EXPLICIT FORMULAE FOR PARAMETERS OF STOCHASTIC MODELS OF A DISCOUNTED EQUITY INDEX USING MAXIMUM LIKELIHOOD ESTIMATION WITH APPLICATIONS

K. FERGUSSON

ABSTRACT. A discounted equity index is computed as the ratio of an equity index to the accumulated savings account denominated in the same currency. In this way, discounting provides a natural way of separating the modelling of the short rate from the market price of risk component of the equity index. In this vein, we investigate the applicability of maximum likelihood estimation to stochastic models of a discounted equity index, providing explicit formulae for parameter estimates. We restrict our consideration to two important index models, namely the Black-Scholes model and the minimal market model of Platen, each having an explicit formula for the transition density function. Explicit formulae for estimates of the model parameters and their standard errors are derived and are used in fitting the two models to US data. Further, we demonstrate the effect of the model choice on the no-arbitrage assumption employed in risk neutral pricing.

1. INTRODUCTION

A discounted equity index is computed as the ratio of an equity index to the accumulated savings account denominated in the same currency. In this way, discounting provides a natural way of separating the modelling of the short rate from the market price of risk component of the equity index. In this vein, we investigate the applicability of maximum likelihood estimation to stochastic models of a discounted equity index, providing explicit formulae for parameter estimates. We restrict our consideration to two important index models, namely the Black-Scholes model and the minimal market model of Platen, each having an explicit formula for the transition density function. The first model is the standard continuous time market model under the classical risk neutral assumption, whereas the second model is the standard model under the benchmark approach, discussed in Platen and Heath [2006]. Explicit formulae for the estimates of model parameters and their standard errors are derived which then are used in fitting the two models to US data. Further, we demonstrate the effect of the model choice on the classical no-arbitrage assumption employed in risk neutral pricing.

The application of maximum likelihood estimation is well studied for stochastic models of equity indices, starting from Mandelbrot [1963] and Fama [1963] and summarised more recently in Behr and Pötter [2009]. However, in the current article we are interested in the application to models of discounted equity indices, also

Date: June 6, 2017.

1991 Mathematics Subject Classification. Primary 62P05; Secondary 60G35, 62P20.

Key words and phrases. Stochastic short rate, maximum likelihood estimation, Black-Scholes model, squared Bessel process, minimal market model, modified Bessel function of the first kind.

JEL Classification. C02, C13.
examined in Baldeaux et al. [2015]. Our motivation stems from the benchmark approach, introduced in Platen [2004], whereby a benchmark is constructed as the “best” (in several ways) performing portfolio for use as a numéraire or reference unit. The benchmark approach uses the growth optimal portfolio (GP) as benchmark, as discussed in Platen and Heath [2006]. The GP achieves the maximum possible expected growth rate at any time, and also the almost surely maximum long-term growth rate, as shown in Platen [2004]. When used as benchmark, each benchmarked non-negative portfolio can be shown to be a supermartingale with its current benchmarked value greater than or equal to the expected future benchmarked values. As such, the GP is the “best” performing portfolio in this sense. It has been studied and employed previously, for example in Kelly [1956], Long [1990], Merton [1971], Karatzas and Shreve [1998], Platen [2002] and by many other authors.

We work on a filtered probability space \((\Omega, \mathcal{A}, (\mathcal{A}_t)_{t \geq 0}, P)\), where \(\Omega\) is the sample space, \(\mathcal{A}\) is the set of events, \((\mathcal{A}_t)_{t \geq 0}\) is the filtration of events modelling the evolution of information and \(P\) is the real world probability measure. It is shown in Platen [2004] that the GP, denoted by \(S^{\delta*}_t\) at time \(t\), for \(t \in [0, T]\), satisfies the SDE

\[
dS^{\delta*}_t = S^{\delta*}_t \left( r_t + \theta^2_t \right) dt + S^{\delta*}_t \theta_t d\hat{W}_t,
\]

where \(r_t\) is the short rate, \(\theta_t\) is the market price of risk and \(\hat{W}\) is a Wiener process.

In Platen [2005] and Platen and Rendek [2012] it is shown that appropriately diversified portfolios represent approximate GPs. Therefore, a number of common, well-diversified stock market indices can be used to approximate the GP, including but not limited to the following: the Standard and Poor’s 500 Index (S&P 500) and the Russell 2000 Index for the US market and the MSCI Growth World Stock Index (MSCI) for global modelling.

In Figure 1 we plot the logarithm of the S&P 500 denominated in United States dollars (USD), and normalised to one at the start, over the period from January 1871 to March 2017, using data sourced from the website of Shiller [1989]. By assuming that the GP for the US equity market is approximated by the S&P 500, Figure 1 can be interpreted as the logarithm of a historical sample path for the GP.

Note from (1.1) that the GP dynamics are completely characterized by the short rate \(r_t\) and the market price of risk \(\theta_t\). Letting \(B_t\) denote the accumulated value of the savings account, satisfying the SDE \(dB_t = r_t B_t dt\), with \(B_0 = 1\), we can separate these two effects by considering the discounted GP process \(\bar{S}^{\delta*}_t = \{\bar{S}^{\delta*}_t, t \in [0, T]\}\), given by

\[
\bar{S}^{\delta*}_t = \frac{S^{\delta*}_t}{B_t},
\]

satisfying the SDE

\[
d\bar{S}^{\delta*}_t = \bar{S}^{\delta*}_t \theta^2_t dt + \bar{S}^{\delta*}_t \theta_t d\hat{W}_t,
\]

for \(t \in [0, T]\).

From, for example, Karatzas and Shreve [1998], one notes that in a complete market, the candidate Radon-Nikodym derivative process \(\Lambda^Q = \{\Lambda^Q_t : t \in [0, T]\}\) for the putative risk neutral measure \(Q\) is equal to the reciprocal of the discounted
GP. That is,

\[ \Lambda_t^Q = \frac{dQ}{dP} \bigg|_{\Delta_t} = \frac{B_t}{\bar{S}_t^\delta} \cdot \frac{S_0^\delta}{\bar{S}_0^\delta}. \]

A necessary condition for \( Q \) to be a probability measure is that

\[ Q(\Omega) = 1. \]

Violation of Condition 1.5 implies that \( Q \) is not equivalent to \( P \) and, therefore, an equivalent risk neutral probability measure does not exist in this case. The Fundamental Theorem of Asset Pricing, as given in Delbaen and Schachermayer [2006], states the classical no-arbitrage condition which is equivalent to the existence of an equivalent risk neutral probability measure. Therefore, violation of Condition 1.5 means that the corresponding dynamics permit some form of classical arbitrage. It is important to understand whether in reality one can support the classical no-arbitrage assumption or whether there is significant evidence from historical data that favour a more general modelling framework where this assumption is not imposed.

In this article we consider a complete market. Two models of the discounted GP \( \bar{S}_t^\delta \) are considered, these being the Black-Scholes (BS) model,

\[ d\bar{S}_t^\delta = \bar{S}_t^\delta \theta_t^2 dt + \bar{S}_t^\delta \theta_t dW_t, \]

where \( \theta_t = \theta > 0 \), and the minimal market model (MMM),

\[ d\bar{S}_t^\delta = \bar{\alpha}_t dt + \sqrt{\bar{S}_t^\delta} \bar{\alpha}_t dW_t, \]
where $\theta_t = \sqrt{\bar{\alpha}_t / \bar{S}^\delta_t}$ with $\bar{\alpha}_t = \bar{\alpha}_0 \exp(\eta t)$, $\eta > 0$.

The Black-Scholes model is the standard market model under classical no-arbitrage assumptions. It derives its name from the geometric Brownian model assumed in Black and Scholes [1973].

The MMM includes the discounted GP process in (1.7) with the requirement that the net GP drift term $\bar{\alpha}_t = \bar{S}^\delta_t |\theta_t|^2$ grows exponentially as $\bar{\alpha}_t = \bar{\alpha}_0 \exp(\eta t)$ with the net growth rate $\eta > 0$ reflecting the long term average growth rate of $\bar{S}^\delta_t$ and, thus, the economy. Both models have one key parameter, the volatility $\theta$ for the BS model and the net growth rate $\eta$ for the MMM.

We will now fit both models to historical data.

2. Black-Scholes Model of Discounted GP

From (1.3) we have the SDE that is assumed to be satisfied by the discounted GP under the BS model. By insisting that the market price of risk $\theta_t$ be a constant $\theta$, the SDE for the discounted GP $\bar{S}^\delta_t$ becomes

\[(2.1) \quad d\bar{S}^\delta_t = \bar{S}^\delta_t (\theta^2 dt + \theta dW_t)\]

From (2.4) the theoretical quadratic variation of $\log(\bar{S}^\delta_t)$ is

\[(2.2) \quad [\log(\bar{S}^\delta_t)]_t = \int_0^t \theta^2 ds = \theta^2 t,\]

which we show in Figure 2 alongside the actual empirical quadratic variation of the logarithm of the discounted GP, computed from Shiller’s series of discounted values of the S&P 500 over the period from 1871 to 2017, given in Appendix C.

