Likelihood evidence on the asset returns puzzle

Efthymios G. Tsionas

Department of Economics,
Athens University of Economics and Business,
76 Patission Street, 104 34 Athens, Greece.
Tel. (++3021) 0820 3388, Fax: (++3021) 0820 3301, Internet: tsionas@aueb.gr

Abstract

Standard equilibrium models are unable to replicate the average return on equity and the risk-free rate during 1889-1978, the well-known asset returns puzzle. The present paper, motivated by the excess of outliers in the data proposes a normal scale mixture stochastic process for output that is compatible with leptokurtosis. Using formal likelihood-based methods, it is shown that observed asset returns are compatible with posterior distributions implied by the model.

Keywords: Asset returns; equity premium puzzle; Bayesian analysis; Gibbs sampling; normal scale mixtures.

JEL Codes: C11, C15, G12.

Acknowledgments: The author wishes to thank but not implicate John Geweke for motivating the study and providing an extraordinary amount of useful comments, Gary Koop, Angelo Melino and Dale Poirier for many useful discussions, and econometrics seminar participants at Minnesota, Rochester and Toronto. Feedback from participants of the Macroeconomics Workshop of the Athens University of Economics and Business, and especially from Tryphon Kollintzas, is gratefully acknowledged. Edward Prescott kindly provided the data. Many thanks are due to the Managing Editor and three anonymous referees for their numerous constructive comments. I wish to express my sincere thanks to one anonymous referee in particular that went out of his way to provide numerous detailed comments that significantly improved every section of this paper.

1. Introduction

Over the past one hundred years, the average real stock return in the United States has been about seven per cent and the average real return on Treasury bills has been about one percent. In a seminal paper, Mehra and Prescott (1985) showed that standard equilibrium asset pricing models are unable to replicate these facts. The idea that crash risk or "peso-problems" could solve the equity premium puzzle put forward by Mehra and Prescott (1985) has traditionally been rejected by economists. One reason is that Mehra and Prescott (1988) rightly replied to Rietz (1988) that very improbable situations are needed to get a successful resolution. There are two additional reasons. The first is Kocherlakota's (1996) conclusion that only market frictions, incompleteness, or non-standard preferences could solve the puzzles and hence, that leptokurtosis could not work. This conclusion is especially important since Kocherlakota (1997) himself has used a leptokurtic model to find that the approach is still unsuccessful. The second reason has to do with the abundance of findings in the GMM literature that the Lucas (1978) model is rejected under fully non-parametric conditions -which, naturally, rules out fat-tailed explanations. Indeed, Hansen, Heaton, and Yaron (1996) seemingly reject even after elaborate correction for the finite-sample properties of GMM tests.

Therefore, the idea that asset return puzzles could be explained by leptokurtosis is not new, and the status of affairs is that confidence for such an explanation is rather limited. The thesis of this paper is that standard equilibrium asset pricing models, when coupled with a reasonable stochastic specification and formal likelihood-based methods, can go far in explaining aspects of asset returns observed during the past hundred years. The methodological approach in this paper can be described in three steps: *(i)* We use a parsimonious and tractable leptokurtic model for the endowment which is supported by the data; *(ii)* We correctly account for the estimation error and the induced effect of estimation uncertainty on asset returns; *(iii)* We use correct finite-sample distributions of finite-sample asset return moments and hence we do not have to rely on dubious asymptotic tests. Cecchetti, Lam and Mark (1993) also rely on *(i)* and *(ii)* to explain asset returns. They used the representative agent model with time-separable utility and a bivariate consumption-dividends Markov switching model which was estimated using GMM. This model is found to dominate both a random walk model and a Mehra-Prescott style model. They use a generalized calibration methodology to incorporate statistical inference. The first source of uncertainty they take into account involves uncertainty in sample asset return moments, and the second arises from the dependence on these moments on parameters that are estimated. They find that the first moments of the data can be matched for a wide range of preference parameters but the model cannot generate second order moments that are statistically close to those in the data. More specifically, "*when the endowment is forced to conform closely to the data, as it is in the Markov-switching model, and leverage is forced to imply that the dividend flow match what we actually observe, then the model cannot match the first and second moments taken together*" (p. 41). Therefore, with a fully articulated model Cecchetti et al. (1993) found that the unconditional first and second moments pose a very stringent test for the model's ability to mimic the characteristics of asset price data. However, taking account of *(iii)* we show that the results obtained are an extension of those presented by Cecchetti et al. (1993) in the sense that second moments of asset returns appear less puzzling than it was thought before.

The paper also takes a different path from the Hansen and Jagannathan (1991) approach. Although the asset return puzzle is indeed about moments, moments are sensitive to leptokurtosis in the data. The premise upon which this paper relies is that it is not possible to think of a model-specific moment in the same way as one thinks of a data-based moment. The model’s implications (be it moments of any order or anything else) cannot be given numbers since they depend on underlying parameters, as well as distributional assumptions that must be taken into account at some point. It turns out that these are parameters of *critical* importance since asset returns are sensitive on such assumptions. One may choose either to ignore this fact, or take up formal inference procedures and ask what they imply, and whether they are reasonable provided they imply something useful.

It is often argued that the Lucas model is rejected using GMM methods, so since GMM is fully non-parametric it is only natural that the Lucas model cannot explain the asset return puzzles. The problem with this argument is that all moment-based methods have to be sensitive to what assumptions are made about the tail behavior of consumption. So it is natural to suppose that there are *some* leptokurtic distributions which lead to *large* finite sample biases of GMM, *do* explain the asset return puzzles, and *are* supported by the data relative to the Gaussian alternative[1]. We will show, indeed, that a particular leptokurtic model is supported by the data, and goes a long way in explaining observed returns, *provided* we allow the model to confront the data in terms of posterior distributions. From another point of view, one way to interpret the Hansen, Heaton, and Yaron (1996) finding that the Lucas model is not rejected (even though parameter estimates are troublesome) is that elaborate correction for the finite-sample properties of GMM can go some way in reversing the trend of rejections that has been observed over the years. It is only reasonable to think that relying on correct finite-sample distributions of finite-sample asset return moments as per (*iii*) can go further in explaining the properties of asset returns. This is certainly good news for macroeconomists that admire the simplicity and elegance of the Arrow-Debreu model.

The methodology adopted in this study can be described as follows:

(i) The study proposes a normal-scale-mixture for endowment growth to model the well-known leptokurtosis of macroeconomic time series. This choice is compelling since as Geweke (1993, 1994a) shows, many U.S. macroeconomic time series strongly favor such specifications. The empirical analysis is based on a *gamma* mixing density.

(ii) Formal Bayesian inference methods are developed, and exact distributions of finite-sample properties of asset returns are derived from the posterior distributions of parameters. The implied posterior distributions of finite-sample properties of asset returns are used to investigate whether the model's predictions are statistically close to what we observe in the data.

It is perhaps worthwhile to explain the method of inference adopted in this study in some more detail. It can be summarized in five steps. In step 1, we fix the risk aversion and discount function parameters, so these parameters are not to be

---

[1] This argument is due to an anonymous referee.

estimated. In step 2, we perform Bayesian inference for the endowment model, which is an AR(1) coupled with a leptokurtic distribution for the disturbance. If we denote the parameters of the model by θ, at this stage we can obtain draws from the posterior distribution $p(\theta \mid \text{data})$. In step 3, for each θ draw, we can simulate an endowment process and compute asset returns for the simulated data set. In step 4, we can compute time-series means, standard deviations, *etc.* of equity premia and real rates. In step 5, we compare observed time-series statistics against their posterior distribution, which is the distribution of time-series statistics computed from step 4 (this distribution is induced by alternative θ draws from the posterior).

It must be emphasized from the outset what is *not* being done in this paper. First, the endowment and preference parameters are not estimated jointly. Second, the agents in the economy are not uncertain about the endowment parameters. The agents know these parameters but the econometrician does not. Third, testing is not based on the distance between finite-sample return moments from their *unconditional* values. In fact, the unconditional values are not needed at all in the context of this paper, and there is always a convincing likelihood-based argument to be made in terms of conditioning on the observed data.

The remaining of the paper is organized as follows. The economic model is presented in section 2. The econometric model is presented in section 3. Time series evidence is presented and discussed in section 4. Some concluding remarks are provided in the final section.

2. The model

Consider an economy with a representative agent endowed with constant relative risk aversion preferences. The preferences of the representative agent are $E_0 \sum_{t=0}^{\infty} \delta^t u(c_t)$ where $\delta > 0$ is the discount factor, and the spot utility function is given by:

$$u(c) = \begin{cases} \dfrac{c^{1-\alpha} - 1}{1-\alpha}, & \alpha > 0,\ \alpha \neq 1 \\ \log(c), & \alpha = 1 \end{cases} \tag{1}$$

where α is the coefficient of relative risk aversion (RRA), and c_t is consumption of date t. There is a single productive unit that produces without cost the consumption good each period. Let Y_t be the date t endowment and

$$y_{t+1} = \frac{Y_{t+1}}{Y_t}. \tag{2}$$

The stochastic process driving endowment is given by

$$\log y_{t+1} = \beta_0 + \beta_1 \log y_t + w_{t+1}^{1/2} \varepsilon_{t+1},\ t=0,1,2,\ldots, \tag{3}$$

where ε_t is *i.i.d* N(0,1), t=1,2,…, and w_t is a random variable with probability density function $\pi(\cdot)$, and support in $(0,\infty)$. The process $\{w_t\}$ is *i.i.d* and w_t is independent of ε_t (t=1,2,…). The AR(1) specification is supported by evidence in Labadie (1986) and Cecchetti *et al.* (1990). Regarding the *i.i.d* specification on $\{w_t\}$, there is evidence that asset return regressions uniformly reject an autoregressive (*i.e.* time varying volatility) specification.[2] It should be noted that since there is no production in this model we cannot separate consumption from dividends or total output.

Endowment in the present model can be measured using aggregate variables like consumption, GNP or dividends, see for example Labadie (1986). This could be avoided in a model with production, but as noted in Mehra and Prescott (1985)

[2] Results are available on request.

production does not by itself solve the equity premium puzzle. The present study uses the annual data for per capita real consumption, dividends and GNP for the time period 1889-1977 of Grossman and Shiller (1981). The specific data set was used since comparisons with previous studies are desirable (for example Mehra and Prescott, 1985, Rietz, 1988 and Labadie, 1986). Some authors have used quarterly data (Kandel and Stambaugh, 1990) but Mehra and Prescott (1985) report that their conclusions were not sensitive to the choice of data frequency.

Clearly the process in (3) is leptokurtic, which is captured by the fact that the disturbance variance is stochastic. One may ask whether or not, and to what extent leptokurtosis is a feature of the data. Jarque and Bera (1987) have provided a test for normality, defined as $JB = (n/6)\left\{S^2 + \tfrac{1}{4}(K-3)^2\right\}$ where S is the coefficient of skewness, K is the coefficient of kurtosis and n denotes the number of observations. The test is distributed as $\chi^2(2)$ under the null hypothesis of normality. For a standard normal distribution we would expect that $S = 0$ and $K = 3$ so the test considers deviations of sample skewness and kurtosis from these baseline values to judge whether the data are sufficiently close to normality. Rejection of normality could be due either to large values of sample skewness or large values of sample kurtosis coefficient. Jarque and Bera (1987) tests for normality yield *p*-values equal to 0.035 for consumption and zero for dividends and GNP (the values of the tests are 6.69, 38.75 and 55.11). Quantile-quantile plots of the data are provided in Figure 1. The plots depict the characteristic S-shape expected for leptokurtic distributions with tails fatter than the normal[3].

The information set for the representative agent at date *t*, is $\Phi_t = \{y_s, s = 1,2,...,t\}$. Notice that w_t is *not* assumed to be part of the agent's information at date t. Expectations with respect to w will be denoted by E in what follows.

The probability density function of $u_t \equiv w_t^{1/2}\varepsilon_t$ is given by

$$p(u) = (2\pi)^{-1/2}\int_0^{\infty} w^{-1/2}\exp\left(-\frac{u^2}{2w}\right)\pi(w)dw\,. \qquad (4)$$

[3] These findings are robust to inclusion of data up to 2002. Quantile-quantile plots of the data as well as Jarque-Bera tests still support the evidence in favor of leptokurtosis. We will not make use of the extended data set in the following for comparability reasons.

Growth rates in (3) exhibit time varying conditional moments that (conditional on the variance) give rise to a normal mixture for $\log y_{t+1}$ with mixing density $\pi(\cdot)$. Moreover, the unconditional distribution of u_{t+1}, exhibits leptokurtosis - excess kurtosis relative to the normal distribution - and, therefore, allows for extreme observations.

Let q_t denote the price of equity and P_t the discount price of risk-free real bonds in terms of the consumption good. The risk-free real interest rate is defined as:

$$R_t^0 = P_t^{-1} - 1, \quad \text{t=0,1,2,....} \tag{5}$$

The real return on equity is:

$$R_t^q = \frac{q_{t+1} + Y_{t+1}}{q_t} - 1, \quad \text{t=0,1,2,....} \tag{6}$$

It can be shown that existence of equilibrium requires existence of the moment generating function (MGF) of u_t and places an upper bound on the risk aversion coefficient α. For details, see part A of the Technical Appendix. Effectively, this places restrictions on what we may consider a valid process for w_t. For example, if w_t follows an inverted gamma distribution, u_t follows a Student-*t* distribution for which the MGF does not exist. Therefore, we have to choose a mixing density for w_t that is sufficiently flexible, yet leads to a density for u_t, such that its MGF exists. Such a density is provided by a *gamma* specification for the distribution of w_t.

