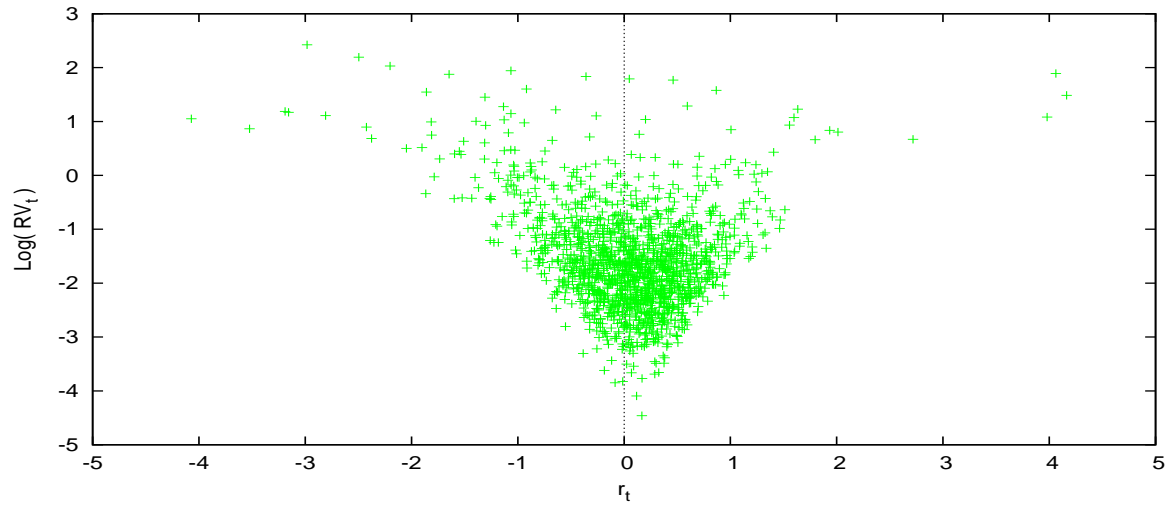


Figure 1: Excess return versus $\log(RV_t)$

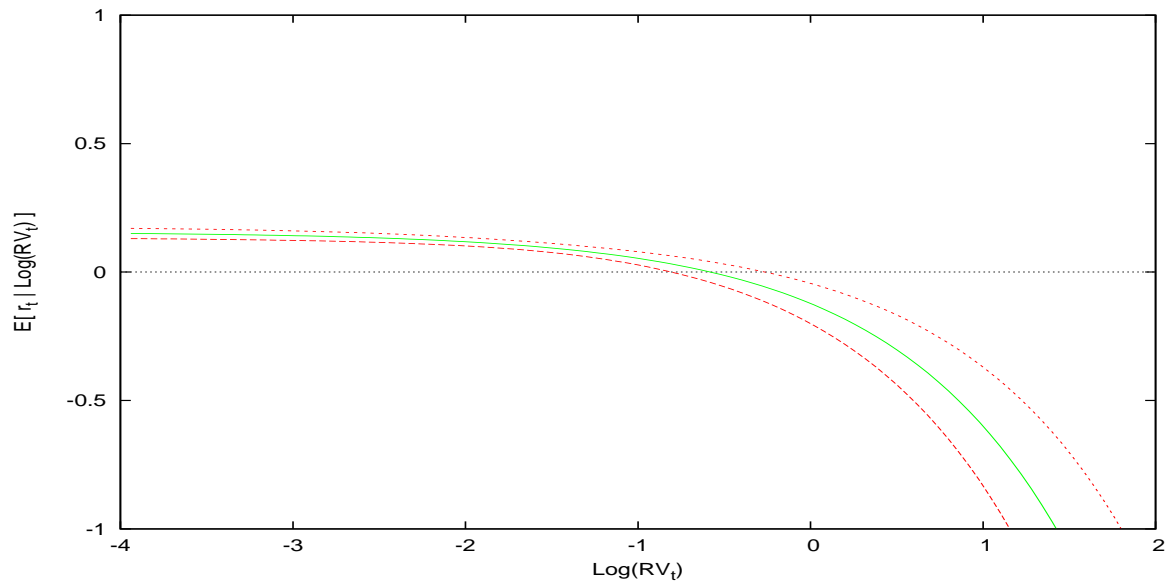


Figure 2: Expected excess return given $\log(RV_t)$ for the parametric model
This figure displays the expected excess return and 0.90 density intervals as a function of $\log(RV_t)$ for the parametric model