**Remark 1.** It is evident from Figure 2 that the Black-Scholes model of the discounted GP fails to capture the stochastic nature of the market price of risk $\theta_t$ since the slope of the quadratic variation of the logarithm of the discounted stock index moves significantly over time.

From (1.4) the Radon-Nikodym derivative of the risk neutral probability measure is equal to the reciprocal of the discounted GP which, under the BS model, is given by

\[(2.3) \quad \Lambda_t^Q = \exp(-\frac{1}{2} \theta^2 t - \theta W_t).\]

**Remark 2.** Risk neutral pricing of derivative contracts relies on the Radon-Nikodym derivative $\Lambda_t^Q$ being a martingale with respect to the real-world probability measure $P$. This condition is indeed satisfied under the model because $E(\Lambda_t^Q | A_s) = \Lambda_s^Q$ for $0 \leq s \leq t < \infty$.

We show in Figure 3 $\Lambda_t^Q = 1/\bar{S}^\delta_t$ for the S&P500, where we note that this trajectory seems unlikely to be that of a true martingale.

In the following section we derive the transition density function of the discounted GP, used to fit the parameters to data.
Figure 2. Quadratic variation of logarithm of discounted stock index.

2.1. Transition Density of Discounted GP. The logarithm of the discounted GP obeys the SDE

\[
d\log (\bar{S}_t^\delta) = \frac{1}{2} \theta^2 dt + \theta dW_t
\]

whose solution is immediately found to be \( \log (\bar{S}_T^\delta) = \log (\bar{S}_0^\delta) + \frac{1}{2} \theta^2 t + \theta (W_t - W_0) \).

This gives rise to the following lemma.

**Lemma 1.** The transition density function of the logarithm of the discounted GP in (2.1) is

\[
p_{\log (\bar{S}_t^\delta)}(t, \log(x_t), \bar{T}, \log(x_{\bar{T}})) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{(\bar{T} - t)\theta^2}} \exp \left( -\frac{1}{2} \frac{\log (\frac{x_t}{x_{\bar{T}}}) - \frac{1}{2} (\bar{T} - t) \theta^2}{\sqrt{(T - t)\theta^2}} \right)^2
\]

and the transition density function of the discounted GP is

\[
p_{\bar{S}_t^\delta}(t, x_t, \bar{T}, x_{\bar{T}}) = \frac{1}{x_T \sqrt{2\pi} \sqrt{(T - t)\theta^2}} \exp \left( -\frac{1}{2} \frac{\log (\frac{x_t}{x_T}) - \frac{1}{2} (T - t) \theta^2}{\sqrt{(T - t)\theta^2}} \right)^2.
\]

By virtue of this lemma we can write the conditional distribution of \( \bar{S}_T^\delta \) given \( \bar{S}_t^\delta \) as

\[
\log (\bar{S}_T^\delta) \sim N \left( \log \bar{S}_t^\delta + \frac{1}{2} (T - t) \theta^2, (T - t) \theta^2 \right)
\]
Below, we plot the transition density function of the logarithm of the discounted GP in Figure 4 for the fitted parameter $\theta$ when starting in January 1871 with $\overline{S}_0^\delta = 1$. We also include in the graph the theoretical mean of the logarithm of the index and the logarithm of the index.

2.2. Fitting the Black-Scholes Model. We use maximum likelihood estimation to fit the model (2.1) to Shiller’s monthly series of discounted values of the S&P 500 over the period from 1871 to 2017, given in Appendix C.

For a time discretisation $t_i = i\Delta$, $\Delta > 0$, our log-likelihood function is

$$\ell_\Delta(\theta) = -\frac{1}{2} \sum_{i=1}^{n} \left\{ \log(2\pi \theta^2 \Delta) + \frac{(\log \overline{S}_{t_i}^\delta - \log (\overline{S}_{t_i-1}^\delta) - \frac{1}{2} \theta^2 \Delta)^2}{\theta^2 \Delta} \right\},$$

and, although widely known (for example, see Chapter 8 of Rice [2007]), an explicit formula for the maximum likelihood estimate of the parameter $\theta$ and its standard error is supplied in the following theorem.

**Theorem 1.** The square of the maximum likelihood estimate $\hat{\theta}_\Delta$ of $\theta$ in the SDE (2.1), with time step size $\Delta$, is given by

$$\hat{\theta}^2_\Delta = \frac{-2 + 2 \sqrt{1 + \frac{1}{n} \sum_{i=1}^{n} (\log \overline{S}_{t_i}^\delta / \overline{S}_{t_{i-1}}^\delta)^2}}{\Delta}$$
Figure 4. Transition density function of logarithm of Black-Scholes discounted GP based at 1871 with trajectory of log discounted S&P500 and theoretical mean.

where there are \( n + 1 \) observations of the discounted GP \( \bar{S}^{\delta*}_{t_i} \), for \( i = 0, 1, 2, \ldots, n \). Further, the standard error of the parameter estimate is given by

\[
S.E.(\hat{\theta}_\Delta) = \frac{1}{\sqrt{n\Delta + \frac{2\theta^2}{\hat{\theta}_\Delta^2}}}.
\]

Proof. We rewrite (2.8) as

\[
\ell(\theta) = -\frac{1}{2} \sum_{i=1}^{n} \left\{ \log(2\pi\Delta) + 2\log(\theta) + \frac{(\log(\bar{S}^{\delta*}_{t_i}/\bar{S}^{\delta*}_{t_{i-1}}))^2}{\theta^2\Delta} - \log(\bar{S}^{\delta*}_{t_i}/\bar{S}^{\delta*}_{t_{i-1}}) + \frac{1}{4} \frac{\theta^2}{\Delta} \right\}
\]

and differentiating with respect to \( \theta \) gives

\[
\ell'(\theta) = -\frac{1}{2} \sum_{i=1}^{n} \left\{ \frac{2}{\theta} - \frac{2(\log(\bar{S}^{\delta*}_{t_i}/\bar{S}^{\delta*}_{t_{i-1}}))^2}{\theta^3\Delta} + \frac{1}{2} \frac{\theta}{\Delta} \right\}.
\]

Rearranging (2.12) gives

\[
\ell'(\theta) = -\frac{1}{4\theta^3} \sum_{i=1}^{n} \left\{ 4\theta^2 - \frac{4(\log(\bar{S}^{\delta*}_{t_i}/\bar{S}^{\delta*}_{t_{i-1}}))^2}{\Delta} + \theta^4 \Delta \right\}
\]

\[
= -\frac{1}{4\theta^3} \left( n\Delta \theta^4 + 4n\theta^2 - \sum_{i=1}^{n} \frac{4(\log(\bar{S}^{\delta*}_{t_i}/\bar{S}^{\delta*}_{t_{i-1}}))^2}{\Delta} \right).
\]
The solution to the equation $\ell'(\theta) = 0$ is given by the formula (2.9). To determine the standard error of the estimate of $\theta$ we take the second derivative of $\ell$, giving

\[
\ell''(\theta) = -\frac{1}{4\theta^4} \left( n\Delta \theta^4 - 4n\theta^2 + 3 \sum_{i=1}^{n} 4 \left( \log(\frac{S_{t_i}^{\delta_i}}{S_{t_{i-1}}^{\delta_i}}) \right)^2 \right).
\]

When $\theta = \hat{\theta}_\Delta$ the second derivative can be simplified to

\[
\ell''(\theta) = -\frac{1}{4\theta^4} \left( 4n\Delta \theta^4 + 8n\theta^2 \right) = -\left( n\Delta + 2n\theta^2 \right).
\]

The standard error is given by the reciprocal of the square root of the negative of the second derivative, giving the result (2.10).

From this theorem we compute the maximum likelihood estimate for various time step sizes. In Table 1 we show the estimate $\hat{\theta} = 0.141002 (0.002380)$ with the standard error shown in brackets.

In Figure 5 we plot the estimate with its 95% confidence interval in dependence on the time step size. We note that for some larger time step sizes the confidence intervals become unrealistic as the normality assumption of the standard error is invalid.
3. Minimal Market Model (MMM)

In this section we consider the minimal market model originally proposed in Platen [2001]. By making the assumption that the drift \( \hat{\alpha}_t = \theta t \bar{S}_t^\delta \) of the discounted GP in (1.3) behaves exponentially we arrive at the SDE (1.7) for the discounted GP, where

\[
(3.1) \quad \bar{\alpha}_t = \bar{\alpha}_0 \exp(\eta t)
\]

and \( \eta \) is the net growth rate. Here the net growth rate \( \eta \) can be viewed as the growth rate of the GP in excess of the short rate.