Asset returns can be obtained using a modification of Labadie's (1986) procedure to obtain them in a series expansion. If $\rho = 1 - \alpha$, $M(\cdot)$ denotes the MGF of w_t, define the recursive coefficients

$$a_{j+1} = \beta_1(\rho + a_j),\ a_1 = \rho\beta_1, \quad \text{j=0,1,2,....} \tag{7}$$

$$A_{j+1} = \delta A_j \exp(\beta_0(\rho + a_j))M(1/2(\rho + a_j)^2),\ A_1 = \delta\exp(\rho\beta_0)M(1/2\rho^2). \tag{8}$$

Then, stock prices are given by:

$$q_t = Y_t \sum_{j=1}^{\infty} A_j y_t^{a_j} \,. \tag{9}$$

Gross real stock returns are given by the expression:

$$1 + R_{t+1}^q = y_{t+1} \frac{1 + \sum_{j=1}^{\infty} A_j y_{t+1}^{a_j}}{\sum_{j=1}^{\infty} A_j y_t^{a_j}} \tag{10}$$

and the gross real risk-free rate is:

$$1 + R_t^0 = \frac{\exp(\alpha\beta_0)\, y_t^{\alpha\beta_1}}{\delta M(\alpha^2/2)} \,. \tag{11}$$

Details and proofs are provided in part A of the Technical Appendix. It is worthwhile to mention that the paper does not rely on existence or other properties of unconditional moments. Since existence of such moments is crucial for sampling-theory treatments of the problem some results in terms of restrictions on the MGF are provided in propositions 3 and 4 in part A of the Technical Appendix.

3. Econometric model

Beginning with de Finetti (1961), Bayesian treatments have generally represented symmetric leptokurtic error distributions as scale mixtures of normal variates, see Lindley (1971), Zellner (1976) and West (1984). For more recent work see Geweke (1993) and Tsionas (1999). Here, these results are used in the context of linear models (like the growth regression in (3)) when the conditional variances are *gamma* distributed.

The model is given by

$$y_i = \mathbf{x}_i' \beta + u_i \,, \quad i = 1,2,...,n \tag{12}$$

where $\mathbf{x}_i$ is a $k \times 1$ vector of covariates, β is a $k \times 1$ parameter vector, $u_i = w_i^{1/2} \varepsilon_i$, $\{\varepsilon_i\}$ is *i.i.d.* $N(0,1)$ and $\{w_i\}$ is *i.i.d* distributed as *gamma*, with probability density

$$\pi(w_i|\lambda,\theta)=\frac{\theta^{\lambda}}{\Gamma(\lambda)}w_i^{\lambda-1}\exp(-\theta w_i),\quad w_i\ge 0,\ \lambda,\theta>0, \tag{13}$$

where λ is the shape parameter and θ is the scale parameter of the distribution. The distribution will be denoted by $G(\lambda,\theta)$. The first two moments of the distribution are

$$E(w_i)=\frac{\lambda}{\theta},\quad Var(w_i)=\frac{\lambda}{\theta^2}. \tag{14}$$

Equations (13) and (14) define the *normal-gamma mixture.* The likelihood function of the normal-gamma mixture is given by

$$L(\beta,\lambda,\theta;\mathbf{y},\mathbf{X})=(2\pi)^{-n/2}\int_0^\infty\ldots\int_0^\infty\prod_{i=1}^n w_i^{-1/2}\exp[-\frac{(y_i-\mathbf{x}_i'\beta)^2}{2w_i}]\pi(w_i\mid\lambda,\theta)dw_1\ldots dw_n. \tag{15}$$

Since $\int_0^\infty x^a\exp(-bx-cx^{-1})dx=2\left(\frac{c}{b}\right)^{\frac{a+1}{2}}K_{a+1}(2\sqrt{bc})$, $a>-1,\ b,c>0$ where $K_{a+1}(\cdot)$ denotes the modified Bessel function of the third kind, the likelihood function can be simplified to

$$L(\beta,\lambda,\theta;\mathbf{y},\mathbf{X})=2^{-n\lambda/2}\theta^{n(2\lambda+1)/4}\Gamma(\lambda)^{-n}\prod_{i=1}^n|y_i-\mathbf{x}_i'\beta|^{\lambda-1/2}K_{\lambda-1/2}(\sqrt{2\theta}\,|y_i-\mathbf{x}_i'\beta|). \tag{16}$$

It can be shown that, for $\lambda=1$, the probability density of u_i is Laplace with scale θ, defined as $f_u(u_i)=(\theta/2)\exp(-\theta|u_i|)$. For fixed λ/θ, and $\theta\to\infty$ the distribution of w_i degenerates, and u_i becomes normally distributed. Thus the model nests two important special cases.

For $\lambda\le 1/2$, the density function is unbounded at the origin, *i.e.* $\lim_{z\to 0}f_u(z)=\infty$. This implies that the likelihood must diverge as $\mathbf{y}_i-\mathbf{x}_i'\beta\to 0$ (for some $i=1,\ldots,n$) and, therefore, is multimodal.

Assuming that

$$\pi(\beta,\lambda,\theta)\propto\pi(\beta)\pi(\lambda)\pi(\theta), \tag{17}$$

and $\pi(\beta) \propto const.$, it remains to specify priors for θ and λ. For θ, a conditionally conjugate *gamma* prior $G(d,b)$, *i.e.*

$$\pi(\theta) \propto \theta^{d-1} \exp(-b\theta), \tag{18}$$

is a practical choice which, at the same time, can represent a sufficiently rich pattern of prior beliefs about this parameter. The prior mean of θ is d/b and the prior variance is d/b^2. It is also assumed that $\lambda^2 \sim G(m,s)$, *i.e.*

$$\pi(\lambda) \propto \lambda^{2m-1} \exp(-s\lambda^2),\ \lambda > 0. \tag{19}$$

This prior places less weight on large values of λ compared to a *gamma* prior and, therefore, makes a statement about the relative prior plausibility of outliers: Restricting λ probabilistically to smaller values, implies that non-normality is more plausible. Of course, this statement can be extremely vague or nearly dogmatic depending on the values of parameters *m* and *s*. The specification of *m, s, d* and *b* will be taken up in section 4.2.

Since the likelihood function in (16) is complicated and prone to numerical problems, an alternative (but equivalent) posterior must be sought. It can be shown that based on (16), the w_i's can be treated as parameters whose priors are given by $\pi(w_i \mid \lambda, \theta)$, see for example Gelfand and Smith (1990) and Tanner and Wong (1987). The joint posterior of all parameters in this case is

$$\begin{aligned}&\pi(\beta,\lambda,\theta,\mathbf{w} \mid \mathbf{y},\mathbf{X}) = \theta^{n\lambda+d-1}\Gamma(\lambda)^{-n} \det(\Omega)^{\lambda-3/2}\lambda^{2m-1} \\ &\exp[-\tfrac{1}{2}(\mathbf{y}-\mathbf{X}\beta)'\Omega^{-1}(\mathbf{y}-\mathbf{X}\beta) - \theta\{\mathrm{b} + tr(\Omega)\} - s\lambda^2],\end{aligned} \tag{20}$$

where $\Omega = diag(w_1,\ldots,w_n)$. The posterior distribution whose kernel is given by (20) is proper despite the fact that an improper prior could be specified for the parameters.

Since the kernel posterior distribution in (20) is not amenable to analytical integration to obtain marginal posterior densities and moments, the need arises to resort to computational methods for Bayesian inference. Although computational Bayesian inference is anything but straightforward, an efficient posterior integration scheme is the Gibbs sampler (Gelfand and Smith, 1990, Tierney, 1994 or Geweke, 1994b, and Tsionas, 1999) modified in the presence of augmentation by latent variables like the w_i's (see Tanner and Wong, 1987). The Gibbs sampler is an iterative scheme that produces successive draws $\{\beta^{(i)}, \lambda^{(i)}, \theta^{(i)}; i = 1,\ldots,M\}$ by utilizing the conditional distributions of a block of parameters conditionally on the block of the remaining parameters by setting up a Markov chain in the parameter space. Under mild regularity conditions (Roberts and Smith, 1994) satisfied in the present application, such draws converge in distribution to the posterior whose kernel is given by $\pi(\beta,\lambda,\theta,\mathbf{w}|\mathbf{y},\mathbf{X})$ in (20). Marginal posterior moments and densities may be computed easily using the Markov chain output. Details of the algorithm are provided in part B of the Technical Appendix.

4. Prior elicitation

Functions of the parameters must be used in order to perform prior elicitation and, also, to assess the model's capability in explaining the asset return puzzles. Growth rate volatility, for example, is crucial since as noted by Mehra and Prescott (1985) sufficiently volatile growth rates could be compatible with nearly any configuration of real equity and risk-free rates for some value of the coefficient of RRA. The functions that will be presented are used to ensure that prior information is reasonable and does not dominate the data in order to produce an artificial resolution of the asset return puzzles.

Volatility and excess kurtosis

Since $E(u_i^2|\lambda,\theta)= E(w_i|\lambda,\theta)= \lambda/\theta$ and $E(u_i^4|\lambda,\theta)= E(w_i^2|\lambda,\theta)= \frac{3\lambda(1+\lambda)}{\theta^2}$ it follows that the coefficient of excess kurtosis, defined as $K(\lambda,\theta)=\frac{E(u_i^4\,|\,\lambda,\theta)}{E(u_i^2\,|\,\lambda,\theta)^2}-3$, is given by

$$K(\lambda,\theta)=3\lambda^{-1}, \tag{21}$$

The standard deviation of u_i is given by

$$v(\lambda,\theta)\equiv(\lambda/\theta)^{1/2}. \tag{22}$$

This is the standard deviation of the error term in the growth rate regression. We follow Mehra and Prescott (1985) and insist on marginal posteriors of v that do not possess heavy tails or that with high probability the support of v contains only “reasonable” values. To that end, several diagnostics are available.

Inter-quartile volatility range

We consider the random variable

$$R_v = \frac{v_{0.75} - v_{0.25}}{v_{0.95} - v_{0.05}}, \tag{23}$$

where v_j is the j*th* quantile of the marginal posterior distribution of v, defined by $\int_0^{v_j} \pi(v|\mathbf{y},\mathbf{X}) = j/100$. Low values of R_v are associated with heavier tails. R_v is about 5/9 for a uniform, 4/10 for a normal and 1/10 for a Cauchy distribution.

Volatility quantiles

We consider moments of $v_{0.95}$ which is, of course, a function of λ and θ. To approximate the posterior expectation of $v_{0.95}$ we proceed as follows. Let $\{w_t^{(j)}, t = 1,..,1000\}$ be a random sample from $G(\lambda^{(j)}, \theta^{(j)})$ where $\lambda^{(j)}$ and $\theta^{(j)}$ denote values of λ and θ respectively in the j*th* Gibbs iteration. Let $v_{0.95}^{(j)}$ be the 95% empirical quantile associated with $\{w_t^{(j)}, t = 1,..,1000\}$. Posterior moments of $v_{0.95}$ can be computed using the sample moments of $v_{0.95}^{(j)}$ s.

Growth rate quantiles

The quantiles of the marginal distribution of u_t (denoted by ξ_m) can be used as additional diagnostics. The m*th* order marginal quantile is computed as follows. Given values of parameters in the j*th* Gibbs iteration, say $[\beta^{(j)}, \lambda^{(j)}, \theta^{(j)}, w^{(j)}]$ a set of residuals $\{u_i^{(j)}, i = 1,...,n\}$ is computed whose order statistics are $\{U_i^{(j)}, i = 1,...,n\}$. The m*th* quantile, $\xi_m^{(j)}$, for the j*th* pass is computed ($0 < m < 1$). The final estimate of ξ_m is given by the sample average of $\xi_m^{(j)}$s over Gibbs iterations. The value m =0.05 is used, implying that interest focuses on the lower 5% value of growth rates. This is a measure of severity of negative crashes in growth rates.

Nominal level of growth rate quantiles

Taking into account Mehra and Prescott's (1988) criticism to Rietz (1988), it is useful to consider the probability that model growth rates are below their 5% counterpart in the data. If r denotes the lower 5% quantile of the empirical distribution of data, the function of interest is $Q(\beta,\lambda,\theta) = \Pr(y_t(\beta,\lambda,\theta) \leq r)$ where $\{y_t(\beta,\lambda,\theta), t = 1,...,n\}$ is a simulated data set. Clearly, the average of this function of interest should not exceed 5% by a significant amount, otherwise the Mehra and Prescott (1988) criticism would apply and a possible resolution of the puzzles would be artificial.

Induced priors for growth rates

The ultimate criterion of comparing alternative prior distributions rests with their implications for growth rates. In other words, we need to consider the implied prior distribution for growth rates, which is really the Bayesian predictive distribution resulting for a given prior on the underlying parameters. To simplify the process somewhat, we assume that β_0 and β_1 are set to their historical values, and consider the implications of different priors of λ and θ for growth rates. This can be done in a straightforward way by simulation. Define $z_t = \log y_t$. Since $\pi(z_t) = \int f(z_t \mid \lambda,\theta)\pi(\lambda,\theta)\, d\lambda\, d\theta$ the following steps are required: (*i*) Draw λ and θ from their prior; (*ii*) Draw z_t from the process in (3) using the simulated values of λ and θ; (*iii*) Present a kernel density estimate of $\{z_t\}$.

All these measures are conditional on β_0 and β_1 which are set equal to their least squares estimates (available in Labadie, 1986) but they are unconditional on all other parameters.

Asset return moments

A complete specification of prior distributions in (18) and (19), involves choice of parameters m, s, d and b. To elicit these parameters we can follow a complementary approach to the one just presented above and consider the implications for the induced prior distribution of endowment's standard deviation (ν in equation (22)). Prior parameters are assigned according to the following requirements:

(i) the tails of ν die off sufficiently fast, so that relatively distant values from the mean become highly unlikely.

(ii) The priors on λ and θ are relatively diffuse.

We choose $m = 2$, $s = 10^3$, $d = 100$, and $b = 1$. We remind again that all priors have been computed by fixing the parameters β_0 and β_1 at their posterior means so the resulting priors are really "conditional priors". Plots of the induced conditional prior on the growth rate's standard deviation ν, and the induced conditional prior of growth rates themselves are provided in Figures 2, 3, and 4. From Figure 2 the conditional prior mean of ν is just above 2%. We remind that the sample standard deviation of growth rates is about 4% for the consumption data. The implied conditional prior on growth rates is excessively smooth: Growth rates turn out to be essentially between –1% and 1% so the prior will not dominate the data to produce an artificial resolution of the asset returns puzzle. That would be the case if the prior implied growth rate crashes significantly more important than their empirical counterparts. Induced conditional priors for asset returns (equity premia and riskless rates specifically) are reported in Figure 4 for $\alpha = 2$. Equity premia are well below 2%, values close to zero are highly probable, and risk-free rates are close to 2%. Equity premia around 6% are far in the tails of marginal priors: The odds of such events are about 4:10,000. To conclude this discussion there is nothing in the prior to "drive" the data to produce large premia and low real rates.