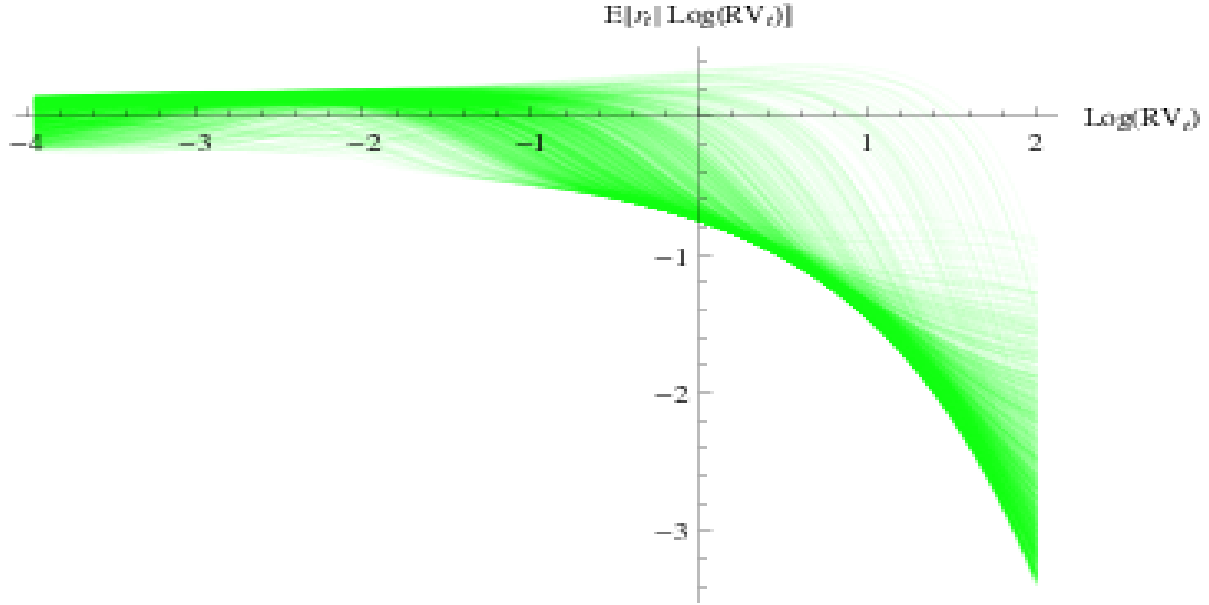


Figure 3: Expected return given $\log(RV_t)$ for each t in the sample

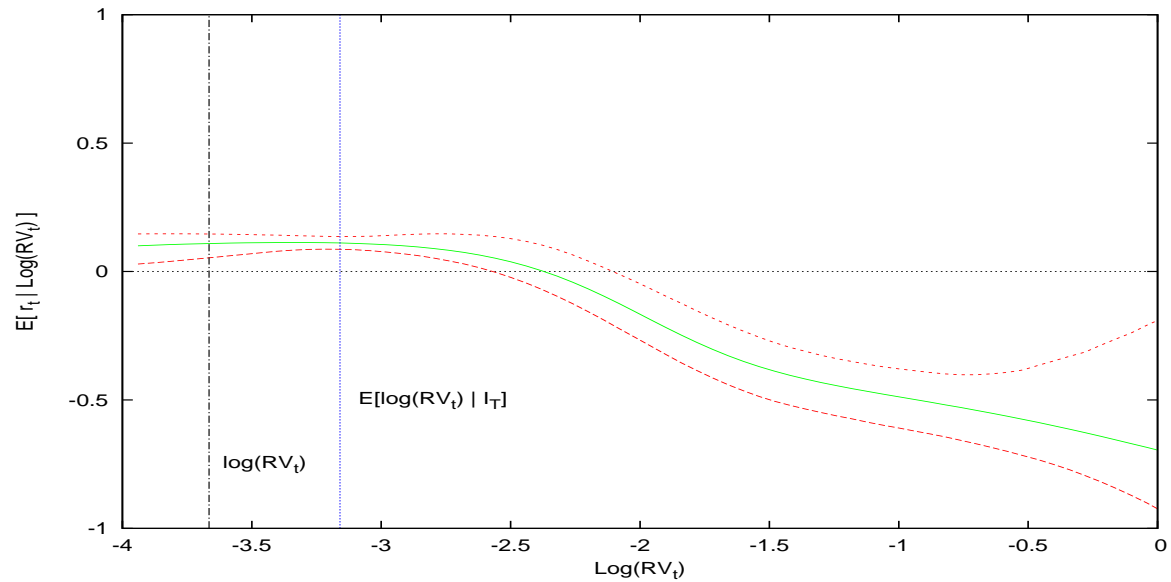


Figure 4: Expected excess return given $\log(RV_t)$ for low volatility

This figure displays the expected excess return and 0.90 density intervals as a function of $\log(RV_t)$ conditional on regressors taken $t = 1964 : 10$ which is an average volatility period. The expected log-realized volatility based on the model is blue while the actual log-realized volatility for $t = 1964 : 10$ is the black vertical line.