The square root of the discounted GP has SDE

\[
(3.2) \quad d(\sqrt{\bar{S}_t^\delta}) = \frac{3}{8\sqrt{\bar{S}_t^\delta}} \bar{\alpha}_t dt + \frac{1}{2} \sqrt{\bar{\alpha}_t} dW_t.
\]

It follows that the discounted GP drift can be observed as four times the first order time derivative of the quadratic variation of the square root of the discounted GP, that is,

\[
(3.3) \quad \bar{\alpha}_t = 4 \frac{d}{dt} [\sqrt{\bar{S}_t^\delta}]_t.
\]

In Figure 8 the quadratic variation of the square root of the discounted stock index is shown. It is clear that the exponential function \( \bar{\alpha}_t = \bar{\alpha}_0 \exp(\eta t) \) provides a good fit and thus confirms the assumptions for the net market trend in (3.1) and constant growth rate \( \eta \) as being reasonable.

Under the MMM, Platen [2002] showed that the discounted GP obeys a time-transformed squared Bessel process of dimension four.

We define the normalised discounted GP process \( Y_t \) as the ratio of the discounted GP to its net market trend. The SDE satisfied by \( Y \) is

\[
(3.4) \quad dY_t = \bar{\alpha}_t^{-1} d\bar{S}_t^\delta - \bar{\alpha}_t^{-2} \bar{S}_t^\delta d\bar{\alpha}_t \\
= dt + \sqrt{\bar{\alpha}_t^{-1} \bar{S}_t^\delta} dW_t - \bar{\alpha}_t^{-1} Y_t \times \eta \bar{\alpha}_t \\
= (1 - \eta Y_t) dt + \sqrt{Y_t} dW_t.
\]

Making the substitutions \( c_t = 1, \nu = 4, \beta_t = -\eta, z_0 = \bar{\alpha}_0 \) in Lemma 7 in Appendix A we see that \( Y_t = \bar{\alpha}_0 \exp(-\eta t) X_{\varphi_t} \), where \( X \) is a squared Bessel process of dimension four with \( X_0 = S_0^\delta \) and \( \varphi_t = (\exp(\eta t) - 1)/(4\eta) \).

We observe that the market price of risk \( \theta_t = \sqrt{\bar{\alpha}_t/S_t^\delta} = 1/\sqrt{Y_t} \) is given as the reciprocal of the square root of the normalised discounted GP. Therefore, the squared market price of risk \( v_t = \theta_t^2 = 1/Y_t \) obeys the SDE

\[
(3.5) \quad dv_t = d(Y_t^{-1}) \\
= -Y_t^{-2} dY_t + \frac{1}{2} \times 2Y_t^{-3} d[Y]_t \\
= -(Y_t^{-2} - \eta Y_t^{-1}) dt - Y_t^{-3/2} dW_t + Y_t^{-2} dt \\
= \eta v_t dt - v_t^{3/2} dW_t.
\]

This SDE is known as the \( 3/2 \) volatility model, proposed in Platen [1997]. The leverage effect, as explained in Black [1976], pertains to the multiplied losses associated with severely adverse movements in the stock index. The negative correlation
between the variable $v_t$ and the discounted GP $\tilde{S}_t^\delta$ shows how the leverage effect is naturally incorporated into the MMM.

### 3.1. Transition Density of Discounted GP

Before we give the transition density function of the discounted GP we prove that the time transformed discounted GP is a squared Bessel process of dimension four.

**Lemma 2.** The discounted GP process $\tilde{S}_t^\delta = \{\tilde{S}_t^\delta, t \in [0,T]\}$ given by the SDE (1.7) is a time-transformed squared Bessel process of dimension four.

**Proof.** Define the $(\mathcal{A}, P)$-local martingale $\tilde{U} = \{\tilde{U}_t, t \in [0,T]\}$ as a solution to the SDE

$$d\tilde{U}_t = \sqrt{\tilde{\alpha}_t} d\hat{W}_t$$

for $t \in [0,T]$. Applying the Dambis, Dubins-Schwarz (DDS) Theorem, for example as given in Klebaner [1998], the DDS Wiener process is $\tilde{U}_t = W_{\tilde{U}_t} = W_{\phi_t}$, where $W = \{W_{\phi}, \phi \in [\phi_0, \phi_T]\}$ is a standard Wiener process in $\phi$-time. The time change itself $\phi = \{\phi_t, t \in [\phi_0, \phi_T]\}$ is defined as

$$\phi_t = \phi_0 + \int_0^t \tilde{\alpha}_s ds,$$

or equivalently

$$d\phi_t = \frac{1}{4} \tilde{\alpha}_t dt$$

for $t \in [0,T]$. Therefore, the discounted GP process $X = \{X_{\phi_t}, \phi_t \in [\phi_0, \phi_T]\}$ with time transform of (3.10) is found by setting

$$X_{\phi_t} = \tilde{S}_t^\delta,$$

for $t \in [0,T]$ with the initial condition $X_{\phi_0} = \tilde{S}_{0}^\delta$. Hence the SDE (1.7) can now be written as

$$dX_{\phi} = 4d\phi + \sqrt{4X_{\phi}} d\hat{W}_\phi$$

for $\phi \in [\phi_0, \phi_T]$. It follows from Revuz and Yor [1999] that $X$ is, in the transformed $\phi$-time (3.10), a squared Bessel process of dimension four. \hfill $\square$

This lemma gives rise to the following lemma concerning the conditional distribution of the discounted GP.

**Lemma 3.** For the discounted GP process $\tilde{S}_t^\delta$, satisfying (1.7) and $\phi_t$ given by (3.10) we have for times $T > t$ and given $\tilde{S}_t^\delta$ that

$$\frac{\tilde{S}_T^\delta}{\phi_T - \phi_t}$$

is non-central chi-squared distributed with four degrees of freedom and with non-centrality parameter $\lambda = \tilde{S}_t^\delta/(\phi_T - \phi_t)$, written as

$$\frac{\tilde{S}_T^\delta}{\phi_T - \phi_t} \sim \chi^2_4, \tilde{S}_t^\delta/(\phi_T - \phi_t).$$

The transition density function of the discounted GP is given explicitly in the following lemma.
Lemma 4. Let \( t \) and \( \bar{T} \) be such that \( \bar{T} > t \geq 0 \). The transition density function of the discounted GP in (1.7) is

\[
p_{S^*_t}(t, x_t, \bar{T}, x_{\bar{T}}) = \frac{1}{2(\phi_{\bar{T}} - \phi_t)} \sqrt{\frac{x_{x_{\bar{T}}}}{x_t}} \exp \left( -\frac{x_t + x_{\bar{T}}}{2(\phi_{\bar{T}} - \phi_t)} \right) I_{\frac{1}{2}} \left( \frac{x_{x_{\bar{T}}}}{\phi_{\bar{T}} - \phi_t} \right),
\]

where \( I_{\nu}(z) = \sum_{m=0}^{\infty} \frac{(\frac{1}{2} z)^{\nu + 2m}}{m! \Gamma(\nu + m + 1)} \) is the modified Bessel function of the first kind with index \( \nu \) and \( \phi_t \) is the quadratic variation of \( \sqrt{S^*_t} \) as given by (3.10).

Remark 3. From (1.7) the logarithm of the GP has the SDE

\[
d\log (\bar{S}^*_t) = \frac{1}{2} Y_t dt + \sqrt{1 \cdot Y_t} dW_t,
\]

where \( Y_t \) obeys SDE (3.4). The stationary density \( \bar{p} \) of \( Y_t \) is that of a scaled chi-square distribution and is given by

\[
\bar{p}(y) = \frac{(2\eta)^2}{\Gamma(2)} y \exp(-2\eta y),
\]

(see, for example, Chapter 4 of Platen and Heath [2006]). Therefore, under the MMM, the distribution of log returns of the discounted GP is Student-t with four degrees of freedom. This contrasts with the corresponding distribution under the BS model being Gaussian. These stationary densities are shown in Figure 6. Also shown, is the transition density of log returns of MMM discounted GP based at January 1871 which demonstrates the ability of the MMM to capture asymmetry in log returns.

The graph of the transition density function of the discounted GP with parameter values \( \tilde{\alpha}_0 = 0.0068370 \) and \( \eta = 0.045486 \) is shown in Figure 7.

3.2. Fitting the MMM Discounted GP. Estimation of the parameters of the MMM discounted GP is achieved using maximum likelihood estimation.

The log-likelihood function in respect of the observed values of the discounted GP \( \bar{S}^*_t \), \( i = 0, 1, 2, \ldots, n \), is

\[
\ell(\tilde{\alpha}_0, \eta) = \sum_{i=1}^{n} \log \left[ \frac{1}{2(\phi_{t_i} - \phi_{t_{i-1}})} \sqrt{\frac{\bar{S}^*_k}{\bar{S}^*_{t_{i-1}}}} \right. \\
\left. \times \exp \left\{ -\frac{1}{2} \left( \frac{\bar{S}^*_{t_i} + \bar{S}^*_{t_{i-1}}}{\phi_{t_i} - \phi_{t_{i-1}}} \right) \right\} I_{\frac{1}{2}} \left( \sqrt{\bar{S}^*_k \bar{S}^*_{t_{i-1}}} / (\phi_{t_i} - \phi_{t_{i-1}}) \right) \right] ,
\]

where \( \phi_t = \frac{1}{2} \tilde{\alpha}_0 (\exp(\eta t) - 1) / \eta \).