[4] Previous versions of the paper have utilized calibration based on the posterior means of parameters.

5. Time series evidence

Having established that the priors are reasonable in a sense that was defined precisely in the previous section, we will examine next evidence from the posterior distribution. Before proceeding to do so we find it necessary to elaborate on a concern of fundamental importance, namely whether there are *reasonable priors for which the historical averages of asset returns are probable given their joint posterior implied by the model.* Clearly, some aspects of the posteriors will be sensitive to prior information adopted while other aspects may be quite robust. For example, the actual joint posteriors might be sensitive but high probability content may still be assigned to historical averages. This is what really concerns one in the present context. *Provided* the reader agrees that the priors adopted here are reasonable (and several measures were offered to convince the reader that this is so) and *provided* the priors when combined with the data, yield posteriors of asset returns which give high probability to historical averages, then we must have succeeded. There is no doubt there could be values of the hyperparameters, which yield priors that do not resolve the puzzle. This cannot be precluded on *a priori* grounds but in that case we would need to refine considerably the mechanism for comparison of prior distributions. Given the computational complexity of the model this task does not appear feasible.

For the prior specification, Table 2 reports statistics of the marginal posterior distributions of functions of interest when the endowment is measured using consumption, dividends and GNP respectively. The functions of interest were computed using the same simulation methodology that we used in section 4 to compute the functions of interest resulting from the prior. In this instance, however, we have the posterior draws for the parameters, so the functions of interest are computed against the posterior. The main posterior results for asset return moments are presented in Table 3 for selected values when the coefficient of RRA equals $\alpha = 2$. The conclusions are as follows:

---

[5] An argument can be made that the post-war data is perhaps a better benchmark because it is smoother and closer to normality. The results do depend on whether we consider the post war data or the data set as a whole. Kocherlakota (1997, p. 565) rightly argues that "*judging from the popularity of doomsday investment books, investors today certainly seem to believe that Depression-like falls in consumption*

1. The posterior expectation of ν is 0.0479 when endowment is measured using consumption. The corresponding data figure is 0.036.[6] Since the marginal posterior of ν is fairly symmetric, this means that values as low as 0.025 are possible given the posterior standard deviation of this parameter. The posterior variation of ν for dividends or GNP is also consistent with what is observed in the data. A well known stylized fact emerges from the posterior means of ν across data sets: Consumption is less volatile than GNP which, in turn, is less volatile than dividends.

2. The 95% quantiles of the posterior distribution of ν provide further information regarding whether ν can assume large values with non-negligible probability. For the three different data sets, these quantities are respectively 0.10, 0.187 and 0.222. They are, typically, associated with small standard errors. Posterior expectations of the range statistic R_ν behave the same way, and indicate that, typically, ν does not possess heavy tails.

3. Further information is provided by the posterior mean of $\xi_{0.05}$, the lower 5% quantile of model disturbances. These quantiles are, typically, associated with high indices of relative numerical efficiency (Geweke, 1992). The $\xi_{0.05}$ statistics are -0.027 for consumption, -0.078 for GNP, and -0.20 for dividends. Compared with the empirical quantiles reported in Table 3, they are not unreasonable at all, and indicate that *the marginal posteriors do not impose extreme tail behavior in growth rates.* Additional interest focuses on posterior expectations of $Q(\beta,\lambda,\theta)$, *i.e.* the nominal level of $\xi_{0.05}$. For consumption, the posterior mean of $Q(\beta,\lambda,\theta)$ is 0.031, with standard deviation 0.02. For GNP and dividends the posterior expectation is 0.054 and 0.017 respectively.

All in all, there is enough evidence to conclude that our posterior analysis is not subject to the Mehra and Prescott (1985) criticism.

4. The indices of excess kurtosis provided by posterior means of K (see (21)) indicate that deviations from normality are important but their standard errors are, typically, large.

---

are possible. If this is true, data on consumption growth that is exclusively from the 'good times' since World War II delivers an inadequate picture of the true IMRS of the representative investor".

[6] Labadie (1986) reports that measurement error in the consumption series could account for a standard deviation about 0.055.

One issue is that although leptokurtosis leads to posterior distributions which make sense in the light of the data, why should one believe in leptokurtosis in the first place? This question can be answered formally based on the Bayes factor of the normal-gamma against the normal model. This is done as follows. Let $\pi_1(\phi_1 \mid Y) \propto \pi_1(\phi_1)L_1(\phi_1;Y)$ and $\pi_2(\phi_2 \mid Y) \propto \pi_2(\phi_2)L_2(\phi_2;Y)$ be two kernel posterior distributions, conditional on the same data Y, with parameters ϕ_1 and ϕ_2 respectively. Here, $\pi_i(\phi_i)$ denotes the prior, and $L_i(\phi_i;Y)$ denotes the likelihood function of the ith model. The marginal likelihood for the ith model is defined as $m_i(Y) = \int \pi_i(\phi_i)L(\phi_i;Y)d\phi_i$ provided the integrals converge. A usual condition to satisfy integrability, is to assume proper priors for the parameters that are not common across models. The marginal likelihood is simply the integrating constant of the kernel posterior. The Bayes factor in favor of model 1 and against model 2 is simply

$$B_{12} = \frac{m_1(Y)}{m_2(Y)},$$

so the issue is how to compute the marginal likelihoods. Here, we follow the technique suggested by Chib (1995), who proposed sub-sampling within the Gibbs sampling scheme. The results of this procedure are provided in Table 5, where reported are Bayes factors in favor of the normal-gamma, and against the normal model. Bayes factors favor the normal-gamma mixture model to varying degrees depending on the prior as well as the data set. However, in all cases the Bayes factors are in excess of unity and between 1.5 and roughly 3. These Bayes' factors are in the intermediate range between "barely worth mentioning" and "positive support" in the terminology of Jeffreys (1961)[7]. This implies that the data do not clearly reject leptokurtic alternatives in favor of the normal distribution for the disturbances so the basic model agrees with some important aspects of the data, as these are summarized by the Bayes factor.[8]

[7] According to Jeffreys (1961) if a Bayes factor is less than 1 we have "negative support" for the hypothesis, if it is between 1 and 3 it is "barely worth mentioning", if it is between 3 and 12 we have "positive" evidence, and if it is between 12 and 150 we have "strong" support.

[8] If we use only the post-war data, the Bayes factors are generally lower: 1.1 for the consumption data, 1.34 for GNP and 2.12 for dividends. This is reasonable because the serious outliers in the data are in the pre-war data set. However, the leptokurtic model still receives support from the data.

6. Posterior evidence on asset return moments

Given the posterior distribution and the evidence that were drawn from it in the previous section, our next objective is to investigate the implications of the posterior for asset return moments. The purpose of this investigation is to assess whether the model's implications for asset return moments are close to the data in a statistical sense.

6.1. Methodology

This section focuses on whether the model matches the observed properties of asset returns when the discount factor is less than unity and the coefficient of RRA is below 10 as suggested by Mehra and Prescott (1985). Likelihood inference for even the simple asset pricing model is highly complicated, which is why calibration is a dominant technique in the literature. For a Bayesian perspective on general equilibrium models, see Geweke (1999). However, deriving posterior distributions of asset returns is important in judging the model’s ability to reproduce key statistics. What is needed is to incorporate all available information in the posterior distributions of parameters to perform asset return inference. Parameters α and δ will remain fixed, because the asset returns puzzle as approached by Mehra and Prescott (1985) is certainly a conditional (on these parameters) concept.

Duffie and Singleton (1994) formalize the moment matching criterion implicit in Mehra and Prescott’s (1985) approach by using a simulated moments estimator. While extremely useful, this technique does not produce distributions of asset returns in finite samples. The method adopted here (see, for example Meng, 1994) is the following: Given a sample $\{\beta^{(i)},\lambda^{(i)},\theta^{(i)};i=1,...,M\}$ from the posterior distribution, asset returns $R^q(\beta,\lambda,\theta;\alpha,\delta)$ and $R^0(\beta,\lambda,\theta;\alpha,\delta)$ in (10) and (11), are functions of interest whose evaluation is possible for each draw. We only keep draws which satisfy the conditions for finite stock prices. Specifically, given $(\beta^{(i)},\lambda^{(i)},\theta^{(i)}\}$ a set of S time series $\{y_t^{(s)},t=1,...,n\}$ is generated ($s=1,...,S$), asset returns $R^q_{(s),t}(\beta^{(i)},\lambda^{(i)},\theta^{(i)};\alpha,\delta)$ and $R^0_{(s),t}(\beta^{(i)},\lambda^{(i)},\theta^{(i)};\alpha,\delta)$ are computed for each date t, and each simulation s, and their time series averages:

$$R^{q}_{(s)}(\beta^{(i)},\lambda^{(i)},\theta^{(i)};\alpha,\delta)=n^{-1}\sum_{t=1}^{n}R^{q}_{(s),t}(\beta^{(i)},\lambda^{(i)},\theta^{(i)};\alpha,\delta)$$

$$R^{0}_{(s)}(\beta^{(i)},\lambda^{(i)},\theta^{(i)};\alpha,\delta)=n^{-1}\sum_{t=1}^{n}R^{0}_{(s),t}(\beta^{(i)},\lambda^{(i)},\theta^{(i)};\alpha,\delta)$$

are saved for each $i=1,..,M$. Standard kernel density procedures can then be used to estimate the posterior distribution of average asset returns. In this case, $n=88$, and $S=500$. Time series standard deviations, or other moments, may be computed similarly. To compute stock prices, q_t, the infinite sum in (9) needs to be truncated. Let $q_t^{(T)}$ denote the series $q_t^{(T)}=Y_t\sum_{j=1}^{T}A_j y_t^{a_j}$, i.e. stock prices computed with the infinite sum truncated at T iterations. We determine T so that $\left|q_t^{(T+1)}-q_t^{(T)}\right|\leq\varepsilon$ for all $t=1,..,n$, and $\varepsilon=10^{-7}$.This criterion is quite stringent since we are interested in absolute, not relative convergence of the series $q_t^{(T)}$.

One issue is whether keeping only draws which satisfy the existence conditions alters drastically the posterior distributions of asset returns. If we ignore the existence conditions the posterior distribution makes very little sense and should be at odds with reality since, in reality, stock prices are finite. On the other hand, by imposing these restrictions we find empirically that (depending on the particular data set) only 10-15% of the available draws are, eventually, rejected so the existence conditions do not place a stringent restriction on the posterior distribution. This is, of course, reassuring because throwing out a large number of draws would only mean that we are trying to impose a straight-jacket on the data and, therefore, the results would not be believable.

Clearly, asset return moments are of critical importance in the investigation of the equity premium and risk-free rate puzzles. Simulations are conducted for given values of the coefficient of RRA α, and the discount factor δ. The discount factor is set to 0.98 resulting in an annual rate of 2.04%. The moment generating function needed in asset return simulations, is given by

$$M(t)=\left(\frac{\theta}{\theta-t}\right)^{\lambda},\text{ for } t<\theta . \qquad (24)$$

6.2. Results

Given the simulation methodology presented above, results can be organized based on *(i)* values of the coefficient of RRA, α, and discount factor, δ, and *(ii)* the data set employed. The results are presented graphically in Figure 5, which provides aspects of the joint distribution of equity premia and risk-free rates for $\alpha = 2$. Asset return moments are reported in Table 3. Clearly, the bivariate posteriors do include the historical averages of 6% and 1%, and *these averages are clearly part of high probability content contours of the joint posterior*. In other words, the historical averages are not far in the tails of the joint posterior. The implication is, of course, that the model is compatible with a 6% premium and risk-free rates close to 1%. One may even have risk-free rates close to zero but these situations are not very probable in the consumption-related posteriors, although they receive non-negligible probability in the GNP and dividends data.

Additional evidence on the adequacy of the model is provided by the behavior of the Sharpe ratio, defined as equity premium over the standard deviation of stock returns. For each parameter draw, a model simulation provides not only an average equity premium but its volatility as well, so the Sharpe ratio can be computed and is a convenient way to summarize the model's implications about two moments simultaneously. The best way to present this evidence is to report the posterior distribution of Sharpe's ratio, as in Figure 6. In this figure, for given values of the risk aversion parameter, a kernel estimate of the posterior is reported. The historical value of the Sharpe ratio is about 0.0.7/0.16=0.44. From the posteriors in Figure 6, this value does not appear unreasonable. One may attach a "*p*-value" to this number, which in light of the posterior turns out to be quite high. Indeed, the *p*-value (computed numerically from the available draws) for the consumption data when the risk aversion parameter is 2 is approximately 0.58. This conveys the message that the historical value is not excessively high, and in this sense we do not face an equity premium or risk-free rate puzzle (conditional, of course, on the model). Naturally, we cannot *precisely* match Sharpe's ratio in the calibration sense but this is immaterial provided we can argue that a model produces a "*p*-value" large enough to claim that the historical value is not large conditional on the model.[9]

An additional concern is whether draws from the posterior, which yield high equity volatility, are compatible with other moments. This is important because high equity volatility might be incompatible with a low and unvolatile risk-free rate and, therefore, it would not be enough to consider exclusively the Sharpe's ratios. The necessity to take account of the joint distribution of all relevant moments has been emphasized in the literature by Cecchetti, Lam and Mark (1993). One summary statistic that can be used in this context is posterior moments for average premium, average real rate and volatility of the real rate conditional on stock return volatility that exceeds the sample value. The results reported in Table 4 indicate that average equity premia are large, and real rates are low and relatively unvolatile. There are two points worth mentioning here. First, a finite-sample volatility realization of the sample value (16% approximately) is not too unlikely according to the evidence presented in Table 3. At least, it is less than two posterior standard deviations away from the posterior mean. But then the evidence presented in Table 4, shows that conditional on that realization, the *other* moments become less puzzling that they were unconditionally[10]: The equity premium goes up, and the real rate and its volatility go down. This is a remarkable feature of the joint posterior that eventually helps us to reconcile the model with the data, in a statistical sense.