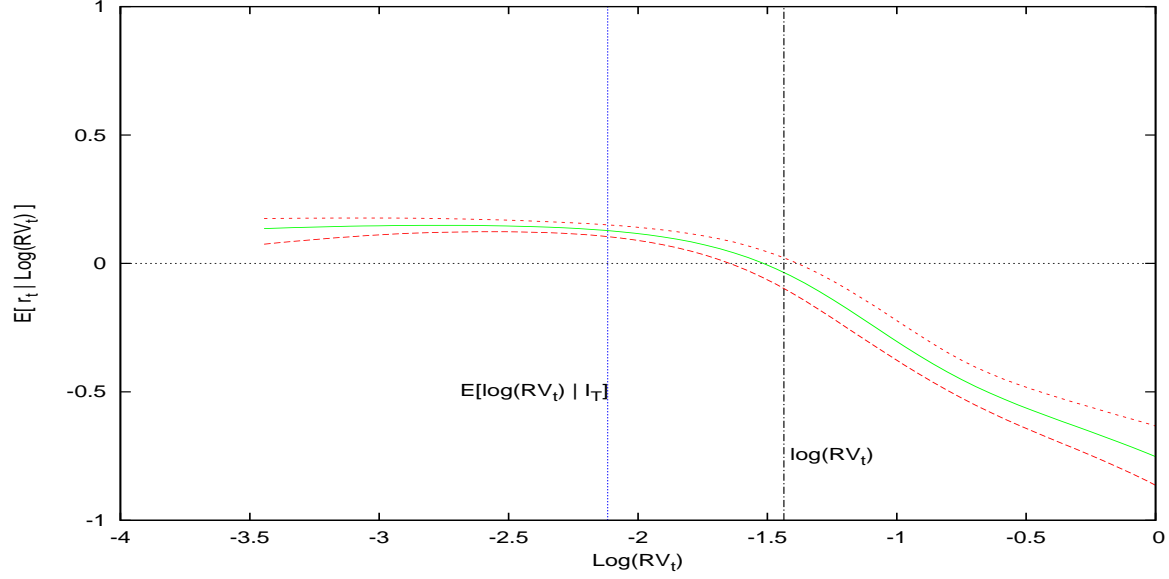


Figure 5: Expected excess return given $\log(RV_t)$ for average volatility
This figure displays the expected excess return and 0.90 density intervals as a function of $\log(RV_t)$ conditional on regressors taken $t = 1996 : 2$ which is an average volatility period. The expected log-realized volatility based on the model is blue while the actual log-realized volatility for $t = 1996 : 2$ is the black vertical line.

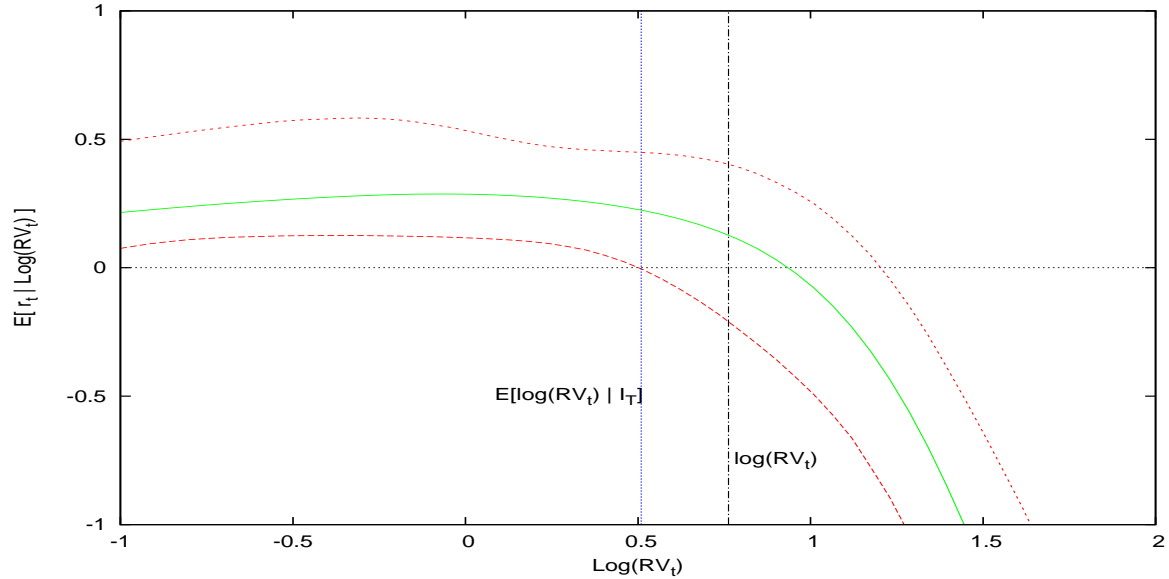


Figure 6: Expected excess return given $\log(RV_t)$ for high volatility
This figure displays the expected excess return and 0.90 density intervals as a function of $\log(RV_t)$ conditional on regressors taken $t = 2008 : 12$ which is a high volatility period. The expected log-realized volatility based on the model is blue while the actual log-realized volatility for $t = 2008 : 12$ is the black vertical line.