The following theorem supplies an accurate approximation to the maximum likelihood estimates as well as an explicit formula for the standard errors of the parameter estimates.

Theorem 2. Assume that the series

\[
\left( \log \left\{ \left( \sqrt{S^*_i} - \sqrt{S^*_{i-1}} \right)^2 \right\} \right)_{i=1}^{n}
\]

is
Figure 6. Graphs of empirical, normal and Student-t stationary density functions as well as the transition density of log returns of MMM discounted GP based at January 1871.

is approximately linear in $t_i$, for $i = 1, 2, \ldots, n$, and that the resulting residuals $\epsilon_i$ of the linear fit have sufficiently small moments, that is

\begin{equation}
\frac{1}{n} \sum_{i=1}^{n} \left( \exp(\epsilon_i) - \epsilon_i - 1 \right) \leq \frac{K_3}{n},
\end{equation}

where

\begin{equation}
\epsilon_i = \log \left\{ \sqrt{S_{t_i}^{\delta_*}} - \sqrt{S_{t_{i-1}}^{\delta_*}} \right\} - K_1 - K_2 t_i
\end{equation}

and where $K_1$, $K_2$ and $K_3$ are positive constants independent of $n$. The maximum likelihood estimates $\hat{\alpha}_{0,\Delta}$ and $\hat{\eta}_{\Delta}$ of $\alpha_0$ and $\eta$ respectively in the SDE (1.7) with time step size $\Delta$ are given by

\begin{align}
\hat{\alpha}_{0,\Delta} & \approx \frac{1}{n} \sum_{i=1}^{n} \frac{1}{4\eta} \left( \exp(\eta \Delta) - 1 \right) \exp(\eta(i-1)\Delta) \\
\hat{\eta}_{\Delta} & \approx \frac{6}{\Delta(n+1)} \left[ \frac{1}{n} \sum_{i \in K} (i-1) \log \left\{ \left( \sqrt{S_{t_i}^{\delta_*}} - \sqrt{S_{t_{i-1}}^{\delta_*}} \right)^2 \right\} \\
& - \frac{1}{n} \sum_{i \in K} \log \left\{ \left( \sqrt{S_{t_i}^{\delta_*}} - \sqrt{S_{t_{i-1}}^{\delta_*}} \right)^2 \right\} \right].
\end{align}
Figure 7. Transition density function of MMM discounted GP based at January 1871, with S&P500 trajectory and theoretical mean.

where summation is over the set $K = \{i = 1, 2, \ldots, n : \bar{S}_{t_i}^{\delta_t} \neq \bar{S}_{t_{i-1}}^{\delta_t}\}$ and where there are $n+1$ observations of the discounted GP $\bar{S}_{t_i}^{\delta_t}$, for $i = 0, 1, 2, \ldots, n$. Further, the standard errors of the parameter estimates are approximately given by

$SE(\hat{\alpha}_0, \Delta) \approx 2\hat{\alpha}_0, \Delta \sqrt{\frac{2n - 1}{n(n + 1)}}$  

$SE(\hat{\eta}, \Delta) \approx \sqrt{\frac{24}{n(n^2 - 1)}\Delta^2}$.

Proof. Firstly, we rewrite the log likelihood function in (3.19) in terms of $\eta$ and a new parameter $a = \frac{1}{\alpha_0} \alpha_0 (\exp(\eta \Delta) - 1)$, giving

$\ell(a, \eta) = -n \log 2 - n \log a - \eta \left(\frac{n}{2}\right) + \frac{1}{2} \log \left\{\frac{\bar{S}_{t_i}^{\delta_t}}{\bar{S}_{t_0}^{\delta_t}}\right\}$

$- \frac{1}{2} \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} f(z_i)$,
where we have used the notation

\begin{equation}
\tag{3.26}
y_i = \frac{\left( \sqrt{S_{t_i}^0} - \sqrt{S_{t_{i-1}}^0} \right)^2}{\varphi_{t_i} - \varphi_{t_{i-1}}} - \frac{\left( \sqrt{S_{t_i}^0} - \sqrt{S_{t_{i-1}}^0} \right)^2}{a \exp(-\eta(i-1)\Delta)} - \frac{\left( \sqrt{S_{t_i}^0} - \sqrt{S_{t_{i-1}}^0} \right)^2}{a \exp(-\eta(i-1)\Delta)}
\end{equation}

and where the function \( f \) is defined as

\begin{equation}
\tag{3.27}
f(x) = \log\{\exp(-x)I_1(x)\},
\end{equation}

which has convenient asymptotic properties as \( x \to \infty \), demonstrated in Appendix B. We straightforwardly obtain the following first order partial derivatives with respect to the parameters \( a \) and \( \eta \), that is,

\begin{equation}
\tag{3.28}
\frac{\partial \ell}{\partial a} = -\frac{n}{a} + \frac{1}{2a} \sum_{i=1}^{n} y_i - \frac{1}{a} \sum_{i=1}^{n} f'(z_i) z_i
\end{equation}

\begin{equation}
\frac{\partial \ell}{\partial \eta} = -\left( \frac{n}{2} \right) \Delta + \frac{\Delta}{2} \sum_{i=1}^{n} y_i (i - 1) - \Delta \sum_{i=1}^{n} f'(z_i) z_i (i - 1)
\end{equation}

from which we obtain the second order partial derivatives

\begin{equation}
\tag{3.29}
\frac{\partial^2 \ell}{\partial a^2} = \frac{n}{a^2} - \frac{1}{a^2} \sum_{i=1}^{n} y_i + \frac{1}{a^2} \sum_{i=1}^{n} \left\{ 2f'(z_i) z_i + f''(z_i) z_i^2 \right\}
\end{equation}

\begin{equation}
\frac{\partial^2 \ell}{\partial \eta \partial a} = -\frac{\Delta}{2a} \sum_{i=1}^{n} y_i (i - 1) + \frac{\Delta}{a} \sum_{i=1}^{n} \left\{ f'(z_i) z_i (i - 1) + f''(z_i) z_i^2 (i - 1) \right\}
\end{equation}

\begin{equation}
\frac{\partial^2 \ell}{\partial \eta^2} = -\frac{\Delta^2}{2} \sum_{i=1}^{n} y_i (i - 1)^2 + \Delta^2 \sum_{i=1}^{n} \left\{ f'(z_i) z_i (i - 1)^2 + f''(z_i) z_i^2 (i - 1)^2 \right\}.
\end{equation}

Note that, from Appendix B, we have as \( y \to \infty \)

\begin{equation}
\tag{3.30}
yf'(y) = -\frac{1}{2} + \frac{3}{8y} + O\left( \frac{1}{y^2} \right)
\end{equation}

\begin{equation}
yf''(y) = \frac{1}{2} - \frac{3}{4y} + O\left( \frac{1}{y^2} \right).
\end{equation}

Inserting these asymptotic estimates into (3.28) we obtain

\begin{equation}
\tag{3.31}
\frac{a}{n} \frac{\partial \ell}{\partial a} = -1 + \frac{1}{2a} \sum_{i=1}^{n} y_i - \frac{1}{n} \sum_{i=1}^{n} \left\{ -\frac{1}{2} + O\left( \frac{1}{y} \right) \right\}
\end{equation}

\begin{equation}
= -1 + \frac{1}{2a} \sum_{i=1}^{n} y_i + O(\Delta)
\end{equation}

\begin{equation}
\frac{1}{\binom{2}{2} \Delta} \frac{\partial \ell}{\partial \eta} = -1 + \frac{1}{2} \sum_{i=1}^{n} y_i (i - 1) - \frac{1}{n} \sum_{i=1}^{n} \left\{ -\frac{1}{2} + O\left( \frac{1}{y} \right) \right\} (i - 1)
\end{equation}