It is important that based on the *consumption* data we find support for a leptokurtic model, and subsequently we find that historical average premia and real rates are supported by joint asset return posteriors. From Figure 5, the GNP and dividends data yield the same conclusions - as expected since these data are more volatile. The engine behind these results is, of course, formal inference techniques but more fundamentally the adoption of a leptokurtic model for output that increases the amount of aggregate risk in the economy beyond the risk adopted by Mehra and Prescott (1985). In combination with risk aversion, leptokurtosis creates a mechanism

---

[9] The argument and the presentation of Sharpe's ratio posterior distributions were suggested by an anonymous referee.

[10] I am grateful to an anonymous referee for suggesting this argument.

in equilibrium that has to compensate investors in order to hold stocks instead of bonds, so it yields higher equity premia and lower risk-free rates compared to models without leptokurtosis, as in Mehra and Prescott (1985).

It should be noted that the model taken up in this study has no transition dynamics at all so it is expected that the model will do a poor job along some dimensions of the data, for example the predictability of asset returns. This is expected because the Mehra-Prescott model is a fairly simple description of reality, and it is not designed to match patterns in excess returns. In fact, Mehra and Prescott claim it cannot match *basic* facts like first moments of the data. The focus of the paper has been on arguing that extending the basic model to allow for leptokurtic disturbances provides a statistical framework that shows that the model is not necessarily deficient in matching asset returns and Sharpe ratios. Of course, the basic model can be extended along the lines of Campbell and Cochrane (1999) to allow for matching the world along other dimensions, but this is a second order issue. The first order issue is to ensure that there is no asset returns puzzle, in a formal statistical sense.[12]

One important question is how the present approach relates to the Mankiw and Zeldes (1991) approach. Mankiw and Zeldes argue that accounting for the consumption of non-stockholders lowers the risk aversion parameter required for a resolution of the equity premium puzzle. This is a valuable contribution. The approach taken up in this paper asks a different question. For a given risk aversion parameter and assuming the model is useful, can we argue that historical values of asset returns are (collectively) plausible? By adopting a statistical framework consistent with Bayesian analysis, it is shown that reasonable priors make the historical values appear plausible from the point of view of the model's asset return posterior distributions. In that sense, the view adopted here is that nothing is *basically* wrong with the model provided we do not ask the model to match exactly first or second moments of the data, but allow an examination of the entire posterior distribution of asset returns. Of course, this does not mean the model is perfect. The model *will* fail along some

---

[11] If we use only the post-war data, the Bayes factors are generally lower: 1.1 for the consumption data, 1.34 for GNP and 2.12 for dividends. This is reasonable because the serious outliers in the data are in the pre-war data set. However, the leptokurtic model still receives support from the data.

[12] The model is also expected to do a poor job along some other dimensions, including but not limited to variance of dividend yields, term premia on real bonds etc. An anonymous referee rightly argued that these term premia are close to zero or even negative so this would pose a serious challenge to this model. This is correct but it is a second-order issue as I argue in the main text.

dimensions of the data, like all models do. The role of theory is to fix the model along these specified dimensions but still the model implications must be sought in terms of return *distributions* conditional on the data.

Conclusions

The purpose of the study was to couple the equilibrium asset pricing model with a stochastic specification that allows for leptokurtosis in aggregate growth, and examine the implications of this approach for the celebrated equity premium puzzle. The methodology is based on formal Bayesian analysis. Prior distributions were selected so that endowment growth is sufficiently smooth and the priors do not dominate the data. For reasonable values of the discount factor and the coefficient of relative risk aversion, posterior distributions of average asset returns support the historical averages of equity premia and real risk-free rates. Moreover, data statistics are compatible with the model-based joint distribution of first and second moments of asset returns.

One important result is that the bivariate posteriors of equity premia and real risk-free rates include the historical averages of 6% and 1%, and *these averages are clearly part of high probability content contours of the joint posterior*. In other words, the historical averages are not far in the tails of the joint posterior. The implication is, of course, that the model is compatible with a 6% premium and risk-free rates close to 1%. One may even have risk-free rates close to zero.

Another important result concerns the behavior of the Sharpe ratio which summarizes two moments simultaneously, namely equity premium and real rate. The historical value of the Sharpe ratio is about 0.44. From the posteriors we computed, the *p*-value (for the consumption data when the risk aversion parameter is 2) is approximately 0.58. Therefore, the historical value is not excessively high, and in this sense we do not face an equity premium or risk-free rate puzzle.

A third result concerns a feature of the joint posterior distribution of asset returns. We focus on whether draws from the posterior, which do yield high equity volatility, are compatible with other moments. One statistic that can be used to address the issue is posterior moments for average premium, average real rate and volatility of the real rate conditional on stock return volatility that exceeds the sample value. The results indicate that a finite-sample volatility realization of the sample value (16% approximately) is not too unlikely and conditional on that realization, the *other* moments become less puzzling that they were unconditionally: The equity premium is higher, and the real rate and its volatility are lower. This is helpful in reconciling the model with the data, in a formal statistical sense.

Moreover, we find support for a normal-gamma error specification in a growth rate autoregression for the consumption, dividends and GNP data. This support is measured using the Bayes factor. We conclude that formal posterior analysis does not provide evidence for an equity premium or risk-free rate puzzle for the equilibrium model when coupled with an econometric model that allows for leptokurtic disturbances.

Table 1. Useful data statistics

Series	*Mean*	*Standard deviation*
Growth rates		
Consumption	0.0183	0.0357
Dividends	0.0061	0.124
GNP	0.0249	0.0791
Real risk-free rate	0.008	0.0567
Real equity return	0.0698	0.165
5% quantiles		
Consumption	-0.046	
Dividends	-0.306	
GNP	-0.088	

Table 2. Posterior statistics for parameters and functions of interest

	Consumption	*GNP*	*Dividends*
β_0	0.022 (0.015)	0.016 (0.018)	0.009 (0.016)
β_1	-0.169 (0.418)	0.121 (0.274)	0.0003 (0.133)
λ	0.042 (0.015)	0.123 (0.069)	2.720 (1.312)
θ	18.12 (3.68)	20.02 (3.871)	122.11 (53.11)
v	0.0479 (0.010)	0.0842 (0.024)	0.138 (0.028)
$v_{0.95}$	0.10 (0.030)	0.187 (0.055)	0.222 (0.040)
$\xi_{0.05}$	-0.027 (0.070)	-0.078 (0.123)	-0.200 (0.081)
K	84.59 (50.48)	35.22 (53.22)	1.612 (2.332)
Q	0.031 (0.020)	0.054 (0.027)	0.017 (0.016)
R_v	0.061 (0.050)	0.260 (0.123)	0.396 (0.022)
Relative numerical efficiency			
β_0	0.77	0.77	0.77
β_1	0.66	0.95	0.95
λ	0.92	1.12	0.92
θ	0.33	0.23	0.13
v	0.44	0.57	0.09
$v_{0.95}$	0.58	0.66	0.05
$\xi_{0.05}$	1.01	0.97	0.31
K	0.93	0.96	0.65
R_v	0.88	0.98	0.75

Notes: Figures in parentheses are posterior standard deviations. Geweke's convergence diagnostic has been computed for each of the parameters reported above. The maximum absolute value of this diagnostic (distributed as standard normal asymptotically in the number of draws) was 1.430 and corresponds to θ, v, and $v_{0.95}$ for the consumption, GNP and dividends data respectively. The model for endowment growth rate is $\log y_{t+1} = \beta_0 + \beta_1 \log y_t + w_{t+1}^{1/2}\varepsilon_{t+1}$ where w_t is distributed iid Gamma with parameters λ and θ. The functions of interest are defined as $v(\lambda,\theta)\equiv(\lambda/\theta)^{1/2}$ -standard deviation of error term, $K(\lambda,\theta) = 3\lambda^{-1}$ -excess kurtosis of error term, $R_v = \frac{v_{0.75} - v_{0.25}}{v_{0.95} - v_{0.05}}$ -the interquartile range of volatility v, v_q denotes the qth quantile of v, and ξ_m denotes the mth quantile of the marginal distribution of error term. These quantities are computed using simulation procedures as described in section 4.

Table 3. Posterior moments of asset returns from the model ($\alpha = 2$)

	Consumption	GNP	Dividends
Equity premia			
Mean	0.039	0.064	0.066
S.D	0.029	0.039	0.033
Real risk-free rate			
Mean	0.031	0.022	0.010
S.D	0.023	0.032	0.032
Volatility of equity premia			
Mean	0.039	0.041	0.035
S.D	0.069	0.085	0.091
Volatility of real risk-free rate			
Mean	0.032	0.034	0.029
S.D	0.020	0.016	0.015

Notes: Given values of the coefficient of relative risk aversion $\alpha = 2$, discount factor $\delta = 0.98$, and posterior draws for the parameters of the model asset returns are functions of interest whose definitions are provided in (10) and (11). Specifically, given the posterior draws $\Theta^{(i)}$ a set of S time series $\{y_t^{(s)}, t = 1,...,n\}$ is generated ($s = 1,...,S$), asset returns $R_{(s),t}^{q}(\Theta^{(i)};\alpha,\delta)$ and $R_{(s),t}^{0}(\Theta^{(i)};\alpha,\delta)$ are computed for each date t, and each simulation s, and their time series averages, $R_{(s)}^{q}(\Theta^{(i)};\alpha,\delta)$ and $R_{(s)}^{0}(\Theta^{(i)};\alpha,\delta)$ are saved for each $i = 1,..,M$. In this case, $n = 88$, and $S = 500$. Time series standard deviations, or other moments, may be computed similarly. To compute stock prices, q_t, the infinite sum in equation (9) needs to be truncated. Let $q_t^{(T)}$ denote the series $q_t^{(T)} = Y_t \sum_{j=1}^{T} A_j y_t^{a_j}$, i.e. stock prices computed with the infinite sum truncated at T iterations. We determine T so that $\left| q_t^{(T+1)} - q_t^{(T)} \right| \le \varepsilon$ for all $t = 1,..,n$, and $\varepsilon = 10^{-7}$. Finally, only draws which satisfy the conditions for finite stock prices are retained.

Table 4.
Asset return moments conditional on stock volatility exceeding sample value ($\alpha = 2$)

	Consumption	GNP	Dividends
Average equity premium	0.069	0.093	0.128
Average real rate	0.019	0.009	0.0001
Volatility of real rate	0.014	0.016	0.015

Notes: Computations reported in this table follow the same methodology described in the notes following Table 3 with one important difference: We only consider the posterior draws for which the stock return volatility exceeds its sample value. Stock returns in this model are functions of the parameters whose definition is provided in equation (10). Stock volatility can be computed for each one of 500 generated endowment time series using standard techniques.

Table 5. Bayes factors in favor of normal-gamma and against normal model

Data set	*Bayes factor*
Consumption	1.515
GNP	2.121
Dividends	2.970

Notes: This table reports the Bayes factor in favor of the normal-gamma model and against a simple normal linear regression model for the endowment growth rate. The Bayes factor summarizes the data and prior evidence in favor of the normal-gamma model and was computed using sub-sampling withing the Gibbs sampler, see Chib (1995).

TECHNICAL APPENDIX

PART A. ASSET RETURNS AND EXISTENCE OF EQUILIBRIUM

In this part, we prove the recursive formulae for asset returns, and examine in some detail conditions for existence of an equilibrium with finite stock prices.

Proposition 1. Let $\rho = 1-\alpha$, $a_1 = \rho\beta_1$, $a_{j+1} = \beta_1(\rho + a_j)$, $A_1 = \exp(\rho\beta_0)M(\frac{1}{2}\rho^2)$, $A_{j+1} = A_j \exp[\beta_0(\rho + a_j)]M\left(\frac{1}{2}(\rho + a_j)^2\right)$, for all $j = 1,2,...$, where $M(t)$ denotes the moment generating function of ω_t (all $t = 1,..,n$), i.e. $M(t) = \int_0^\infty \exp(t\omega)\pi(\omega)d\omega$, and $\pi()$ is the probability density of ω_t as in (4). The solution for stock prices is given by $q_t = Y_t \sum_{j=1}^{\infty} A_j y_t^{a_j}$ where y_t is defined in (3), and Y_t denotes the endowment (see the discussion following (1)). Returns are then defined as in (5) and (6).

Proof.

Let s_t denote the date t measure of equity held by the representative agent and carried over from the end of period t-1, b_t the date t measure of Treasury bills carried over from the end of period t-1, q_t the price of equity and P_t the discount price of risk-free real bonds in terms of the consumption good. There is one unit of equity outstanding, and zero units of bonds. The household's optimization problem is:

$$\text{max:} \quad E_0 \sum_{t=0}^{\infty} \delta^t u(c_t)$$

by choice of $\{c_t\}_{t=0}^{\infty}, \{s_{t+1}\}_{t=0}^{\infty}, \{b_{t+1}\}_{t=0}^{\infty}$, and subject to:

$$c_t + q_t(s_{t+1} - s_t) + P_t b_{t+1} \leq s_t Y_t + b_t$$

$$c_t, b_{t+1} \geq 0 .$$

The first order conditions to the household's optimization problem give the following condition

$$h_t = \delta E_t (h_{t+1} + Y_{t+1}^{\rho})$$

where $h_t \equiv Y_t^{-\alpha} q_t$, t=0,1,2,….

The unique solution of this martingale difference equation is given by

$$h_t = E_t \sum_{j=1}^{\infty} \delta^j Y_{t+j}^{\rho} .$$

We start by solving this equation. This requires computing conditional expectations of the form

$$E_t Y_{t+j}^{\rho} \text{ for j=1,2,}\ldots$$

We start at j=1. Since $Y_{t+1} = y_{t+1} Y_t$ it follows $Y_{t+1}^{\rho} = Y_t^{\rho} \exp(\rho\beta_0) y_t^{\rho\beta_1} \exp(\rho u_{t+1})$,

so if we take expectations conditional at date t, we get

$$E_t Y_{t+1}^{\rho} = Y_t^{\rho} \exp(\rho\beta_0) y_t^{\rho\beta_1} M(1/2\rho^2), \qquad \text{(A.1)}$$

where M() is the moment generating function of ω. If we define

$a_1 = \rho\beta_1$ and $A_1 = \exp(\rho\beta_0) M(1/2\rho^2)$, then equation (A.1) can be written as

$$E_t Y_{t+1}^{\rho} = Y_t^{\rho} A_1 y_t^{a_1}. \qquad \text{(A.2)}$$

We now do the j=2 case to get some intuition about the nature of the solution.