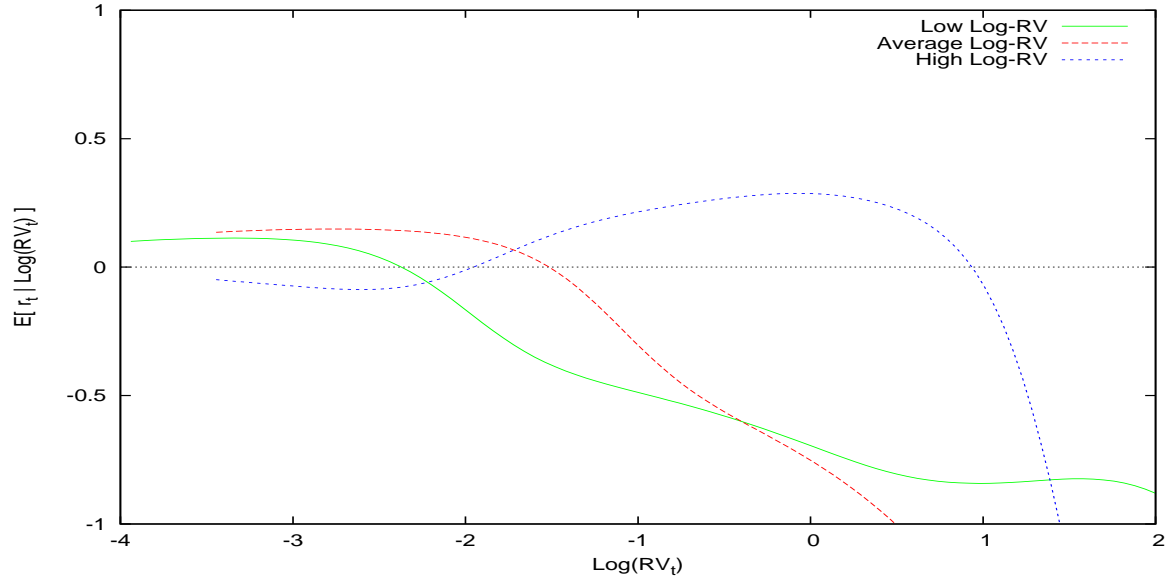


Figure 7: Expected excess return given $\log(RV_t)$ for various periods
This figure displays the expected excess return as a function of $\log(RV_t)$ conditional on regressors taken from $t = 1964 : 10$ “Low Log-RV”, $t = 1996 : 2$, “Average Log-RV” and $t = 2008 : 12$ “High Log-RV”.

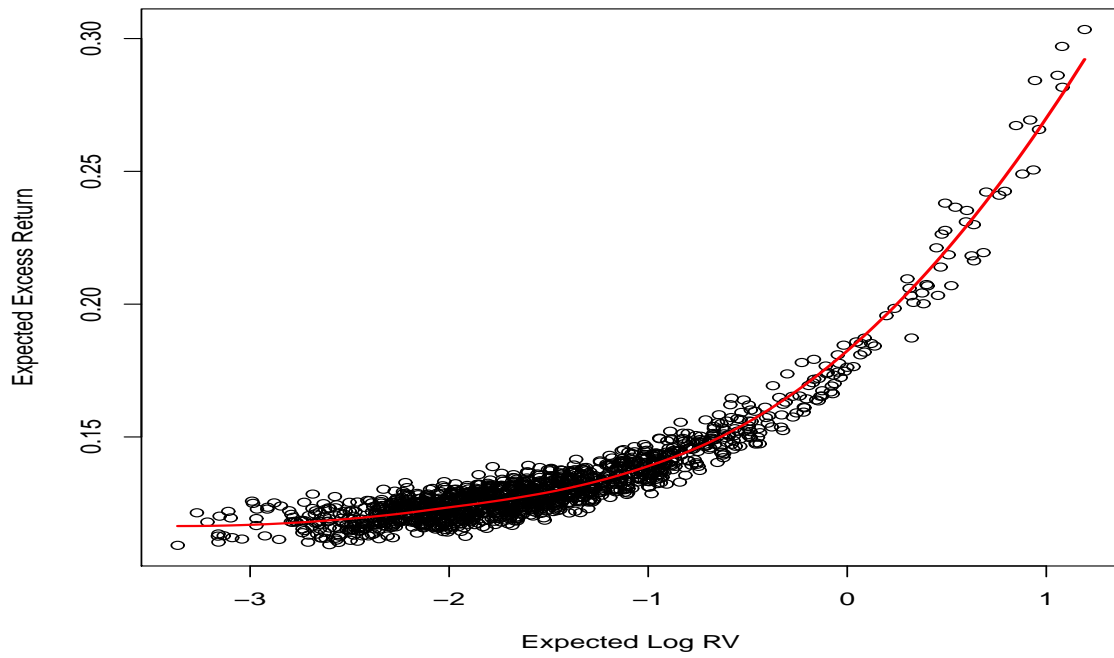


Figure 8: Expected excess return when volatility feedback is zero

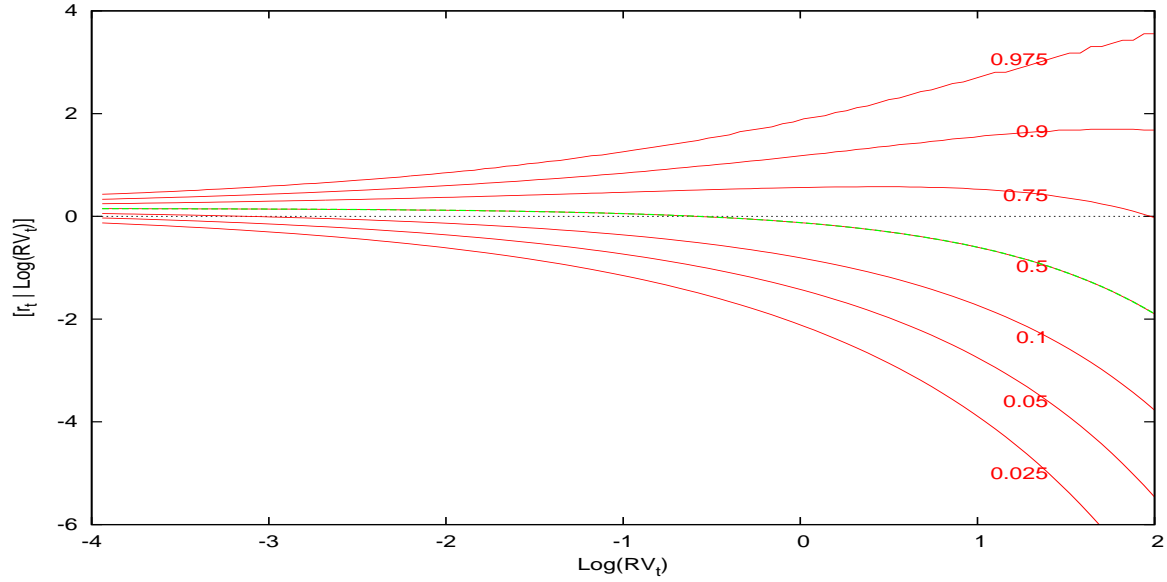


Figure 9: Quantiles of excess returns given $\log(RV_t)$ for the parametric model
This figure displays the quantiles of the distribution of excess returns conditional on $\log(RV_t)$ for the parametric model. The green dotted line is the expected excess return given $\log(RV_t)$.