\begin{equation}
= -1 + \frac{1}{2} \sum_{i=1}^{n} y_i (i - 1) + O(\Delta),
\end{equation}
Here we have employed the two approximations \( \tilde{S}_t^\delta \approx \tilde{\alpha}_0 \exp(\eta i \Delta) \) and
\[
(3.32) \quad z_i \approx \frac{\tilde{\alpha}_0 \exp(\eta (i - 1/2) \Delta)}{\frac{1}{4\eta} \tilde{\alpha}_0 \exp(\eta \Delta) - 1 \exp(\eta (i - 1) \Delta)} \\
\approx \frac{4\eta \exp(\eta \Delta/2)}{\eta \Delta + O((\eta \Delta)^2)} \\
\approx \frac{4}{\Delta} (1 + O(\Delta))
\]
to arrive at \( 1/z_i = O(\Delta) \). Our maximum likelihood equations become
\[
(3.33) \quad \frac{1}{n} \sum_{i=1}^{n} y_i = 1 + O(\Delta) \\
\frac{1}{\binom{n}{2}} \sum_{i=1}^{n} y_i (i - 1) = 1 + O(\Delta).
\]
The solutions to these equations are close to the exact ones by virtue of the smallness of \( \Delta \) for large values of \( n \). If the series \( (y_i)_{i=1}^{n} \) is nearly constant, as per our assumption, Jensen’s inequality \( E[\log X] \leq \log E[X] \) is nearly strict so that we have the approximations
\[
(3.34) \quad \frac{1}{n} \sum_{i \in K} \log(y_i) \approx \log(1 + O(\Delta)) \\
\frac{1}{\binom{n}{2}} \sum_{i \in K} \log(y_i) (i - 1) \approx \log(1 + O(\Delta)),
\]
where summation is over those \( i \) in the set \( K = \{ i : y_i > 0 \} \). The solutions to these equations are straightforwardly shown to be those given in (3.23). Using the approximations (3.31) in (3.29) we obtain the simplified expressions
\[
(3.35) \quad \frac{\partial^2 \ell}{\partial a^2} \approx -\frac{1}{2} + O(\Delta) \\
\frac{\partial^2 \ell}{\partial a \partial \eta a} \approx -\frac{1}{2} + O(\Delta) \\
\frac{1}{n(n-1)(2n-1)\Delta^2} \frac{\partial^2 \ell}{\partial \eta^2} \approx -\frac{1}{2} + O(\Delta).
\]
Fisher’s information matrix is computed to be
\[
(3.36) \quad -\left( \frac{\partial^2 \ell}{\partial a^2} \frac{\partial^2 \ell}{\partial \eta a} \right)^{-1} \approx \frac{24\alpha^2}{n^2(n^2 - 1)\Delta^2} \left( \frac{n(n-1)(2n-1)\Delta}{\binom{n}{2}} \frac{\Delta^2}{\alpha} - \frac{n}{\binom{n}{2}} \frac{\Delta}{\alpha} \right)
\]
and the standard errors of \( \hat{\alpha} \) and \( \hat{\eta} \Delta \) are
\[
(3.37) \quad SE(\hat{\alpha}) \approx 2\hat{\alpha} \sqrt{\frac{2n-1}{n(n+1)}} \\
SE(\hat{\eta} \Delta) \approx \sqrt{\frac{24}{n(n^2 - 1)\Delta^2}}.
\]
Because \( a = \frac{1}{4\eta} \hat{\alpha}_0 (\exp(\eta \Delta) - 1) \approx \frac{1}{4} \hat{\alpha}_0 (1 + O(\Delta)) \) we have

\[
(3.38) \quad SE(\hat{\alpha}_0, \Delta) \approx 4 \Delta SE(\hat{a}) \approx \frac{8 \hat{a}}{\Delta} \sqrt{\frac{2n - 1}{n(n + 1)}} \approx 2 \hat{\alpha}_0, \Delta \sqrt{\frac{2n - 1}{n(n + 1)}}.
\]

We remark that the assumptions (3.21) and (3.22) underlying the maximum likelihood estimation in Theorem 2 are likely to be satisfied in empirical applications where the quadratic variation of the square root of elements of the time series has approximate logarithmic growth, as posited for the MMM discounted GP, and where this approximation is sufficiently good, as specified in (3.21). For an arbitrary series, a suitable transformation may need to be applied so that the transformed series satisfies the assumptions (3.21) and (3.22).

For Shiller’s monthly data set of US one-year deposit rates and stock index values from 1871 to 2017, given in Appendix C, the following estimates are obtained by applying the Newton-Raphson root-finding method to the first-order partial derivatives of the log-likelihood function,

\[
(3.39) \quad \hat{\alpha}_0 = 0.006837 \ (0.000462), \quad \eta = 0.045486 \ (0.000800),
\]

where the standard errors are shown in brackets.

We note that the estimate of the net growth rate \( \eta \) is close to the empirical finding of \( \eta \approx 0.06 \) derived from the estimates 10.1% and 4.1% of annualised returns for the last century of equities and short-dated treasury bills respectively, given in Chapter 16 of Dimson et al. [2002].

The quadratic variation of \( \sqrt{S_{\delta t}} \) is

\[
(3.40) \quad \left[ \sqrt{S_{\delta t}} \right]_t = \frac{1}{4} \int_0^t \hat{\alpha}_s ds = \frac{1}{4\eta} \hat{\alpha}_0 (\exp(\eta t) - 1).
\]

We can visually assess the accuracy of the two parameters \( \hat{\alpha}_0 \) and \( \eta \) by comparing the theoretical quadratic variation of the square root of the discounted GP, namely \( \frac{1}{4\eta} \hat{\alpha}_0 (\exp(\eta t) - 1) \), with the quadratic variation of \( \sqrt{S_{\delta t}} \). The graphs are shown in Figure 8.

**Remark 4.** Similar to the parameter estimation method employed in Section 4 of Hulley and Platen [2012], a crude estimate of \( \hat{\alpha}_0 \) and \( \eta \) can be obtained by solving the two equations \( \hat{\varphi}_n - \varphi_0 = \hat{\alpha}_0 (\exp(\eta t_n) - 1)/4n \) \( = QV_{t_n} \) and \( \hat{\varphi}_{n/2} - \varphi_0 = \hat{\alpha}_0 (\exp(\eta t_{n/2}) - 1)/4n \) \( = QV_{t_{n/2}} \) for \( \eta \), giving \( \exp(\eta t_{n/2}) - 1 = QV_{t_{n/2}}/QV_{t_n} \) and hence \( \eta = 0.041917 \), with log-likelihood \(-957.907584\). The estimate of \( \eta \) provided by Theorem 2 is \( \hat{\eta} = 0.046020 \), with log-likelihood \(-943.885294\) which is very close to the optimum in (3.39).

### 3.3. Fitting the MMM Discounted GP using a Lognormal Approximation.

We will see that good estimates of the parameters of the MMM discounted GP can be obtained by approximating the non-central chi-squared distribution by a lognormal distribution having the same mean and variance. The following lemma supplies formulae for the conditional mean and variance of the discounted GP.
Lemma 5. For the discounted GP obeying the SDE (1.7) and for times \( s, t \) such that \( s < t \) we have

\[
\begin{align*}
    m_s(t) &= E(S_t^\delta | A_s) = \bar{\bar{S}}_s^\delta + \frac{\bar{\alpha}_0}{\eta} (\exp(\eta t) - \exp(\eta s)) \\
v_s(t) &= \text{VAR}(S_t^\delta | A_s) = \bar{\bar{S}}_s^\delta \frac{\bar{\alpha}_0}{\eta} (\exp(\eta t) - \exp(\eta s)) \\
    &\quad + \frac{\bar{\alpha}_0^2}{2\eta^2} (\exp(\eta t) - \exp(\eta s))^2.
\end{align*}
\]

Proof. Integrating the SDE (1.7) from time \( s \) to time \( t \) and taking expectations conditional on \( S_s^\delta \) gives

\[
\begin{align*}
    E(S_t^\delta | A_s) &= \bar{\bar{S}}_s^\delta + \int_s^t \tilde{\alpha}_u \, du \\
    &= \bar{\bar{S}}_s^\delta + \bar{\alpha}_0 \int_s^t \exp(\eta u) \, du \\
    &= \bar{\bar{S}}_s^\delta + \frac{\bar{\alpha}_0}{\eta} (\exp(\eta t) - \exp(\eta s)),
\end{align*}
\]
which is the first equation. The SDE for \((\bar{S}_u^\delta)^2\) is, by Ito’s Lemma,

\[
(3.43) \quad d(\bar{S}_u^\delta)^2 = 2\bar{S}_u^\delta \cdot d\bar{S}_u^\delta + \frac{1}{2} \times 2 \times d[\bar{S}_u^\delta]_u \\
= 2\bar{S}_u^\delta \cdot \bar{\alpha}_u du + 2\sqrt{\bar{\alpha}_u(\bar{S}_u^\delta)^3} dW_u + \bar{\alpha}_u \bar{S}_u^\delta du \\
= 3\bar{S}_u^\delta \cdot \bar{\alpha}_u du + 2\sqrt{\bar{\alpha}_u(\bar{S}_u^\delta)^3} dW_u.
\]