Since

$$Y_{t+2} = Y_{t+1} y_{t+2} = Y_t y_{t+1} y_{t+2} =$$
$$Y_t \exp(\beta_0) y_t^{\beta_1} \exp(u_{t+1}) \exp(\beta_0)[\exp(\beta_0 y_t^{\beta_1} \exp(u_{t+1})]^{\beta_1} \exp(u_{t+2}) = (\ldots) =$$
$$Y_t \exp(2\beta_0 + \beta_0\beta_1) y_t^{(\beta_1+\beta_1^2})\exp[(1+\beta_1)u_{t+1} + \rho u_{t+2}],$$

it follows easily that

$$E_t Y_{t+2}^{\rho} = Y_t^{\rho} \exp[\rho(2\beta_0 + \beta_0\beta_1)] y_t^{\rho\beta_1(1+\beta_1)} M(1/2\rho^2 + 1/2\rho^2(1+\beta_1)^2). \qquad \text{(A.3)}$$

Let $a_2 = \beta_1(\rho + a_1) \equiv \beta_1\rho + \beta_1\rho^2$. Notice that MGF's satisfy

$$M(c_1 + c_2) = M(c_1) M(c_2), \forall c_1, c_2$$

and rewrite (3) in the following form:

$$E_t Y_{t+2}^{\rho} = Y_t^{\rho} \exp(\beta_0(a_1 + \rho)) \; \mathrm{y}_t^{\rho\beta_1(1+\beta_1)} \; \exp(\rho\beta_0) M(1/2\rho^2) \quad M(1/2\rho^2(1+\beta_1)^2) =$$
$$Y_t^{\rho} A_1 \exp(\beta_0(\rho + a_1)) M(1/2(\rho + a_1)^2) \quad y_t^{\rho\beta_1(1+\beta_1)} \qquad \text{(A.4)}$$

Compared to (3) notice that one of the $\exp(\rho\beta_0)$ s is merged with $M(1/2\rho^2)$ to form A_1 while the other is merged with the "new" MGF expression $M(1/2(\rho + a_1)^2)$ to advance, as we will see, from A_1 to A_2.

Regarding the last MGF term in (4), notice that $(\rho + a_1)^2 = \rho^2(1+\beta_1)^2$. It is therefore clear that we should set

$$a_2 = \rho\beta_1(1+\beta_1) \equiv \beta_1(\rho + a_1),$$

which is the exponent associated with y_t in (4). It is clear that we should set:

$$A_2 = A_1 \exp(\beta_0(\rho + a_1))M(1/2(\rho + a_1)^2).$$

So it seems reasonable to guess

$$a_{j+1} = \beta_1(\rho + a_j),\ a_1 = \rho\beta_1 \quad \text{(A.5a)}$$

and

$$A_{j+1} = A_j \exp(\beta_0(\rho + a_j))M(1/2(\rho + a_j)^2),\ A_1 = \exp(\rho\beta_0)M(1/2\rho^2). \quad \text{(A.5b)}$$

One important point is that the martingale includes a term δ^j which I didn't carry along to ease notation. The presence of this term requires modification of the A_j formulae as follows:

$$A_1 = \delta\exp(\rho\beta_0)M(1/2\rho^2) \text{ and } A_{j+1} = \delta A_j \exp(\beta_0(\rho + a_j))M(1/2(\rho + a_j)^2) \quad \text{(A.5c)}$$

I will not enforce this in what follows but final expressions have to include δ. I justify this by noticing that the presence of δ^j in the martingale requires getting a recursion for A_j $= r(A_{j-1})$ and then setting A's as in (A.5b) to allow for δ^j in the martingale. Alternatively we could carry δ along whenever $E_t Y_{t+j}$ is computed.

We now do the case for general j. The objective is to compute $E_t Y_{t+k}^{\rho}$ for $k = 1,2,...$ Since

$$Y_{t+k} = Y_t y_{t+1} y_{t+2} ... y_{t+k} = Y_t \prod_{j=1}^{k} y_{t+j}$$

and

$$\log y_{t+j} = \beta_0(1 + \beta_1 + \beta_1^2 + ... + \beta_1^{j-1}) + \beta_1^j \log y_t + \sum_{s=1}^{j} \beta_1^{j-s} u_{t+s},$$

which implies

$$y_{t+j} = \exp[\beta_0(1 + \beta_1 + \beta_1^2 + ... + \beta_1^{j-1})] y_t^{\beta_1^j} \exp[\sum_{s=1}^{j} \beta_1^{j-s} u_{t+s}]$$

it follows that:

$$Y_{t+k}^{\rho} = Y_t^{\rho} \prod_{j=1}^{k} \{\exp[\rho\beta_0(1 + \beta_1 + \beta_1^2 + ... + \beta_1^{j-1})] y_t^{\rho\beta_1^j} \exp[\rho\sum_{s=1}^{j} \beta_1^{j-s} u_{t+s}]\}. \quad \text{(A.6)}$$

Consider first the term Q= $\exp[\sum_{s=1}^{j} \beta_1^{j-s} u_{t+s}]$. Written out in detail this term is

$$\begin{aligned} Q = {} & \exp(u_{t+1})\ \exp(\beta_1 u_{t+1} + u_{t+2})\ \exp(\beta_1^2 u_{t+1} + \beta_1 u_{t+2} + u_{t+3}) \\ & \exp(\beta_1^3 u_{t+1} + \beta_1^2 u_{t+2} + \beta_1 u_{t+3} + u_{t+4}) + (......) + \\ & \exp(\beta_1^{j-1} u_{t+1} + \beta_1^{j-2} u_{t+2} + ... + \beta_1 u_{t+j-1} + u_{t+j}) \end{aligned}$$

If we collect similar terms we arrive at the following:[13]

---

[13] The following expression proves useful if *u*s are not independent. Dependence of *u*s is not well founded in the data but nevertheless one may want to experiment with certain dependence structures.

$$Q = \exp[u_{t+1}(1+\beta_1+\beta_1^2+\ldots+\beta_1^{j-1})$$
$$\exp[u_{t+2}(1+\beta_1+\beta_1^2+\ldots+\beta_1^{j-2}]$$
$$\exp[u_{t+3}(1+\beta_1+\ldots+\beta_1^{j-3}]$$
$$(.......)$$
$$\exp[u_{t+j}].$$

Therefore (for $j = k$) we get

$$E_t \exp[\rho\sum_{s=1}^{j}\beta_1^{j-s}u_{t+s}] =$$
$$E_t\{\exp[\rho u_{t+1}(1+\beta_1+\beta_1^2+\ldots+\beta_1^{j-1})$$
$$\exp[\rho u_{t+2}(1+\beta_1+\beta_1^2+\ldots+\beta_1^{j-2}]$$
$$\exp[\rho u_{t+3}(1+\beta_1+\ldots+\beta_1^{j-3}]$$
$$(.......)$$
$$\exp[\rho u_{t+j}]\}$$

=

$$\begin{aligned}&M[1/2\rho^2(1+\beta_1+\beta_1^2+\ldots+\beta_1^{j-1})^2]\\&M[1/2\rho^2(1+\beta_1+\beta_1^2+\ldots+\beta_1^{j-2})^2]\\&(..............)\\&M(1/2\rho^2).\end{aligned} \qquad (A.7)$$

Notice that

$$a_j = \rho\beta_1+\rho\beta_1^2+\ldots+\rho\beta_1^j, \qquad (A.8a)$$

which implies

$$(a_j+\rho)^2 = \rho^2(1+\beta_1+\beta_1^2+\ldots+\beta_1^j)^2. \qquad (A.8b)$$

But then the previous expression involving the MGF's can be written as:

$$M[1/2(a_j+\rho)^2]\ \ M[1/2(a_{j-1}+\rho)^2]\ (\ldots)\ M(1/2\rho^2),\ j=k. \qquad (A.9)$$

Now consider the exponents of y_t in (A.6). Terms that involve y_t can be collected together as

$$W = \prod_{j=1}^{k} y_t^{\beta_1^j} \equiv y_t^{\rho\beta_1+\rho\beta_1^2+..+\rho\beta_1^k}. \qquad (A.10)$$

But the exponent of y_t in the above expression is just a_k because of (A.8a). Therefore, we can write

$$E_t Y_{t+k}^{\rho} = A_k Y_t^{\rho} y_t^{a_k} \qquad (A.11)$$

where it is apparent that $a_{j+1} = \beta_1(\rho + a_j)$ as in equation (A.5a). The recursion for A_j in (A.5b) follows if we notice the multiplicative way they enter into (A.9). To justify the term $\exp(\beta_0(a_j + \rho))$ that appears in A_{j+1} notice that the constant term in (A.6) is

$$C = \prod_{j=1}^{k} \exp\{\rho\beta_0(1+\beta_1+..+\beta_1^{j-1})\} = \exp(\rho\beta_0)\exp(\rho\beta_0[1+\beta_1])...\exp(\rho\beta_0[1+\beta_1+...+\beta_1^{k-1}]).$$

But $\beta_0(a_j + \rho) = \rho\beta_0(1+\beta_1+...+\beta_1^j)$ and therefore

$$C = \exp(\rho\beta_0)\exp(\beta_0(a_1+\rho)).....\exp(\beta_0(a_{k-1}+\rho))$$

If we merge the different exponential terms with respective MGF terms in (A.9) we see that the guess in (A.5) was correct.

Next we consider existence of an equilibrium with finite stock prices.

Proposition 2. Provided $\delta \exp[\beta_0\rho/(1-\beta_1)]M(\rho^2/(1-\beta_1)^2) < 1$, there exists an equilibrium with finite stock prices.

Proof. It suffices to have $q_t < \infty$, all $t = 1,2,....$ First notice that $a_j \to a^*$ provided $|\beta_1|<1$. Also notice that from (A.5c) A_j is summable provided the following holds:

$$\delta \exp[\beta_0(a^* + \rho)]M(1/2(a^* + \rho)^2)<1.$$

Since $a^* + \rho \equiv \rho/(1-\beta_1)$ we obtain the stated condition.

For $\beta_0 = \beta_1 = 0$, equilibrium existence requires $\delta M(1/2\rho^2)<1$. This places restrictions on discount factors, CRRAs as well as distributional parameters. Of importance is here the role of the MGF as well.

In the next proposition, we provide results about unconditional moments of asset returns.

Proposition 3. If the moment generating function exists at $\max\left\{\frac{\alpha^2}{2}, \left(\frac{\beta_1 m}{1-\beta_1}\right)^2\right\}$ then the unconditional m th order moment of the risk-free rate is finite, i.e. $E\left(R_t^0\right)^m < \infty$.

Proof.

The risk-free rate satisfies

$$\left(1+R_t^0\right)^m = \frac{\exp(\alpha\beta_0 m)}{\left\{\delta M(\alpha^2/2)\right\}^m} y_t^{\alpha\beta_1 m}.$$

Since

$$y_{t+1} = \exp(\beta_0) y_t^{\beta_1} \exp(u_{t+1})$$

the unconditional process is

$$y_t = \exp\left(\frac{\beta_0}{1-\beta_1}\right)\exp\left(\frac{u_{t+1}}{1-\beta_1}\right)$$

and, therefore $y_t^{\alpha\beta_1 m} = \exp\left(\frac{\alpha\beta_1 m}{1-\beta_1}(\beta_0 + u_{t+1})\right)$. Taking expectations we obtain

$$E\left(y_t^{\alpha\beta_1 m}\right) = \exp\left(\frac{\alpha\beta_0\beta_1 m}{1-\beta_1}\right) E\exp\left\{\tfrac{1}{2} m^2 \left(\frac{\alpha\beta_1}{1-\beta_1}\right)^2 w\right\}$$

which implies

$$E\left(y_t^{\alpha\beta_1 m}\right) = \exp\left(\frac{\alpha\beta_0\beta_1 m}{1-\beta_1}\right) M\left(\frac{1}{2}\left\{\frac{\alpha\beta_1 m}{1-\beta_1}\right\}^2\right).$$

By the definition of the risk-free rate we see that the moment generating function must exist at $\alpha^2/2$. The last expression implies that it must exist at $\frac{1}{2}\left\{\frac{\alpha\beta_1 m}{1-\beta_1}\right\}^2$, so it suffices to exist at the maximum of the two.

If we assume $m = 1$, so that we are interested in the first unconditional moment of the risk-free rate and $\beta_1 = 0.8$, $\alpha = 4.5$ under gamma distributed variances w_t with shape parameter λ and scale θ the MGF is $M(t) = \frac{1}{(1 - t/\theta)^\lambda}$, we need existence of the MGF at 10.125 and 162 implying $\theta > 162$. Clearly, risk-free rates cannot have moments of every order.

In the next proposition, we consider existence of the m*th* order unconditional expectation of stock returns.

Proposition 4. Suppose the moment generating function of w_t exists at $t^* = \frac{m^2}{2(1-\beta_1)^2} \max_j \left\{(1-a_j)^2,\ a_j^2\right\}$. Then the m th order unconditional expectation of stock returns is finite, i.e. $E\left(R_t^q\right)^m < \infty$.

Proof.

We have $z_t \equiv 1 + R_t^q = y_t \dfrac{1 + \sum_{j=1}^{\infty} A_j y_{t+1}^{a_j}}{\sum_{j=1}^{\infty} A_j y_t^{a_j}}$.

