An Efficient Estimation for Switching Regression Models: A Monte Carlo Study

Dinghai Xu*

Abstract

This paper investigates an efficient estimation method for a class of switching regressions based on the characteristic function (CF). We show that with the exponential weighting function, the CF based estimator can be achieved from minimizing a closed form distance measure. Due to the availability of the analytical structure of the asymptotic covariance, an iterative estimation procedure is developed involving the minimization of a precision measure of the asymptotic covariance matrix. Numerical examples are illustrated via a set of Monte Carlo experiments examining the implentability, finite sample property and efficiency of the proposed estimator.

Keywords: Switching Regression model, Characteristic Function; Integrated Squared Error; Gaussian Mixtures.

This project is partially supported by the SSHRC funding from University of Waterloo and the IQFI centre.

*Department of Economics, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada. Email: dhxu@uwaterloo.ca; Tel: 001-519-888-4567 ext. 32047.

1 Introduction

In recent years, increasing attention has been focused on the problem of discontinuous jumps or shifts among regimes in the data. One popular class of the approaches to accommodate this characteristics is through the finite mixture modeling. The most intuitive underlying assumption is that the sample data is drawn from different sub-group components. In other words, the population can be viewed as a mixture of distinct sources. In biology, for instance, the heterogeneities often come from genders, species, genetics classes and etc; see McLachlan and Peel (2000) and references therein. In finance applications, for example, the stock returns can be viewed as mixtures of different information components; see Kon (1984). We shall see that there are many other applications across different research disciplines, including astronomy, engineering, economics, psychiatry and etc. Some of the excellent surveys are collected in Everitt and Hand (1981), Titterington, Smith and Makov (1985) and Mclachlan and Peel (2000). One main characteristics of adopting mixture models is due to the flexibility nature. As a note, it is well known that any continuous distribution can be approximated arbitrarily well by an appropriate finite Gaussian mixtures. In addition, the interpretation of mixture models is straightforward. In this paper, we investigate a class of mixture models, namely switching regressions (SWR), which can be viewed as natural extensions of the mixtures of normal (MN) settings.

In general, if a random variable Y is drawn from a K-component MN, its probability density function (pdf) is defined as,

$$f(y) = \sum_{k=1}^{K} p_k \phi_k(y; \mu_k, \sigma_k^2) \tag{1}$$

where p_k is the mixing proportion or weight of the kth component in the mixture with the restrictions that $0 \leq p_k \leq 1$ and $\sum_{k=1}^{K} p_k = 1$. For $k = 1, 2, ..., K$,

$$\phi_k(y; \mu_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp(-\frac{(y-\mu_k)^2}{2\sigma_k^2})$$

The SWR generalizes (1) by allowing the means changing across observations. Consequently, it extends to the following set up, for $i = 1, 2, ..., n$,

$$f(y_i) = \sum_{k=1}^{K} p_k \phi_k(y_i; \mu_{ki}, \sigma_k^2) \tag{2}$$

where $\mu_{ki} = x_i'\beta_k$. Or equivalently, it can be expressed as the following regression system,

$$\begin{aligned} y_i &= x_i'\beta_1 + u_{1i} && \text{with probability } p_1 \\ y_i &= x_i'\beta_2 + u_{2i} && \text{with probability } p_2 \\ &... \\ y_i &= x_i'\beta_K + u_{Ki} && \text{with probability } p_K \end{aligned} \tag{3}$$

with $u_{ki} \sim N(0, \sigma_k^2)$. In general, x_i' is $n \times m$ and the corresponding coefficient β_k is $m \times 1$.[1] Then, we have $(K \times m + 2K - 1)$ unknown parameters, which is specified as $\theta = (p_1, ..., p_{K-1}; \beta_1, ..., \beta_K; \sigma_1^2, ..., \sigma_K^2)'$.