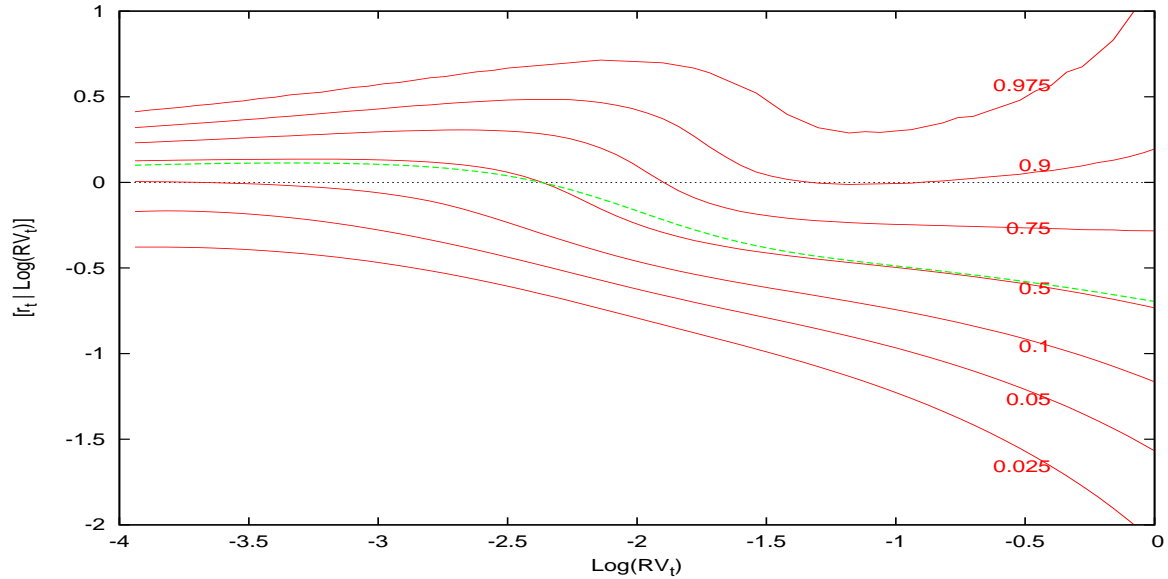


Figure 10: Quantiles of excess returns given $\log(RV_t)$ for low volatility
This figure displays the quantiles of the distribution of excess returns conditional on $\log(RV_t)$ for $t = 1964 : 10$. The green dotted line is the expected excess return given $\log(RV_t)$.

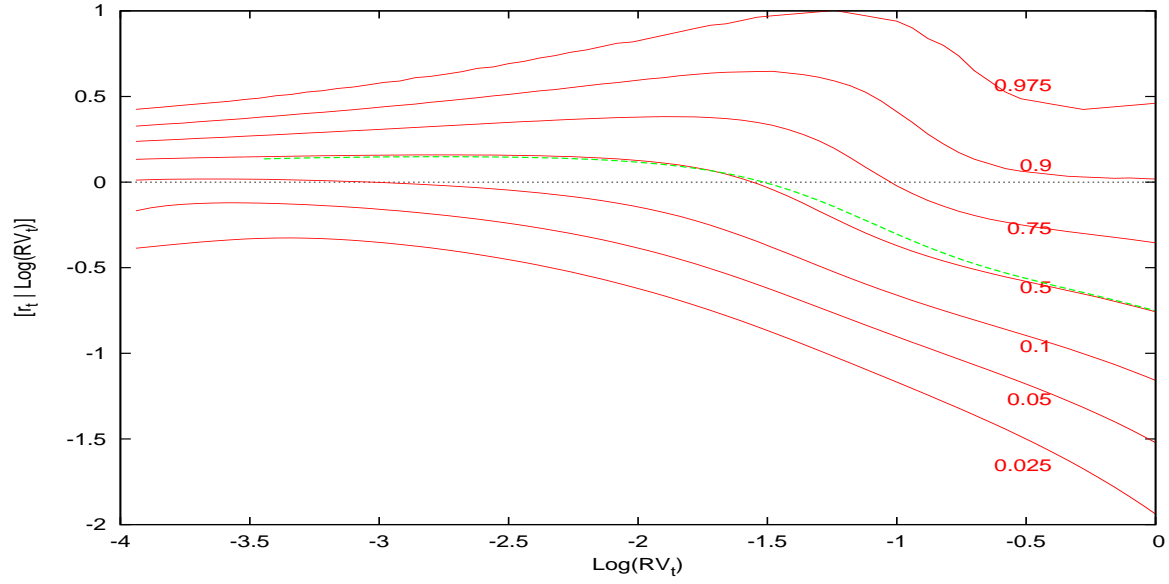


Figure 11: Quantiles of excess returns given $\log(RV_t)$ for average volatility
This figure displays the quantiles of the distribution of excess returns conditional on $\log(RV_t)$ for $t = 1996 : 2$. The green dotted line is the expected excess return given $\log(RV_t)$.