Let $k \in \{1,2,..\}$. Then $\sum_{j=1}^{\infty} A_j y_t^{a_j} > A_k y_t^{a_k}$, and we have the bound

$$z_t < A_k^{-1} y_t^{1-a_k} \left[1 + \sum_{j=1}^{\infty} A_j y_{t+1}^{a_j} \right].$$

It is clear that for any $m \in \{0,1,2,..\}$ we have

$$z_t^m < \left\{ A_k^{-1} y_t^{1-a_k} \left[1 + \sum_{j=1}^{\infty} A_j y_{t+1}^{a_j} \right] \right\}^m .$$

The unconditional process is $y_t = C \exp\left(\dfrac{u_t}{1-\beta_1} \right)$ where $C = \exp\left(\dfrac{\beta_0}{1-\beta_1} \right)$. Therefore, we have

$$z_t^m < \left[A_k^{-1} \left\{ C \exp\left(\frac{u_t}{1-\beta_1} \right) \right\}^{1-a_k} \left[1 + \sum_{j=1}^{\infty} A_j \left\{ C \exp\left(\frac{u_{t+1}}{1-\beta_1} \right) \right\}^{a_j} \right] \right]^m .$$

Let $A_k^{-1} \left\{ C \exp\left(\dfrac{u_t}{1-\beta_1} \right) \right\}^{1-a_k} = \chi_t$ and $\chi_t \sum_{j=1}^{\infty} A_j \left\{ C \exp\left(\dfrac{u_{t+1}}{1-\beta_1} \right) \right\}^{a_j} = \psi_t$

so that $z_t^m < (\chi_t + \psi_t)^m$. If we use a binomial expansion we have

$$z_t^m < \sum_{h=0}^{m} \binom{m}{h} \chi_t^{m-h} \psi_t^h .$$

Because u_t and u_{t+1} are independent, by taking expectations we have

$$E\left(\chi_t^{m-h}\right) \propto M\left(\tfrac{1}{2} \left(\frac{(1-a_k)(m-h)}{1-\beta_1} \right)^2 \right)$$

and

$$E\left(\psi_t^h\right) \propto E\left(\chi_t^h\right) \sum_{j=1}^{\infty} A_j M\left(\left(\frac{a_j h}{1-\beta_1} \right)^2 \right) \propto M\left(\tfrac{1}{2} \left(\frac{(1-a_k)h}{1-\beta_1} \right)^2 \right) \sum_{j=1}^{\infty} A_j M\left(\left(\frac{a_j h}{1-\beta_1} \right)^2 \right).$$

Therefore,

$$E\left(z_t^m\right) < K\sum_{h=0}^{m}\left\langle \binom{m}{h} M\left(\tfrac{1}{2}\left\{\frac{(1-a_k)(m-h)}{1-\beta_1}\right\}^2\right) M\left(\tfrac{1}{2}\left\{\frac{(1-a_k)h}{1-\beta_1}\right\}^2\right) \sum_{j=1}^{\infty} A_j M\left(\tfrac{1}{2}\left\{\frac{a_j h}{1-\beta_1}\right\}^2\right)\right\rangle$$

where K is a finite constant. Provided the moment generating function exists at $t^* = \frac{1}{2(1-\beta_1)^2}\max_{k,h,j}\left\{(1-a_k)^2(m-h)^2,\ (1-a_k)^2 h^2,\ (ha_j)^2\right\}$ stock returns will have finite unconditional expectations. It is clear that the maximum of $(1-a_k)^2(m-h)^2$ with respect to h is attained at $h=0$, and the maximum of $(ha_j)^2$ is attained at $h=m$, and the maximum of $(1-a_k)^2 h^2$ is attained at $h=m$, for any given m. Since k was arbitrary, we have $t^* = \frac{m^2}{2(1-\beta_1)^2}\max_j\left\{(1-a_j)^2,\ a_j^2\right\}$. It remains to show that $\sum_{j=1}^{\infty} A_j M\left(\tfrac{1}{2}\left\{\frac{a_j h}{1-\beta_1}\right\}^2\right)$ converges. For any given $h \in \{0,1,2,..,m\}$, we have $A_{j+1} < A_j$ for all $j = 0,1,2,..$ which implies $\lim_{j\to\infty} A_j = 0$ and we also have $\lim_{j\to\infty} a_j = a^* = \frac{\rho\beta_1}{1-\beta_1}$. By continuity of the moment generating function at a point where it exists, we have $\lim_{j\to\infty} M\left(\tfrac{1}{2}\left\{\frac{a_j h}{1-\beta_1}\right\}^2\right) = M\left(\tfrac{1}{2}\left\{\frac{a^* h}{1-\beta_1}\right\}^2\right)$ and, therefore we have that the ratio of terms $j+1$ and j in the infinite sum is

$$\frac{A_{j+1} M\left(\tfrac{1}{2}\left\{\frac{a_{j+1} h}{1-\beta_1}\right\}^2\right)}{A_j M\left(\tfrac{1}{2}\left\{\frac{a_j h}{1-\beta_1}\right\}^2\right)} \to \frac{A_{j+1}}{A_j} < 1 \text{ as } j \to \infty .$$

By the ratio test for convergence of sums, we have $\sum_{j=1}^{\infty} A_j M\left(\tfrac{1}{2}\left\{\frac{a_j h}{1-\beta_1}\right\}^2\right) < \infty$ and therefore the stated conditions imply existence of the mth order unconditional moment for stock returns.

The proposition places lower bounds on θ in order for various unconditional moments to exist. These unconditional moments are sampling-theory expectations, *i.e.* they are taken against all possible data generating mechanisms. From the Bayesian point of view, we do *not* need these conditions because we do not require that unconditional moments exist since we condition on the available data. However, sampling-theory treatments of the equity premium puzzle do rely heavily on such conditions.

Next, we consider the issue of existence of equilibrium with finite stock prices in some more detail. This issue is of obvious importance since model implications based on simulation cannot have any validity unless conditions for equilibrium with finite prices can be established.

Proposition 5. If ω_ts follow a gamma distribution with shape λ and scale θ and the endowment process is as in equation (3) the condition for finite equilibrium stock prices is $\ln\delta + \frac{\rho\beta_0}{1-\beta_1} < \lambda \ln\left[1 - \frac{\rho^2}{2\theta(1-\beta_1)^2}\right]$ provided $\rho^2 < 2\theta(1-\beta_1)^2$.

Proof.

The condition for finite stock prices by Proposition 2, is

$$\delta \exp\left[\beta_0(a^* + \rho)\right] M\left(\frac{(a^*+\rho)^2}{2}\right) < 1$$

where $(a^* + \rho) = \frac{\rho}{1-\beta_1}$, and $M(t) = (1 - t/\theta)^{-\lambda}$, $t < \theta$. Therefore, the condition becomes

$$\lambda < \frac{\ln\delta}{\ln\left[1 - \frac{\rho^2}{2\theta}\right]}$$

or

$$\ln\delta + \frac{\rho\beta_0}{1-\beta_1} < \lambda \ln\left[1 - \frac{\rho^2}{2\theta(1-\beta_1)^2}\right] \tag{A.12}$$

and for this to be well defined, we need $\rho^2 < 2\theta(1-\beta_1)^2$.

It is clear that condition (A.12) is easily satisfied if $\theta \to \infty$, $\rho = 1-\alpha$ is sufficiently negative, and $\beta_0(1-\beta_1) > 0$.

Define $\eta = \beta_0/(1-\beta_1)$, $\tau = 1/(1-\beta_1)$, and

$$f(\tau) = \frac{\lambda \ln\left[1 - \frac{\rho^2\tau^2}{2\theta}\right] - \ln\delta}{\rho}$$

so that the existence conditions become

$$n < f(\tau), \text{ if } \rho > 0$$

$$n > f(\tau), \text{ if } \rho < 0$$

with $\tau < \tau^* = \frac{\sqrt{2\theta}}{|\rho|}$. Taking derivatives of the $f(\tau)$ function, we can show easily that:

$$f'(\tau) < 0 \text{ and } f''(\tau) < 0, \text{ if } \rho > 0$$

$$f'(\tau) > 0 \text{ and } f''(\tau) > 0, \text{ if } \rho < 0.$$

We consider, the case $\rho = 1 - \alpha < 0$, and $\beta_0 \in [-1,1]$. It is also clear that we must have $\eta \geq -\frac{1}{2}$ and $\eta = \beta_0 \tau \leq \tau$. If we take account of these restrictions, existence can be analyzed in the (η, τ) space as in Figure A1.

The region where stock prices are finite is shown by the shaded area. The convex curve has a slope $f'(\tau)$ which is approximately constant if $\lambda \approx 0$ or $\rho = 1 - \alpha \approx 0$. Analysis in terms of (β_0, β_1) is somewhat more involved but can be facilitated if we use a simulation approach. This approach has been utilized as well but results are not reported to save space.

PART B. GIBBS SAMPLING FOR THE NORMAL-GAMMA MODEL

The following assumption presents the normal-gamma linear model by treating the heteroscedastic weights w_i as parameters with appropriate priors, see (12) and (13).

Assumption 1. Let $\mathbf{y}=\mathbf{X}\beta+\mathbf{u}$, where $\mathbf{u}$ is *i.i.d* $\mathrm{N}(0,\Omega)$, $\Omega = diag(w)$, and w=$[\mathrm{w}_1, w_2, \ldots., w_n]'$ is a vector of parameters, whose prior distribution is $\pi(w_i|\lambda,\theta)$, *i.e.* $w_i \sim$ *i.i.d* $G(\lambda,\theta)$ (i=1,…,n).

Assumption 1 gives rise to posterior distribution $\pi(\beta,\lambda,\theta,\mathbf{w}|\mathbf{y},\mathbf{X})$ whereas (16) produces a posterior $\pi^*(\beta,\lambda,\theta|\mathbf{y},\mathbf{X})$ It is not difficult to show that the two posteriors are equivalent in the sense that $\int \pi(\beta,\lambda,\theta,\mathbf{w}|\mathbf{y},\mathbf{X})d\mathbf{w} = \pi^*(\beta,\lambda,\theta|\mathbf{y},\mathbf{X})$. Therefore, to perform inferences about β,λ and θ we can use the augmented posterior $\pi(\beta,\lambda,\theta,\mathbf{w}|\mathbf{y},\mathbf{X})$ and keep only the marginal in (β,λ,θ). This idea is known as data augmentation, and is further elaborated in Tanner and Wong (1987) and Geweke (1993, 1994b). For other applications of this idea, see Tsionas (1999).

The augmented posterior, written out in detail is

$$\pi(\beta,\lambda,\theta,\mathbf{w}\mid\mathbf{y},\mathbf{X})=\theta^{n\lambda}\Gamma(\lambda)^{-n}\det(\Omega)^{\lambda-3/2}\exp[-\tfrac{1}{2}(\mathbf{y}-\mathbf{X}\beta)'\Omega^{-1}(\mathbf{y}-\mathbf{X}\beta)-\theta\, tr(\Omega)]\pi(\beta,\lambda,\theta)$$

where $\pi(\beta,\lambda,\theta)$ is the joint prior of parameters.

Given the joint prior, the augmented posterior is given in the paper as equation (20). The posterior conditional distributions required for Gibbs sampling are described in the following.

Posterior conditional distributions

The posterior conditional distribution of β is given by

$$\beta|\lambda,\theta,\mathbf{w},\mathbf{y},\mathbf{X}\sim N(\tilde{\beta}(\Omega),[\mathbf{X}'\Omega^{-1}\mathbf{X}]^{-1}) \quad \text{(B.1)}$$

where $\tilde{\beta}(\Omega)=(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y}$. Drawing random numbers from this distribution is straightforward.

The posterior conditional distribution of λ

$$\pi(\lambda|\beta,\theta,\mathbf{w},\mathbf{y},\mathbf{X})\propto\lambda^{2m-1}\theta^{n\lambda}\Gamma(\lambda)^{-n}\exp[\sum_{i=1}^{n}\log w_i-s\lambda^2]. \quad \text{(B.2)}$$

This is not a standard density, but obtaining random samples can be accomplished by using acceptance techniques, employing as source density an exponential whose parameter is chosen to maximize the acceptance rate. This method closely follows the method described in Geweke (1994a).

The posterior conditional distribution of θ, has kernel density given by:

$$\pi(\theta|\beta,\lambda,\mathbf{w},\mathbf{y},\mathbf{X})\propto\theta^{n\lambda+d-1}\exp[-(b+tr(\Omega))\theta], \quad \text{(B.3)}$$

which implies that $\theta\mid\beta,\lambda,\mathbf{w},\mathbf{y},\mathbf{X}\sim G(n\lambda+d,b+tr(\Omega))$. Drawing random numbers from this distribution is straightforward.

[14] The number of draws is determined so that the plots are as informative as possible.

The posterior conditional distribution of w can be written as $\pi(\mathbf{w}|\beta,\lambda,\theta,\mathbf{yX})=\prod_{i=1}^{n}\pi(w_i|\beta,\lambda,\theta,\mathbf{yX})$, where

$$\pi(w_i|\beta,\lambda,\theta,\mathbf{yX})\propto w_i^{\lambda-3/2}\exp[-\theta w_i - q_i w_i^{-1}],\ w_i>0,\ \text{i=1,2,…,n} \qquad \text{(B.4)}$$

and $q_i=(y_i-\mathbf{x}_i'\beta)^2/2$.

This is not a standard density but acceptance methods can be used. Since it is the product of a $G(\lambda,\theta)$ and an inverted gamma $IG(1/2,q_i)$ we can draw from the first distribution, and accept the draw with probability $(2q_i e/w_i)^{1/2}\exp(-q_i/w_i)$. Another idea would be to approximate (B.4) by a *gamma* density, and use optimal acceptance sampling techniques, or take a Metropolis step. This approach was not pursued.

Initial values of parameters were selected from burn-in runs of 500 Gibbs iterations. The Gibbs sampler was implemented using 5,000 subsequent iterations. Based on the convergence diagnostic of Geweke (1992) it was not possible to reject the hypothesis that means of the parameters and functions of interest are the same in the first 500 and last 2,500 iterations. The evaluation of numerical accuracy proceeds using the numerical standard error (NSE) and relative numerical accuracy (RNE), see Geweke (1992). Both require estimates of the spectral density of draws at the origin, that were formed using an AR(20) pre-filter enforcing stationarity. The ordinates around zero were smoothed using a Daniel window of width 20 ordinates, as detailed in the appendix of Geweke (1994a).

Implementation of Gibbs sampling

Starting from initial conditions $\beta^{(0)},\theta^{(0)},\lambda^{(0)}$ the Gibbs sampling is an iterative sampling scheme which updates these values as follows. For $i=0,1,2,...M$

Draw $\beta^{(i)}|\theta^{(i-1)},\lambda^{(i-1)},\mathbf{w}^{(i-1)},\mathbf{y,X}$ from its conditional distribution in (B.1).