As done for \(\tilde{S}_u^\delta\), we integrate this SDE for \((\bar{S}_u^\delta)^2\) from time \(s\) to time \(t\) and take expectations conditional on \(\bar{S}_u^\delta_s\). By noticing that the integral with respect to the Wiener process is a martingale, we obtain

\[
(3.44) \quad E((\bar{S}_t^\delta)^2 | A_s) = (\bar{S}_s^\delta)^2 + \int_s^t 3\bar{\alpha}_u E(\bar{S}_u^\delta | A_s) du \\
= (\bar{S}_s^\delta)^2 + \int_s^t 3\bar{\alpha}_u \bar{S}_u^\delta du + \int_s^t 3\bar{\alpha}_u \frac{\bar{\alpha}_0}{\eta} (\exp(\eta u) - \exp(\eta s)) du \\
= (\bar{S}_s^\delta)^2 + (3\bar{S}_s^\delta - 3\frac{\bar{\alpha}_0}{\eta} \exp(\eta s)) \int_s^t \bar{\alpha}_u du + 3\frac{\bar{\alpha}_0}{\eta} \int_s^t \bar{\alpha}_u \exp(\eta u) du \\
= (\bar{S}_s^\delta)^2 + (3\bar{S}_s^\delta - 3\frac{\bar{\alpha}_0}{\eta} \exp(\eta s)) \frac{\bar{\alpha}_0}{\eta} (\exp(\eta t) - \exp(\eta s)) \\
+ 3\frac{\bar{\alpha}_0^2}{2\eta^2} (\exp(2\eta t) - \exp(2\eta s)) \\
= (\bar{S}_s^\delta)^2 + 3\bar{\alpha}_s^\delta \frac{\bar{\alpha}_0}{\eta} (\exp(\eta t) - \exp(\eta s)) \\
+ 3\frac{\bar{\alpha}_0^2}{2\eta^2} (\exp(\eta t) - \exp(\eta s))^2.
\]

The formula for the variance is straightforwardly given by

\[
(3.45) \quad VAR(\bar{S}_t^\delta | A_s) = E((\bar{S}_s^\delta)^2 | A_s) - (E(\bar{S}_s^\delta | A_s))^2 \\
= \bar{S}_s^\delta \frac{\bar{\alpha}_0}{\eta} (\exp(\eta t) - \exp(\eta s)) \\
+ \frac{\bar{\alpha}_0^2}{2\eta^2} (\exp(\eta t) - \exp(\eta s))^2,
\]

which is the second equation. □

We approximate the transition density of the square root process from time \(t_{i-1}\) to time \(t_i\) by a lognormal distribution that matches the mean and variance. The lognormal distribution for the exponential of a Gaussian random walk with mean \(\mu\) and variance \(\sigma^2\) has mean

\[
(3.46) \quad m = \exp(\mu + \frac{1}{2}\sigma^2)
\]

and variance

\[
(3.47) \quad v = \exp(2\mu + \sigma^2) (\exp(\sigma^2) - 1).
\]
It is straightforward to show that the approximating lognormal distribution has, according to (3.46) and (3.47), parameters $\mu_{t_i}^{(LN)}$, $\sigma_{t_i}^{(LN)}$ given by

\begin{align}
(3.48) \quad \mu_{t_i}^{(LN)} &= \log (m_{t_{i-1}}(t_i)) - \frac{1}{2}(\sigma_{t_i}^{(LN)})^2 \\
(\sigma_{t_i}^{(LN)})^2 &= \log \left(1 + \frac{v_{t_{i-1}}(t_i)}{m_{t_{i-1}}(t_i)^2}\right),
\end{align}

respectively, where

\begin{align}
(3.49) \quad m_{t_{i-1}}(t_i) &= \bar{S}_{t_{i-1}}^{\delta_{t_i}} + \frac{\bar{S}_{t_{i-1}}^{\delta_{t_i}}}{\eta} \left(\exp(\eta t_i) - \exp(\eta t_{i-1})\right) \\
v_{t_{i-1}}(t_i) &= \bar{S}_{t_{i-1}}^{\delta_{t_i}} - \frac{\bar{S}_{t_{i-1}}^{\delta_{t_i}}}{\eta} \left(\exp(\eta t_i) - \exp(\eta t_{i-1})\right) \\
&+ \frac{\bar{S}_{t_{i-1}}^{\delta_{t_i}}}{2\eta^2} \left(\exp(\eta t_i) - \exp(\eta t_{i-1})\right)^2.
\end{align}

Therefore, our approximating log-likelihood function on the set of observed values of the discounted GP $\bar{S}_{t_i}^{\delta_{t_i}}$, for $i = 0, 2, \ldots, n$ is

\begin{align}
(3.50) \quad \ell(\bar{S}_{t_i}^{\delta_{t_i}}, \eta) &= -\frac{1}{2} \sum_{i=1}^{n} \left\{ \log (2\pi(\sigma_{t_i}^{(LN)})^2) + 2 \log \left(\frac{\log(\bar{S}_{t_i}^{\delta_{t_i}} - \mu_{t_i}^{(LN)})}{(\sigma_{t_i}^{(LN)})^2}\right) \right\}
\end{align}

**Remark 5.** The skew of the lognormal distribution is

\begin{align}
(3.51) \quad m_{t_{i-1}}(t_i)^3 (\exp((\sigma_{t_i}^{(LN)})^2) - 1)^2 \left(\exp((\sigma_{t_i}^{(LN)})^2) + 2\right) \\
&\approx 16(\phi_{t_i} - \phi_{t_{i-1}})^2 m_{t_{i-1}}(t_i) (1 + O((\phi_{t_i} - \phi_{t_{i-1}})/m_{t_{i-1}}(t_i)))
\end{align}

and the skew of the noncentral chi-squared distribution is

\begin{align}
(3.52) \quad -64(\phi_{t_i} - \phi_{t_{i-1}})^3 + 24(\phi_{t_i} - \phi_{t_{i-1}})^2 m_{t_{i-1}}(t_i) \\
&\approx 24(\phi_{t_i} - \phi_{t_{i-1}})^2 m_{t_{i-1}}(t_i) (1 + O((\phi_{t_i} - \phi_{t_{i-1}})/m_{t_{i-1}}(t_i)))
\end{align}

we see that the lognormal approximation accounts for roughly two-thirds of the skew present in the noncentral chi-squared distribution which is not captured by the normal approximation, for example.

We fit the MMM to the monthly discounted GP series derived from Shiller’s data set given in Appendix C, obtaining the maximum likelihood estimates

\begin{align}
(3.53) \quad \bar{\alpha}_0 &= 0.006904 (0.000466) \\
\eta &= 0.045555 (0.000800).
\end{align}

We note that the estimates for $\bar{\alpha}_0$ and $\eta$ are in close agreement with those in (3.39).

### 4. Comparison of Models

The two models considered in this chapter have explicit formulae for their transition density functions and this has allowed the fitting of parameters using maximum likelihood estimation. The Black-Scholes model is most easily fitted to the data because it has a closed form expression for its parameter estimate. In contrast, the MMM requires two-dimensional grid searches to find the best fitting parameters.

In fitting the two models to the US discounted GP data we can identify which model provides the best fit to the data by looking at the Akaike Information Criterion, shown in Table 1, where the MMM appears to be the best fitting model.
Table 1. Values of the AIC in respect of the discounted GP models (Shiller’s monthly US data set).

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Log Likelihood</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black-Scholes</td>
<td>$\theta = 0.141002$</td>
<td>$-1060.320193$</td>
<td>$2122.640385$</td>
</tr>
<tr>
<td></td>
<td>$SE(\theta) = 0.002380$</td>
<td>$-1060.320193$</td>
<td>$2122.640385$</td>
</tr>
<tr>
<td>MMM</td>
<td>$\bar{\alpha}_0 = 0.006837$</td>
<td>$-943.670030$</td>
<td>$1891.340059$</td>
</tr>
<tr>
<td></td>
<td>$SE(\bar{\alpha}_0) = 0.000462$</td>
<td>$-943.670030$</td>
<td>$1891.340059$</td>
</tr>
<tr>
<td></td>
<td>$\eta = 0.045486$</td>
<td>$-943.670030$</td>
<td>$1891.340059$</td>
</tr>
<tr>
<td></td>
<td>$SE(\eta) = 0.000800$</td>
<td>$-943.670030$</td>
<td>$1891.340059$</td>
</tr>
</tbody>
</table>

Figure 9. Comparison of quantile-quantile plots of discounted GP models (Shiller’s monthly US data set).

To establish whether the MMM is a good fitting model we consider Pearson’s goodness-of-fit chi-squared statistic, described in Kendall and Stuart [1961].

Given a time series of discounted GP values $\{\bar{S}_t^\alpha : j = 1, 2, \ldots, n\}$ and a hypothesised transition density function with corresponding cumulative distribution function $F$ we compute the $n-1$ quantiles $q_j = F(t_{j-1}, \bar{S}_{t_{j-1}}^\alpha, t_j, \bar{S}_t^\alpha)$ for $j = 2, 3, \ldots, n$. Under the hypothesised model the quantiles $q_j$ are independent and uniformly distributed. These quantiles are graphed against those of the uniform distribution in Figure 9. One notes that the MMM model remains in some sense visually closest over the $[0, 1]$ interval.

A similar comparison is shown in Figure 10 for the daily data series of the S&P500 total return index values and Federal Funds Rates from January 1970 to
Figure 10. Comparison of quantile-quantile plots of discounted GP models (US daily data from Bloomberg).

May 2017, sourced from the Bloomberg data services and shown in Appendix D, where a similar conclusion follows.