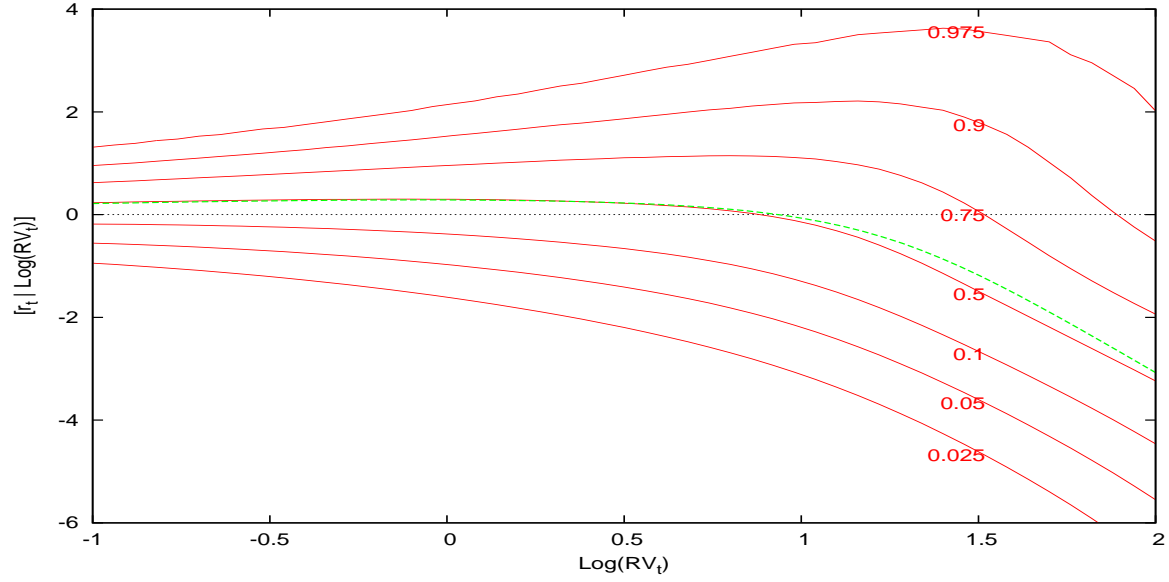


Figure 12: Quantiles of excess returns given $\log(RV_t)$ for high volatility
This figure displays the quantiles of the distribution of excess returns conditional on $\log(RV_t)$ for $t = 2008 : 12$. The green dotted line is the expected excess return given $\log(RV_t)$.

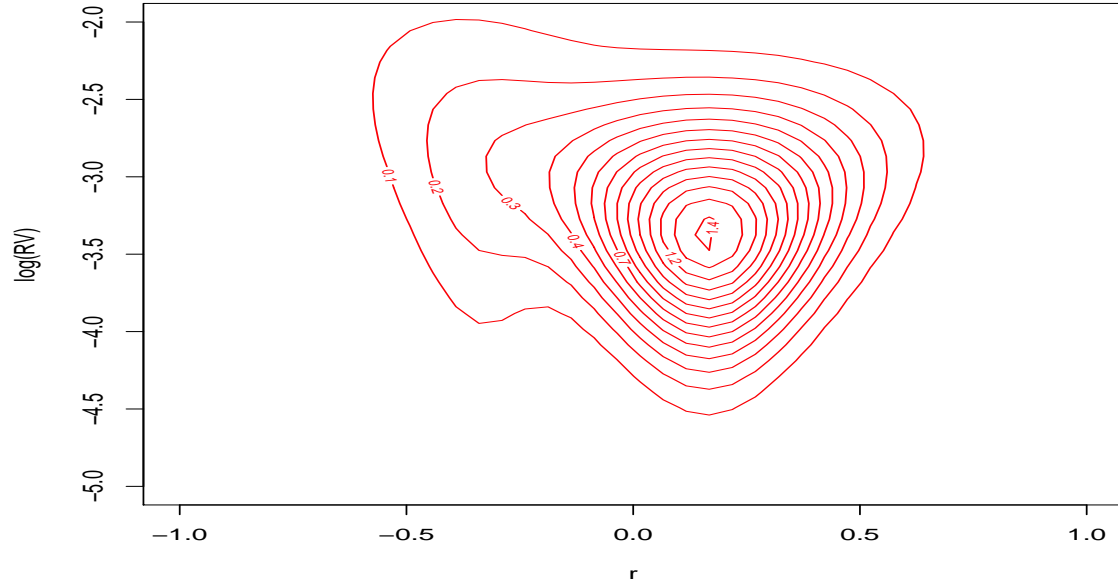


Figure 13: Predictive density for $r_t, \log(RV_t)$ for low volatility $t = 1964 : 10$