Draw $\theta^{(i)}|\beta^{(i)},\lambda^{(i-1)},\mathbf{w}^{(i-1)},\mathbf{y,X}$ from its conditional distribution in (B.3).

Draw $\lambda^{(i)}|\beta^{(i-)},\theta^{(i)},\mathbf{w}^{(i-1)},\mathbf{y,X}$ from its conditional distribution in (B.2).

Draw $\mathbf{w}^{(i)}|\beta^{(i)},\theta^{(i)},\lambda^{(i)},\mathbf{y,X}$ from its conditional distribution in (B.4).

This produces a (non-random) sample $\{\beta^{(i)},\theta^{(i)},\lambda^{(i)};i=1,..,M\}$ that converges in distribution to the posterior. In a sense, this Gibbs sample reconstructs the posterior distribution and therefore, marginal posterior moments or densities can be computed. For example, the marginal posterior density of λ can be estimated by using kernel density estimation techniques for the Gibbs sub-sample $\{\lambda^{(i)}; i = 1,\ldots,M\}$. The posterior expectation of a function of the parameters, say $f(\beta,\lambda,\theta)$, is given by

$$E\{f(\beta,\lambda,\theta) \mid \mathbf{y},\mathbf{X}\} = \int f(\beta,\lambda,\theta)\pi(\beta,\lambda,\theta \mid \mathbf{y},\mathbf{X})d\beta d\lambda d\theta$$

and can be approximated by

$$M^{-1}\sum_{i=1}^{M} f(\beta^{(i)},\lambda^{(i)},\theta^{(i)}).$$

Such functions of interest include volatilities of growth regressions, asset return averages *etc*., as detailed in section 3.3 of the paper.

References

Campbell, J.Y., and J.H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, Journal of Political Economy 107, 205-251.

Cecchetti S.G., P.-S. Lam and N.C. Mark, 1993, The equity premium and the risk-free rate: Matching the moments, Journal of Monetary Economics 31, 21-45.

Cecchetti S.G. and N.C. Mark, 1990, Evaluating empirical tests of asset pricing models: Alternative interpretations, American Economic Review 80, 48-51.

Chib, S., 1995, Marginal likelihood from the Gibbs output, Journal of the American Statistical Association 90, 1313-1321.

de Finetti B., 1961, The Bayesian approach to the rejection of outliers, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1, 199-210, University of California Press.

Duffie D. and K.J.Singleton, 1994, Simulated moments estimation of Markov models of asset prices, Econometrica 62, 929-952.

Gelfand A.E. and A.F.M. Smith, 1990, Sampling based approaches to calculating marginal densities, Journal of the American Statistical Association 85, 398-409.

Geweke J., 1992, Evaluating the accuracy of sampling based approaches to the calculation of posterior moments, in Berger J.O., J.M. Bernardo, A.P. Dawid and A.F.M. Smith (editors), Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics: 169-194, Oxford University Press.

Geweke J., 1993, Bayesian treatment of the independent Student-t linear model, Journal of Applied Econometrics 8, S19-S40.

Geweke J., 1994a, Priors for macroeconomic time series and their application, Econometric Theory 10, 609-632 [technical appendix in Federal Reserve Bank of Minneapolis Discussion Paper No.64, May 1992].

Geweke J., 1994b, Monte Carlo simulation and numerical integration, in Amman H., D. Kendrick and J. Rust, (editors) Handbook of Computational Economics, North Holland, chapter 15.

Geweke J., 1999, Computational experiments and reality, manuscript, Department of Economics, University of Minnesota.

Grossman S.J., and R.J.Shiller, 1981, The determinants of the variability of stock market prices, American Economic Review: 71, 222-227.

Hansen, L.P., H. Heaton, and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, Journal of Business and Economic Statistics 14, 262-280.

Hansen, L.P., and R. Jagannathan, 1991, Implications of security market data for models of dynamic economies, Journal of Political Economy 99, 225-262.

Jarque, C.M., and A.K. Bera, 1987, A test for normality of observations and regression residuals, International Statistical Review 55:163-172.

Jeffreys, H., 1961, Theory of Probability, third edition, Oxford: Clarendon Press.

Kandel S. and R.F. Stambaugh, 1990, Expectations and volatility of consumption and asset returns, Review of Financial Studies 3, 207-232.

Kocherlakota N.R., 1996, The equity premium: It's still a puzzle, Journal of Economic Literature 34, 42-71.

Kocherlakota, N.R., 1997, Testing the consumption CAPM with heavy-tailed pricing errors, Macroeconomic Dynamics 1, 551-567.

Labadie P., 1986, Stochastic inflation and the equity premium, Journal of Monetary Economics 24, 277-298.

Lindley D.V., 1971, The estimation of many parameters, in Godambe V.P. and D.A. Sprott (editors), Foundations of Statistical Inference, Holt, Rinehart and Winston Inc.

Lucas, R.E. Jr., 1978, Asset prices in an exchange economy, Econometrica 66, 1429-1455.

Mankiw, G.N., and S.P. Zeldes, 1991, The consumption of stockholders and non-stockholders, Journal of Financial Economics 17, 211-219.

Mehra R. and E.C. Prescott, 1985, The equity premium: A puzzle, Journal of Monetary Economics 15, 145-161.

Mehra R. and E.C. Prescott, 1988, The equity risk premium: A solution?, Journal of Monetary Economics 18, 133-136.

Meng, X.L., 1994, Posterior predictive p-values, Annals of Statistics 22, 1142-1160.

Rietz T.A., 1988, The equity premium: A solution, Journal of Monetary Economics 22, 17-33.

Roberts G.O. and A.F.M. Smith, 1994, Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms, Stochastic Processes and their Applications 49, 207-216.

Tanner M.A. and W.H.Wong, 1987, The calculation of posterior distributions by data augmnetation, Journal of the American Statistical Association 82, 528-550.

Tierney L., 1994, Markov chains for exploring posterior distributions (with discussion), Annals of Statistics 22, 1701-1762.

Tsionas E.G., 1999, Monte Carlo inference in econometric models with symmetric stable disturbances, Journal of Econometrics 88, 365-401.

West M., 1984, Outlier models and prior distributions in Bayesian linear regression, Journal of the American Statistical Association, Series B, 46, 431-439.

Zellner A., 1976, Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms, Journal of the American Statistical Association 71, 400-405.

CAPTIONS FOR FIGURES 1 - 6 AND FIGURE A1.

FIGURE 1. This Figure provides quantile-quantile plots for the consumption, dividends and GNP data. The horizontal axis provides the empirical quantiles and the vertical axis provides the theoretical quantiles that would result from the normal distribution. If the data were normally distributed, all points should lie on a straight line.

FIGURE 2. The upper panel reports the induced prior distribution of consumption growth's standard deviation, ν, and the lower panel provides the induced prior for output growth. These quantities are functions of the parameters so a prior on the parameters implies a prior on the standard deviation of consumption growth and growth itself. The purpose of computing these priors is to ensure that the priors do not yield unreasonable predictions about crashes in consumption. These figures are generated using random drawings of parameter values from the prior.

FIGURE 3. This Figure reports the induced prior distribution of excess kurtosis in consumption growth. Excess kurtosis is function of the parameters of the model so a prior on the parameters implies an excess kurtosis prior. This prior is computed using random drawings of parameter values from their prior.

FIGURE 4. The upper panel reports the induced prior distribution of equity premium and the lower panel reports the induced prior of real risk-free rates from the model. Setting the regression parameters to zero, these quantities are functions of the parameters whose definitions are provided in equations (10) and (11). Using random drawings for the parameters of the model, these quantities can be computed and marginal distributions can be approximated using kernel density estimation techniques using the available draws.

FIGURE 5. This Figure reports an aspect of the bivariate posterior distribution of equity premium and real risk-free rate when the coefficient of relative risk aversion is 2 and the discount factor is 0.98. We use the three data sets (consumption, dividends and GNP). The construction is based on the following scheme. 500 endowment time series were generated for each one of the posterior draws of the parameters. For each time series we computed equity premia and real rates using equations (10) and (11) and recorded their sample averages. The averages are plotted in the two panels of these figures to provide an aspect of the joint distribution of asset returns.

FIGURE 6. This Figure provides the marginal posterior distribution of Sharpe's ratio for the three data sets, consumption, dividends, and GNP, in the upper, middle and lower panel respectively. The Sharpe ratio is defined as equity premium divided by the standard deviation of stock returns. For each parameter draw, a model simulation provides not only an average equity premium but its volatility as well, so the Sharpe ratio can be computed and is a convenient way to summarize the model’s implications about two moments simultaneously.

FIGURE A1. This Figure summarizes graphically the conditions for existence of an equilibrium in the asset pricing model with a normal-gamma mixture distribution for the error term in a consumption-growth autoregression of order one. The existence conditions are stated in terms of parameters η and τ, which are defined in Appendix A.