The above representations allow great flexibility to capture various density shapes, however, the flexibility causes the problems in the likelihood function. More specifically, the likelihood function under the above mixture settings is unbounded over the parameter space and consequently, the conventional maximum likelihood estimator (MLE) is not well defined.[2] Attempting to solve this fundamental statistical problem alternative to likelihood-based approach dates back to Cohen (1967) and Day (1969), who apply the method of moments (MOM) for estimating the mixture parameters. Due to the inefficiency of the MOM estimator, Quant and Ramsey (1978) and Schimdt (1982) propose an approach based on the moment generating function (MGF). Alternatively, a class of estimation methods based on the characteristic function (CF) have been used in this context, see Tran (1998), Knight and Yu (2002) and Xu and Knight (2008).[3]

In this paper, an iterative estimation procedure is developed based on both the closed form objective distance function and the asymptotic covariance measure. This proposed iterative procedure essentially updates the weighting function and consequently, it improves the estimation efficiency and reduces the computational cost.

The paper is organized as follows. Section 2 derives an closed form distance function between the theoretical CF and its empirical counterpart

[1] If x_i' is a set of different order of lagged values of y_i, (3) assumes the similar set up as the mixture autoregressions (MAR) structure, see Wong and Li (2000).

[2] We will provide more discussions on the failure of the ordinary MLE procedure in the next section.

[3] Another popular class of the estimation approaches for mixture models are the Bayesian methods. In this paper, we are not trying to conduct a general survey. An excellent survey regarding to Bayesian type approaches is documented in McLanchlan and Peel (2000).

(ECF) under a general exponential weighting kernel for the SWR. It also discusses the asymptotic properties of the proposed estimator along with the iterated procedure. Section 3 conducts a set of Monte Carlo simulations and demonstrate the improvement of the asymptotic efficiency. Section 4 concludes the paper. All the proofs, Tables and Figures are collected in the Appendix.

2 SWR Model and An Iterative Estimation Procedure

We start this section with examining the unboundedness in the mixture likelihood function and for illustrative purposes, we restrict our attention to the case where the mixtures have two components. Consider the case where K=2 in (1), i.e., the mixture pdf of y_i $(i = 1, 2, ..., n)$ can be written as,

$$f(y_i) = \frac{p_1}{\sqrt{2\pi\sigma_1^2}} \exp(-\frac{(y_i - \mu_1)^2}{2\sigma_1^2}) + \frac{1 - p_1}{\sqrt{2\pi\sigma_2^2}} \exp(-\frac{(y_i - \mu_2)^2}{2\sigma_2^2}) \tag{4}$$

We arbitrarily set the parameter values of μ_2 and σ_2^2 and choose μ_1 to equal to the jth element in the data, i.e., $\mu_1 = y_j$. In other words, the jth residual vanishes from the first regime. Then, for any $p_1 \in (0, 1)$ with the chosen μ_1, μ_2 and σ_2^2, we examine the behavior of the likelihood function over a sequence of points as σ_1^2 approaches to 0. In (4), for $i = 1, 2, ..., n$, $\frac{1-p_1}{\sqrt{2\pi\sigma_2^2}} \exp(-\frac{(y_i-\mu_2)^2}{2\sigma_2^2}) > 0$. As $\sigma_1^2 \to 0$, for $i = 1, ..., j - 1, j + 1, ..., n$, $\frac{p_1}{\sqrt{2\pi\sigma_1^2}} \exp(-\frac{(y_i-\mu_1)^2}{2\sigma_1^2}) \to 0$. When $i = j$, $\frac{p_1}{\sqrt{2\pi\sigma_1^2}} \exp(-\frac{(y_j-\mu_1)^2}{2\sigma_1^2}) = \frac{p_1}{\sqrt{2\pi\sigma_1^2}} e^0$. As σ_1^2 approaches to 0, the first part of jth term in (4) will become arbitrarily large. Similar analysis applies to the SWR case where $\mu_{1i} = x_i'\beta_1$ and $\mu_{2i} = x_i'\beta_2$. Therefore, in general, the conventional MLE may falsely maximize an unbounded likelihood, which leads to a numerical instability problem.[4]

Observing the difficulties encountered in the standard MLE, in this paper, we derive an alternative estimator based on the CF to solve the estimability problem. The approach is also referred as an integrated squared error (ISE) method, proposed in Heathcote (1977). In essence, the ISE estimation is

[4]Quant and Ramsey (1978) and Quant (1988) also point out that the problem is essentially due to the singularity of the matrix of second partial derivatives of the log-likelihood function, which is equivalent to a vanishing Jacobian for Gaussian mixtures model from the conventional MLE approach.

similar to the GMM and MGF approaches, which minimize a certain distance measure between the population components and the corresponding empirical counterparts. In the general set-up of the SWR in (2) or (3), the associated theoretical CF for y_i is defined as follows,

$$G_i(y_i; t, \theta, x_i) = E(e^{jty_i}) = \sum_{k=1}^{K} p_k \exp(jx_i'\beta_k t - \frac{1}{2}\sigma_k^2 t^2) \tag{5}$$

where $j = \sqrt{-1}$. θ is a vector of the unknown parameters in the SWR, which has $(K \times m + 2K - 1)$ elements in general.