For a fixed integer \( k \) satisfying \( 2 \leq k \leq (n - 1)/5 \) we partition the unit interval into \( k \) equally sized subintervals. Hence we compute the number of observations \( O_i \) in the \( i \)-th subinterval \( ((i - 1)/k, i/k] \) for \( i = 1, 2, \ldots, k \). The corresponding expected number of observations \( E_i \) in the \( i \)-th subinterval is \( (n - 1)/k \). Our test statistic is thus computed as

\[
S = k \sum_{i=1}^{k} \frac{(O_i - (n - 1)/k)^2}{(n - 1)}
\]

which is approximately chi-squared distributed with \( \nu = k - 1 - n_{\text{parameters}} \) degrees of freedom.

The value of Pearson’s chi-squared statistic and corresponding p-value for each model and for a range of partition sizes is shown in Table 2. It is evident that both the BS model and MMM can be rejected at the 1% level of significance.

Another test of goodness-of-fit is the Kolmogorov-Smirnov test, as described by Stephens [1974]. Under the null hypothesis that the set of \( n \) observations \( u_1, u_2, \ldots, u_n \) emanate from a uniform distribution, the Kolmogorov test statistic is

\[
D_n = \sup_{x \in \{u_1, u_2, \ldots, u_n\}} \max \left( F^{(n)}(x) - x, x - F^{(n)}(x) - \frac{1}{n} \right)
\]
Table 2. Pearson’s chi-squared statistics and p-values in respect of various discounted GP models (Shiller’s monthly US data set).

<table>
<thead>
<tr>
<th>k</th>
<th>BS</th>
<th>ν</th>
<th>p-value</th>
<th>MMM</th>
<th>ν</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>128.9133</td>
<td>3</td>
<td>1.1768E-14</td>
<td>93.7822</td>
<td>2</td>
<td>1.5767E-10</td>
</tr>
<tr>
<td>10</td>
<td>147.8016</td>
<td>8</td>
<td>0.0000E0</td>
<td>113.7537</td>
<td>7</td>
<td>0.0000E0</td>
</tr>
<tr>
<td>15</td>
<td>165.1619</td>
<td>13</td>
<td>0.0000E0</td>
<td>116.1425</td>
<td>12</td>
<td>0.0000E0</td>
</tr>
<tr>
<td>20</td>
<td>179.3181</td>
<td>18</td>
<td>0.0000E0</td>
<td>130.9715</td>
<td>17</td>
<td>0.0000E0</td>
</tr>
<tr>
<td>25</td>
<td>180.7206</td>
<td>23</td>
<td>0.0000E0</td>
<td>130.4356</td>
<td>22</td>
<td>0.0000E0</td>
</tr>
</tbody>
</table>

Table 3. Kolmogorov-Smirnov test statistics in respect of various discounted GP models (Shiller’s monthly US data set).

<table>
<thead>
<tr>
<th>BS</th>
<th>MMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dn</td>
<td>0.096819</td>
</tr>
<tr>
<td>n</td>
<td>1754</td>
</tr>
<tr>
<td>Kn</td>
<td>4.054863</td>
</tr>
<tr>
<td>F(Kn)</td>
<td>1.000000</td>
</tr>
<tr>
<td>p-value</td>
<td>1.0436E-14</td>
</tr>
</tbody>
</table>

and the modified test statistic \( K_n = \sqrt{n}D_n \) has the limiting distribution function, as \( n \to \infty \),

\[
F(x) = \frac{\sqrt{2\pi}}{x} \sum_{k=1}^{\infty} \exp(-(2k-1)^2\pi^2/(8x^2)),
\]

where

\[
F^{(n)}(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{u_i \leq x}
\]

is the empirical cumulative distribution function. We compute the test statistics in Table 3 where we see both models can be rejected at the 1% level of significance.

Finally, another test of goodness-of-fit is the Anderson-Darling test, as described in Stephens [1974]. Under the null hypothesis that the set of \( n \) observations \( u_1 \leq u_2 \leq \ldots \leq u_n \) emanate from a uniform distribution, the test statistic \( A \) is given by

\[
A = \sqrt{-n - S},
\]

where

\[
S = \sum_{i=1}^{n} \frac{2i-1}{n} \left( \log(u_i) + \log(1 - u_{n+1-i}) \right).
\]

We compute the test statistics in Table 4 where, as for the Kolmogorov-Smirnov test, we see that both of the discounted GP models can be rejected at the 1% level of significance. The p-values of the test statistic \( A \) in Table 4 were estimated using sample Anderson-Darling statistics of 1,000,000 simulations of sets of 1754 uniformly distributed observations.

5. Conclusions

In this article we have demonstrated the applicability of maximum likelihood estimation of parameters of discounted equity index models, giving explicit formulae
Table 4. Anderson-Darling test statistics in respect of various discounted GP models (Shiller’s monthly US data set).

<table>
<thead>
<tr>
<th></th>
<th>BS</th>
<th>MMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>-1783.440143</td>
<td>-1774.793996</td>
</tr>
<tr>
<td>$n$</td>
<td>1754</td>
<td>1754</td>
</tr>
<tr>
<td>$A^2 = -n - S$</td>
<td>29.440143</td>
<td>20.793996</td>
</tr>
<tr>
<td>$A$</td>
<td>5.425877</td>
<td>4.560043</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.0000E0</td>
<td>5.0092E-10</td>
</tr>
</tbody>
</table>

for maximum likelihood estimates of model parameters. The model parameters fitted to the US data set have values consistent with estimates obtained by others, for example Dimson et al. [2002]. Also, we have demonstrated several ways of assessing the goodness-of-fits of the models.

As mentioned in Remark 2, the log-returns of the discounted GP under the Black-Scholes model are Gaussian with constant variance, whereas under the MMM they are Student-t distributed. Empirical evidence in Fergusson and Platen [2006] and in Platen and Rendek [2008] indicates that the log returns of discounted equity indices are Student-t distributed with four degrees of freedom. This also shows that the MMM is a more realistic model that appears to capture better the stochastic volatility and leptokurtic log-returns observed in the market than the classical Black-Scholes model.

As mentioned in Remark 3, for the Black-Scholes model, the Radon-Nikodym derivative is a martingale and therefore risk neutral pricing of contingent claims is possible. Importantly, however, the Radon-Nikodym derivative pertaining to the MMM is a strict supermartingale and therefore the MMM does not admit an equivalent risk neutral probability measure. Thus, for the MMM we must resort to another valuation framework, such as the Benchmark Approach, where some arbitrage is permitted.

Acknowledgments

The author is thankful for the support of Professor Eckhard Platen. This research is supported by an Australian Government Research Training Program Scholarship.

Appendix A. Squared Bessel Processes

A squared Bessel process $X$ of dimension $\nu$ is a stochastic process defined by

$$X_t = \sum_{j=1}^{\nu} (\lambda^{(j)} + Z_t^{(j)})^2$$  \hspace{1cm} (A.1)

for $t \geq 0$ and for a fixed integer $\nu \geq 0$.

We first state some lemmas which lead to a solution to the SDE (1.7), the proofs of which can be found in Platen and Heath [2006].

Lemma 6. Let the stochastic process $X$ be defined by (A.1) for $\nu > 2$. Then $X$ has the SDE

$$dX_t = \nu dt + 2\sqrt{X_t} dZ_t$$  \hspace{1cm} (A.2)

with initial condition $X_0 = \sum_{j=1}^{\nu} (\lambda^{(j)})^2$. 

Lemma 7. Let $Y$ be the process defined by
\begin{equation}
Y_t = z_t X_{\varphi_t}
\end{equation}
for $t \geq 0$, where $X$ is as in (A.2), $z_t = z_0 \exp\left(\int_0^t b_u du\right)$ and $\varphi_t = \varphi_0 + \frac{1}{4} \int_0^t c_u^2 z_u du$, for deterministic functions $b$ and $c$. Then $Y$ has the SDE
\begin{equation}
dY_t = \left(\nu c_t^2 + b_t Y_t\right) dt + c_t \sqrt{Y_t} dU_t.
\end{equation}

For completeness we give here the proof.

Proof. From (A.3) we have the SDE
\begin{equation}
dY_t = X_{\varphi_t} dz_t + z_t dX_{\varphi_t}
\end{equation}
\begin{align*}
&= \frac{Y_t}{z_t} dz_t + z_t (\nu d\varphi_t + 2 \sqrt{X_{\varphi_t}} dZ_{\varphi_t}) \\
&= Y_t b_t dt + z_t (\nu c_t^2 dt + 2 \sqrt{X_{\varphi_t}} dZ_{\varphi_t}) \\
&= (Y_t b_t + \nu c_t^2) dt + 2 z_t \sqrt{X_{\varphi_t}} dZ_{\varphi_t}.
\end{align*}

We define the process $U_t = \int_0^t \frac{1}{\sqrt{\varphi_u}} dZ_{\varphi_u}$ and note that $d[U]_t = dt$, which means that $U$ is a standard Wiener process. Here $\varphi'(t) = \frac{c_t^2}{4t^3}$ and so
\begin{equation}
dU_t = \frac{1}{\sqrt{\varphi_t}} dZ_{\varphi_t} = \frac{2 \sqrt{z_t}}{c_t} dZ_{\varphi_t}.
\end{equation}

Hence
\begin{equation}
dY_t = \left(\nu c_t^2 + b_t\right) dt + c_t \sqrt{Y_t} dU_t
\end{equation}
which proves (A.4).