Figure 1. Quantile-quantile plots of the data

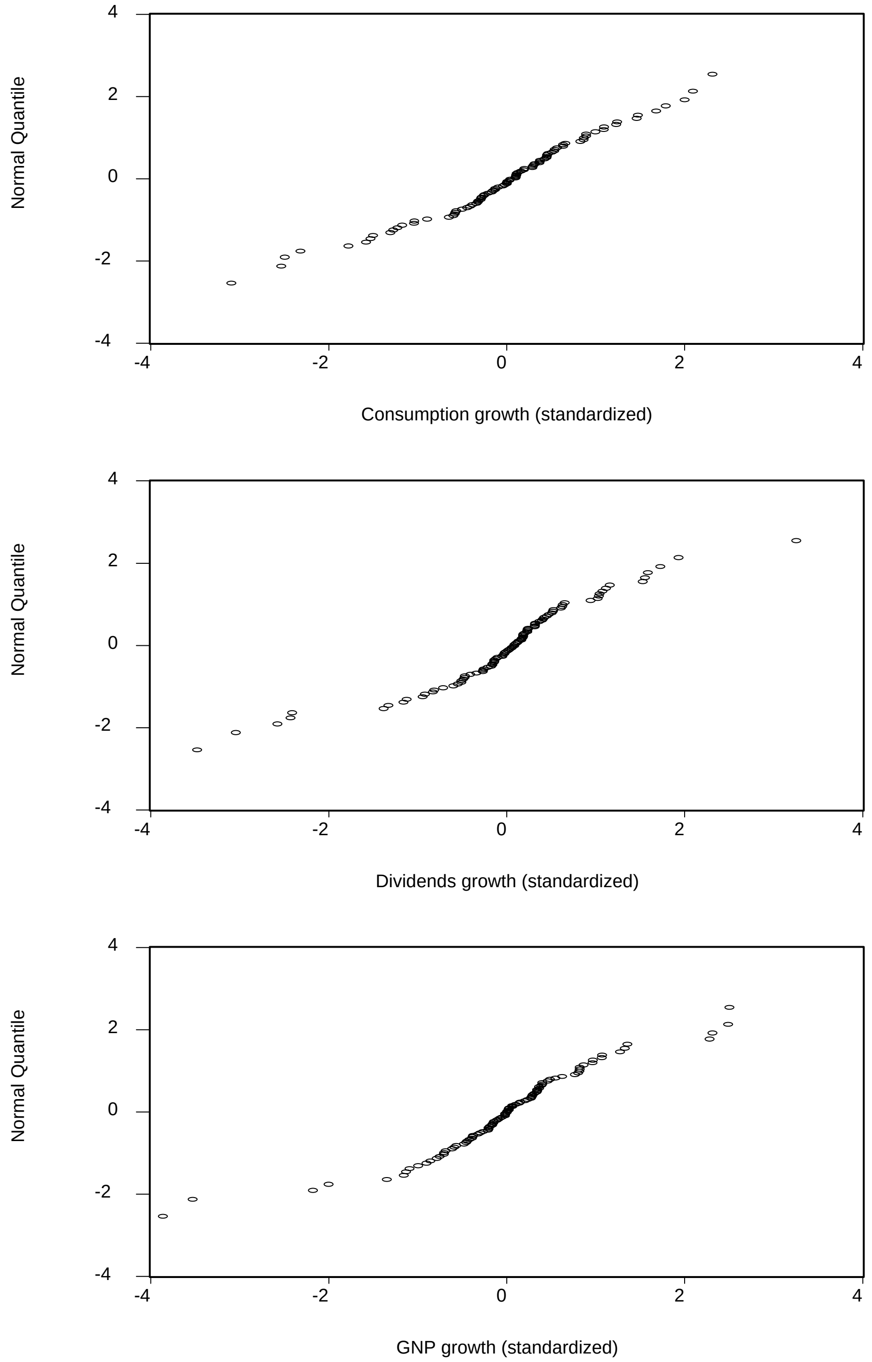


Figure 2. Induced priors for ν and output growth

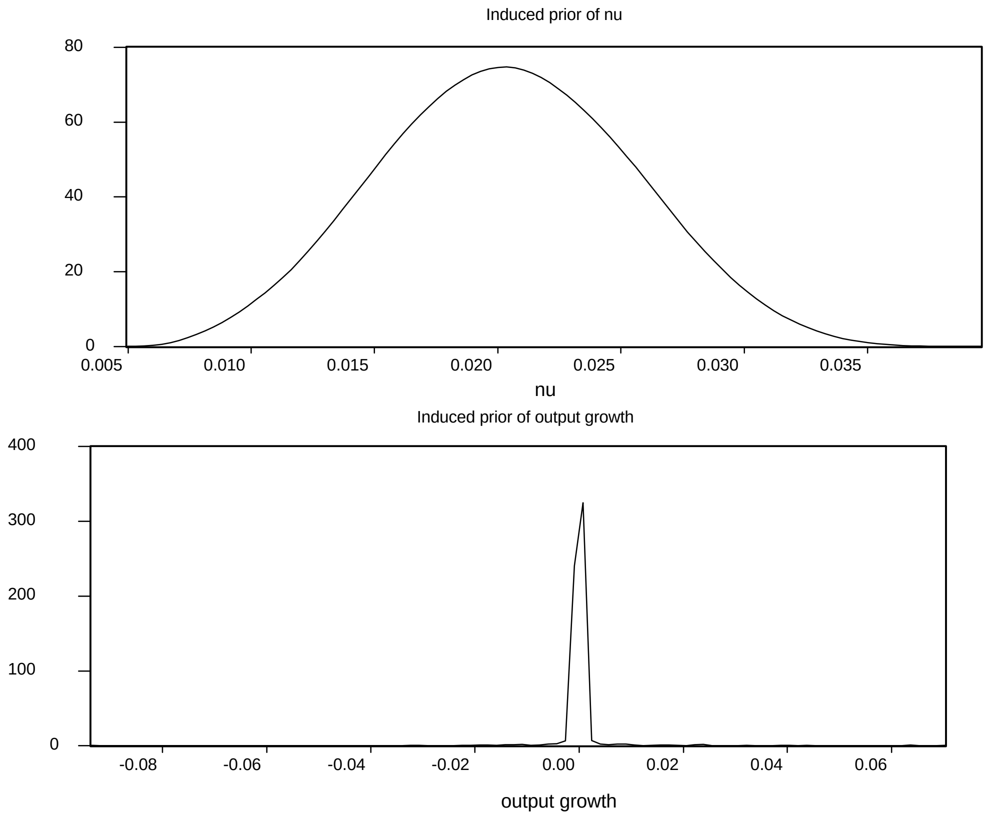


Figure 3. Induced prior for excess kurtosis

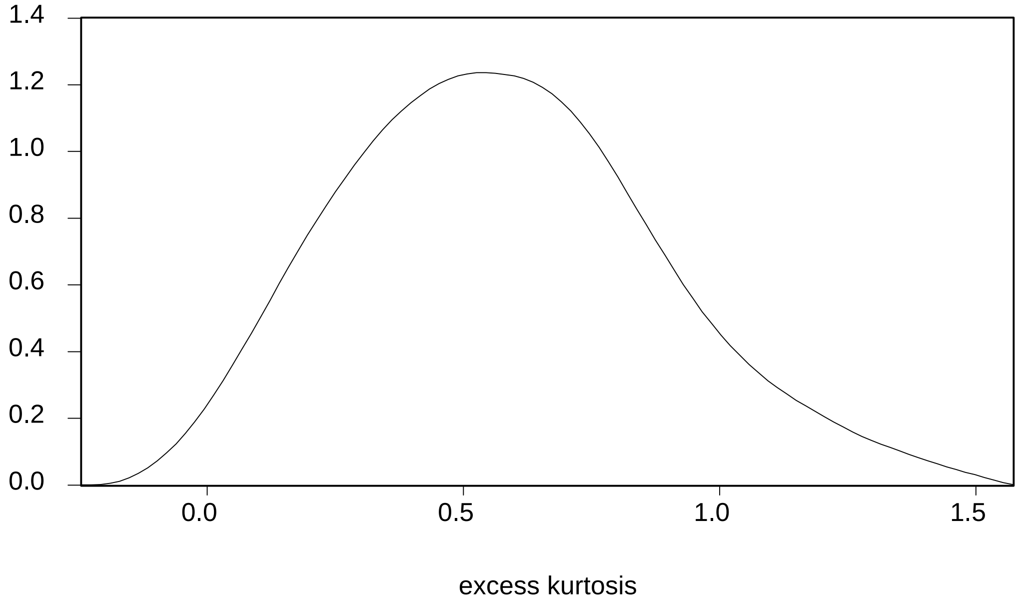


Figure 4. Induced prior for asset returns, $\alpha = 2$

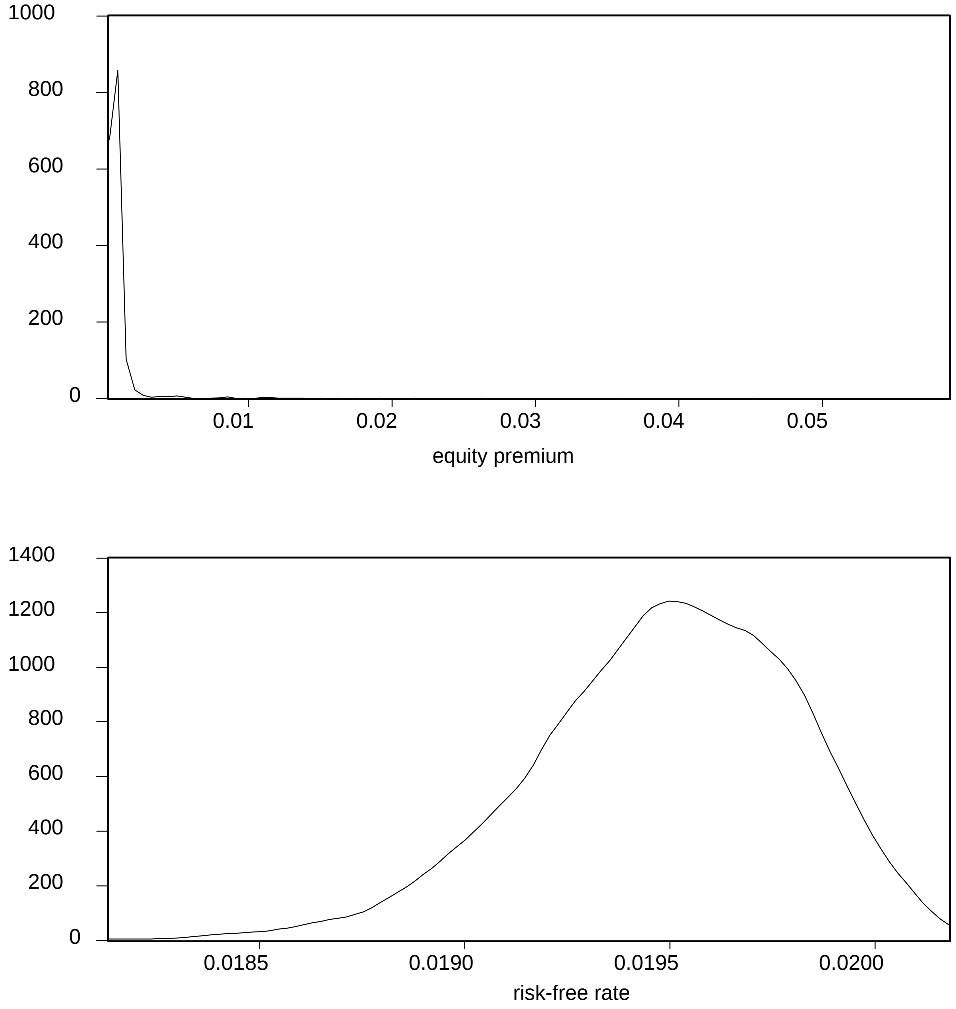


Figure 5. Bivariate posterior distribution of asset returns

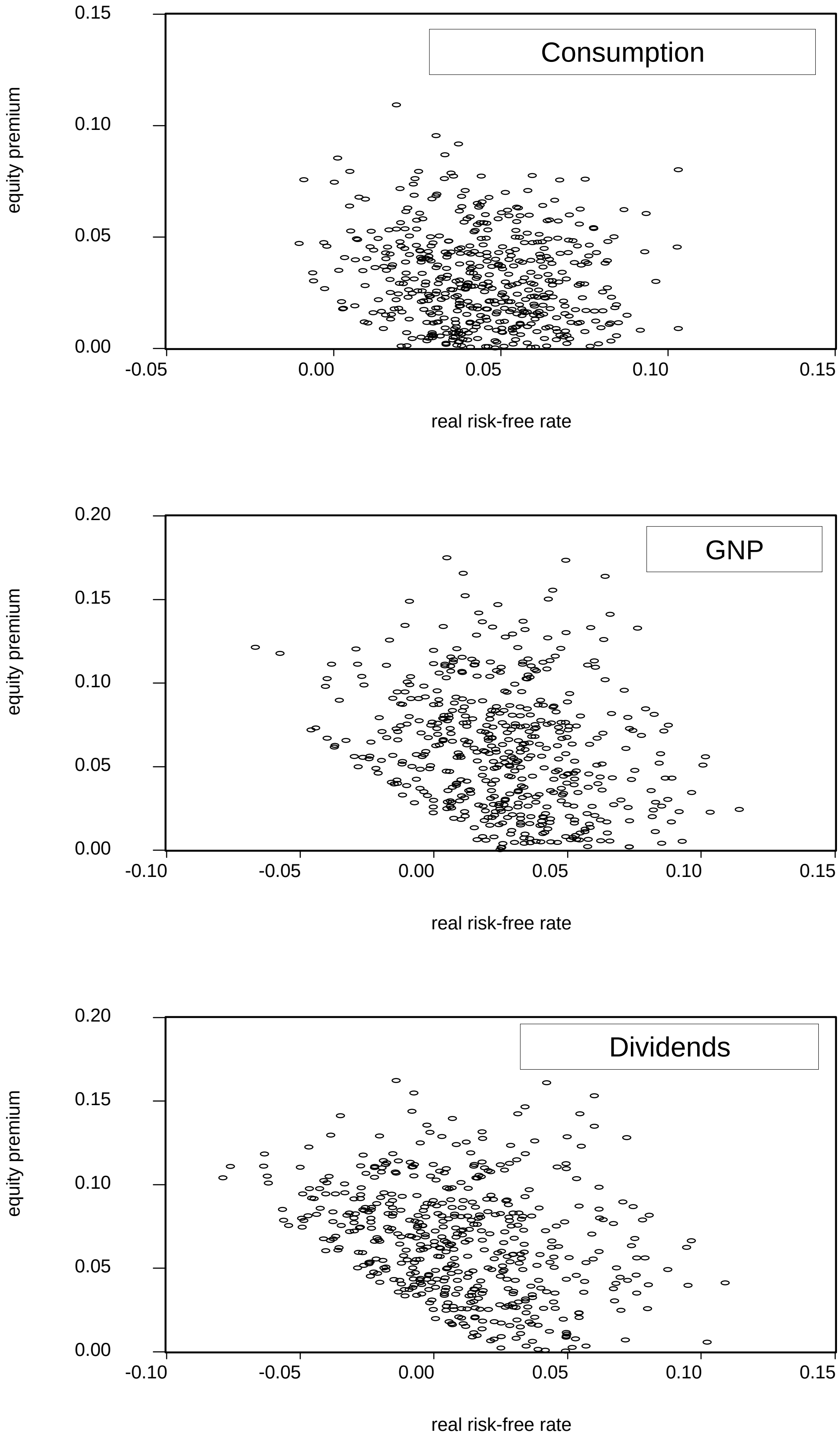
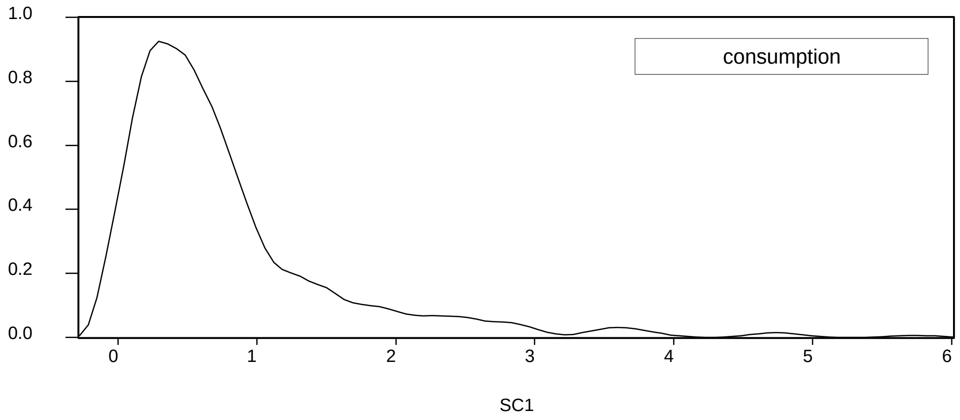
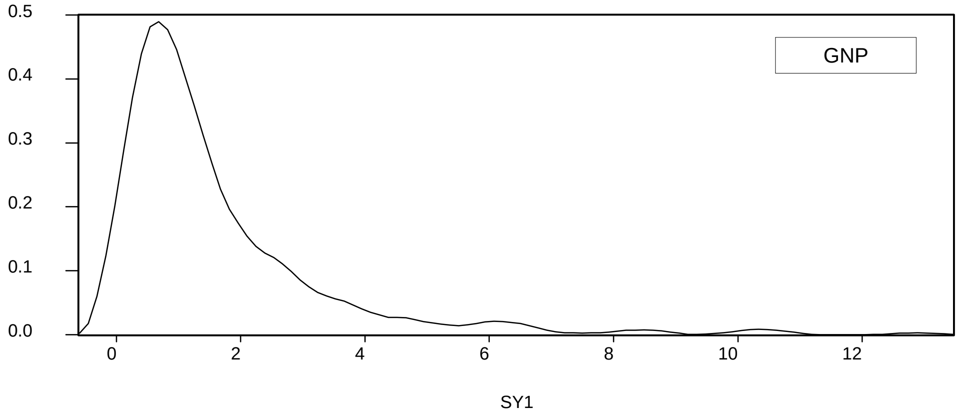


Figure 6. Posterior distributions of Sharpe ratio





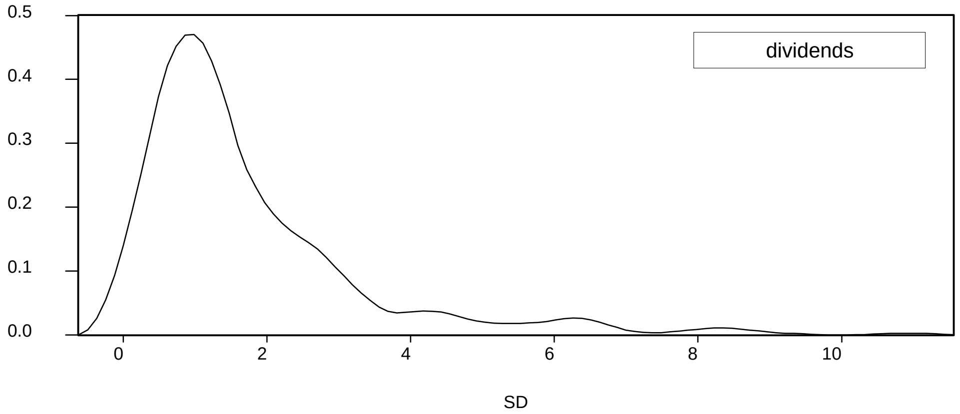


Figure A1. Exact conditions for existence of an equilibrium