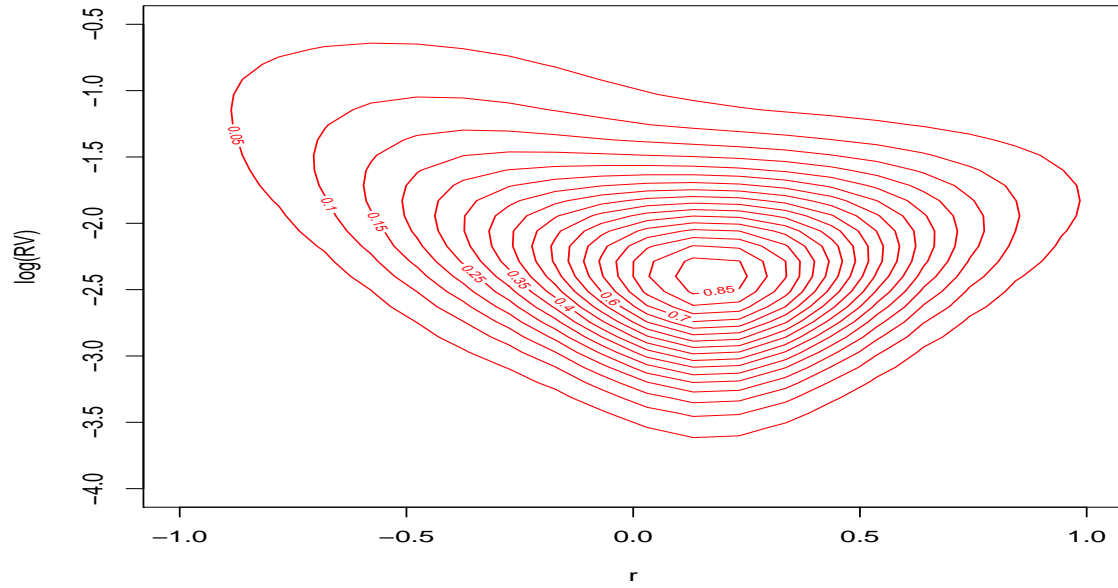


Figure 14: Predictive density for $r_t, \log(RV_t)$ for average volatility $t = 1996 : 2$

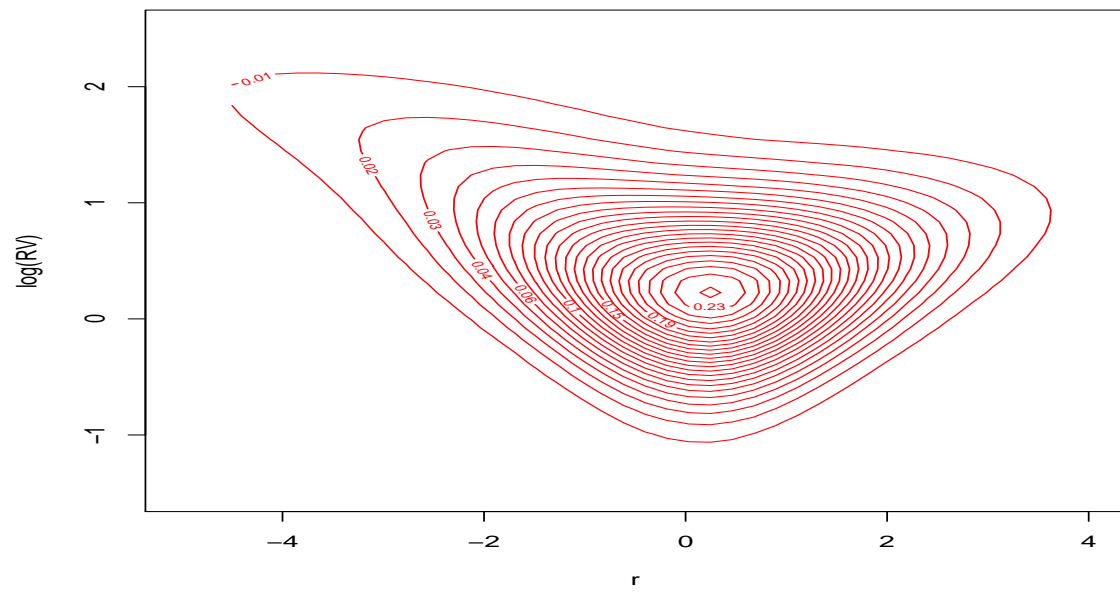


Figure 15: Predictive density for $r_t, \log(RV_t)$ for high volatility $t = 2008 : 12$