\bigskip

\textbf{Appendix B. Asymptotic Properties of Functions Involving the Modified Bessel Function of the First Kind}

We give some asymptotic properties of the function $f(x) = \log\{\exp(-x)I_1(x)\}$ given in (3.27).

Lemma 8. For the function $f$ in (3.27) we have the asymptotic formulae
\begin{equation}
f'(x) = -\frac{1}{2x} + \frac{3}{8x^2} + \frac{3}{8x^3} + O\left(\frac{1}{x^4}\right)
\end{equation}
\begin{equation}
f''(x) = \frac{1}{2x^2} - \frac{3}{4x^3} + O\left(\frac{1}{x^4}\right)
\end{equation}

Proof. As given in Chapter 6 of Watson [1966], the modified Bessel function of the first kind can be written as
\begin{equation}
I_\nu(x) = \frac{1}{\pi} \int_0^\pi \exp(x \cos(t)) \cos(\nu t) dt - \frac{\sin \nu \pi}{\pi} \int_0^\infty \exp(-x \cosh t - \nu t) dt,
\end{equation}
which simplifies to
\begin{equation}
I_1(x) = \frac{1}{\pi} \int_0^\pi \exp(x \cos(t)) \cos(t) dt
\end{equation}
when \( \nu = 1 \). Given \( f(x) = \log(\exp(-x)I_1(x)) \) we have

\[
(B.4) \quad f'(x) = \frac{d}{dx} \left\{ -x + \log I_1(x) \right\} = -1 + \frac{I'_1(x)}{I_1(x)}
\]

\[
(f''(x) = \frac{d}{dx} \left\{ -1 + \frac{I'_1(x)}{I_1(x)} \right\} = \frac{I''_1(x)}{I_1(x)} - \frac{I'_1(x)^2}{I_1(x)^2}.
\]

Making the change of variables \( u = 2\sqrt{x} \sin(t/2) \) in (B.3) gives

\[
(B.5) \quad I_1(x) = \frac{1}{\pi} \exp(x) \int_{0}^{\pi} \exp(x \cos(t) - x) \cos(t) \, dt
\]

\[
= \frac{1}{\pi} \exp(x) \int_{0}^{\pi} \exp(-2x \sin^2(t/2)) \cos(t) \, dt
\]

\[
= \frac{1}{\pi} \exp(x) \int_{0}^{2\sqrt{x}} \exp(-u^2/2) \cos(t) \sqrt{x} \cos(t/2) \, du
\]

\[
= \frac{1}{\pi} \exp(x) \int_{0}^{2\sqrt{x}} \exp(-u^2/2)(1 - \frac{1}{2} u^2/x)(1 - \frac{1}{4} u^2/x)^{-\frac{1}{2}} \, du
\]

\[
= \exp(x) \frac{2}{\sqrt{2\pi x}} \left\{ \int_{0}^{\infty} - \int_{2\sqrt{x}}^{\infty} \right\} \exp(-u^2/2)(1 - \frac{u^2}{2x})(1 + u^2/8x + \frac{3u^4}{128x^2} + \frac{5u^6}{1024x^3} + O(\frac{u^8}{x^4})) \, du
\]

\[
= \exp(x) \frac{2}{\sqrt{2\pi x}} \left\{ \int_{0}^{\infty} \exp(-u^2/2)(1 - \frac{3u^2}{8x} - \frac{5u^4}{128x^2} - \frac{7u^6}{1024x^3} + O(\frac{u^8}{x^4})) \, du
\]

\[
= \exp(x) \left\{ 1 - \frac{3}{8x} - \frac{15}{128x^2} - \frac{105}{1024x^3} + O\left(\frac{1}{x^4}\right) \right\},
\]

where we have made use of

\[
(B.6) \quad \sqrt{\frac{x}{\pi}} \int_{0}^{\infty} \exp(-u^2/2)u^{2k} \, du = (2k - 1)(2k - 3) \ldots 1,
\]

for \( k \in \{0, 1, 2, \ldots\} \) and

\[
(B.7) \quad \int_{2\sqrt{x}}^{\infty} \exp(-u^2/2) \, du \leq \int_{2\sqrt{x}}^{\infty} \frac{u}{2\sqrt{x}} \exp(-u^2/2) \, du = \frac{\exp(-2x)}{2\sqrt{x}}.
\]

Similarly we have

\[
(B.8) \quad I'_1(x) = \frac{1}{\pi} \exp(x) \int_{0}^{\pi} \exp(x \cos(t) - x) \cos^2(t) \, dt
\]

\[
= \frac{1}{\pi} \exp(x) \int_{0}^{2\sqrt{x}} \exp(-u^2/2)(1 - \frac{1}{2} u^2/x)^2(1 - \frac{1}{4} u^2/x)^{-\frac{1}{2}} \, du
\]

\[
= \exp(x) \frac{2}{\sqrt{2\pi x}} \left\{ \int_{0}^{\infty} \exp(-u^2/2)(1 - \frac{7u^2}{8x} + \frac{19u^4}{128x^2} + \frac{13u^6}{1024x^3} + O(\frac{u^8}{x^4})) \, du
\]

\[
= \exp(x) \left\{ 1 - \frac{7}{8x} + \frac{57}{128x^2} + \frac{195}{1024x^3} + O\left(\frac{1}{x^4}\right) \right\}.
\]
and

\[ I_1''(x) = \frac{1}{\pi} \exp(x) \int_0^\pi \exp(x \cos(t) - x) \cos^3(t) \, dt \]

\[ = \frac{1}{\pi} \int_0^{2\sqrt{x}/\pi} \exp(-u^2/2)(1 - \frac{1}{2}u^2/x)^3(1 - \frac{1}{4}u^2/x)^{-\frac{3}{2}} \, du \]

\[ = \frac{\exp(x)}{\sqrt{2\pi x}} \int_0^\infty \exp(-u^2/2)(1 - \frac{11u^2}{8x} + \frac{75u^4}{128x^2} - \frac{63u^6}{1024x^3} + O(\frac{u^8}{x^4})) \, du \]

\[ = \frac{\exp(x)}{\sqrt{2\pi x}} \left\{ 1 - \frac{11}{8x} + \frac{225}{128x^2} - \frac{945}{1024x^3} + O\left(\frac{1}{x^4}\right) \right\}. \]

From these expressions we have

\[ \frac{I_1'(x)}{I_1(x)} = 1 - \frac{1}{2x} + \frac{3}{8x^2} - \frac{3}{8x^3} + O\left(\frac{1}{x^4}\right) \]

\[ \frac{I_1'(x)^2}{I_1(x)^2} = 1 - \frac{1}{x} + \frac{1}{x^2} + \frac{3}{8x^3} + O\left(\frac{1}{x^4}\right) \]

\[ \frac{I_1''(x)}{I_1(x)} = 1 - \frac{1}{x} + \frac{3}{2x^2} - \frac{3}{8x^3} + O\left(\frac{1}{x^4}\right) \]

and the result follows. \(\square\)

**Appendix C. Shiller’s Monthly US Data Set 1871 to 2017**

Shiller’s monthly US data set contains cash rates and S&P500 total return values over the period from 1871 to 2012, as given in Shiller [1989] and his website. Values from 2013 to 2017 have been spliced using data obtained from Bloomberg data services. The cash savings account \(B_t\) commences with a value of one and is iteratively computed using the formula

\[ B_{t+\Delta} = (1 + r_t^{(1)} \frac{d}{360}) B_t \]

where \(r_t^{(1)}\) is the one-year cash rate and \(d\) is the number of days in the period of length \(\Delta\). The stock index, used as a proxy for the growth optimal portfolio, commences with a value of one and is iteratively computed using the formula

\[ S_t^{\delta_t} = S_t^{\delta_t} \times (I_{t+\Delta} + D_{t+\Delta})/I_t \]

where \(I_t\) is the stock price index at time \(t\) and \(D_t\) is the gross value of dividends paid during period \((t, t + \Delta]\). We assume that a financial institution’s hedging profits can be offset against its product offerings and therefore attract zero taxation.

**Appendix D. Daily Data Set January 1970 to May 2017**

The daily US data set comprises the Federal Funds Rate and S&P 500 total return index values over the period from January 1970 to May 2017, sourced from Bloomberg data services.
References


(K. Fergusson) **University of Melbourne, Centre for Actuarial Studies, Carlton VIC 3059, Australia**  
*E-mail address*, K. Fergusson: kevin.fergusson@unimelb.edu.au