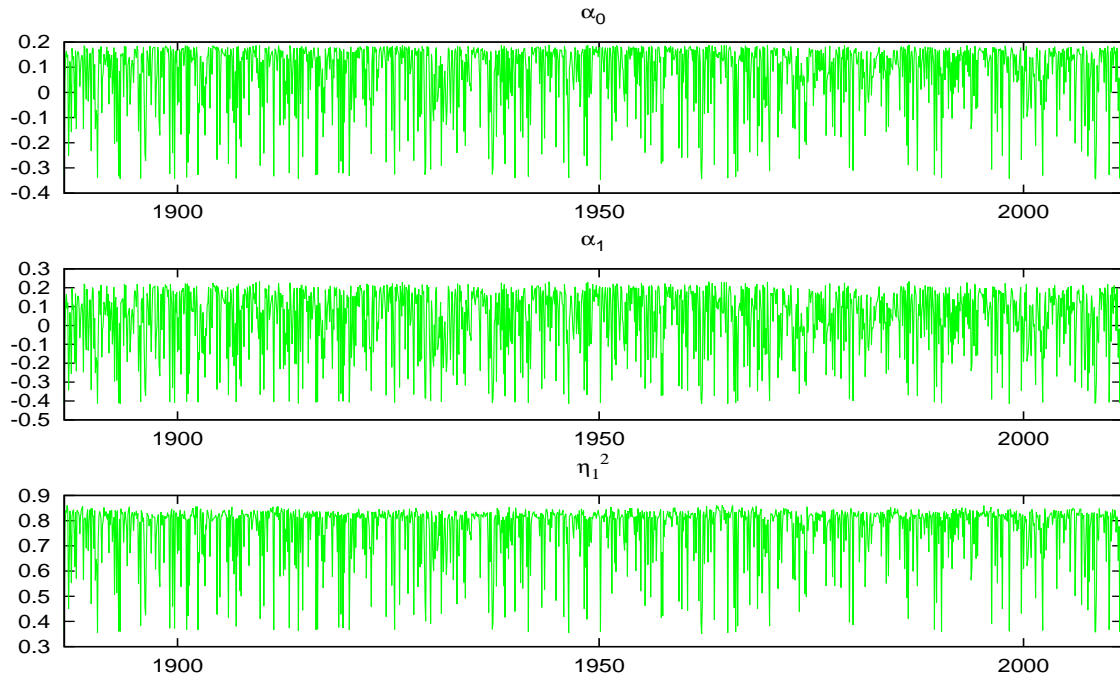


Figure 16: Posterior means of α_{0,s_t} , α_{1,s_t} and η_{1,s_t}^2 .

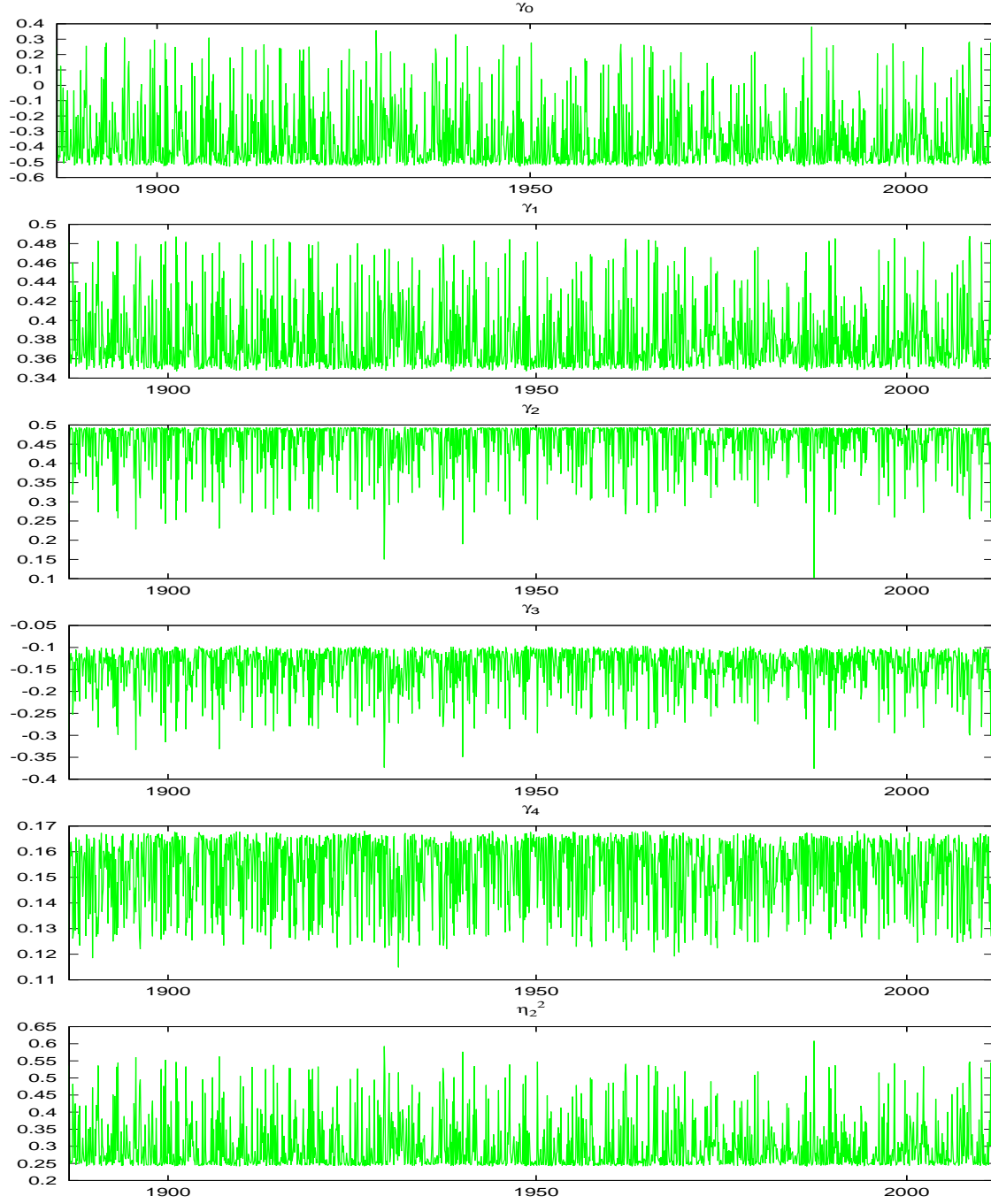


Figure 17: Posterior means of $\gamma_{0,st}, \dots, \gamma_{4,st}$ and $\eta_{2,st}^2$.