The ECF for the sample y_i $(i = 1, ..., n)$ from (2) or (3) is of the following form,

$$g_i(y_i; t) = \exp(jty_i) \tag{6}$$

Continuously matching (5) and (6) under a general form of continuous weighting functions, the distance measure is constructed as the objective function for optimization.

$$D(y; \theta, x) = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{+\infty} |G_i(x_i, y_i; t, \theta) - g_i(y_i; t)|^2 w(t) dt \tag{7}$$

where $w(t)$ is a weighting function which ensures the convergence of the integral. In this paper, we adopt a Gaussian kernel form for $w(t)$, i.e., $w(t) = \exp(-bt^2)$. Parameter b can be viewed as a non-negative bandwidth. This kernel form has been used extensively in the literature, see Paulson et al. (1975), Heathcote (1977), Knight and Yu (2002), Besbeas and Morgan (2002) and Xu and Knight (2008). There are several advantages choosing this weighting kernel. In general, the exponential function tends to put more weight around the origin, which is consistent with the CF theory that the CF contains most information around the origin. In the SWR context, we find that under this Gaussian kernel, both the distance function in (7) and the asymptotic covariance structure of the estimator can be derived in closed-form expressions. This would reduce the computational burden in the practical implementation significantly. In addition, this weighting function continuously evaluates the distance between the theoretical CF and the ECF, which avoids the two major problems arising from the discrete type of methods: (i) the choice of the size of the evaluating grids and (ii) the choice of the distance among the grids, see Schmidt (1982). However, as Paulson et al. (1975) and Yu (2004) point out that, with a special weighting form

$\exp(-t^2)$, the estimation may lead to a poor efficiency. One way of improving the efficiency is to use the cross-validation method for the selection of the bandwidth, see Besbeas and Morgan (2002). In this paper, we develop an efficient iterated procedure to continuously update the bandwidth via minimizing a certain precision measure of the asymptotic covariance matrix. This procedure will be discussed in details in the Monte Carlo section.

Proposition 1 If a random sample y_i is generated from a K-regime regression system defined in (2) or (3) and the distance measure is defined as in (7), then the integral can be solved analytically and is given by:

$$\begin{aligned} D(y;\theta,x) &= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{K}\sum_{h\neq k}^{K} 2p_kp_h\sqrt{\frac{\pi}{b+\frac{1}{2}(\sigma_k^2+\sigma_h^2)}}\exp(-\frac{(x_i'\beta_k-x_i'\beta_h)^2}{4b+2(\sigma_k^2+\sigma_h^2)})\right. \\ &+ \left.\sum_{k=1}^{K}p_k^2\sqrt{\frac{\pi}{b+\sigma_k^2}}-2\sum_{k=1}^{K}p_k\sqrt{\frac{\pi}{b+\frac{1}{2}\sigma_k^2}}\exp(-\frac{(y_i-x_i'\beta_k)^2}{4b+2\sigma_k^2})\right) \qquad (8) \end{aligned}$$

Proof: See the Appendix.

The ISE-based estimates result from minimizing (8) with respect to (w.r.t) the unknown parameter, θ, i.e., $\hat{\theta} \in \text{argmin}[D(y;\theta,x)]$. As a note, the bandwidth parameter b is not treated as an unknown in (8). The reason is, referring to Aitchison and Aitken (1976), that a naive optimization of the distance function w.r.t b will yield meaningless result ($b \to \infty$). However, it can be seen that b may play an important role for the estimation since $\hat{\theta}$ is a function of b. Therefore, a robust selection procedure for b is needed for improving the estimation efficiency. As mentioned earlier, with the closed form distance measure in (8), the asymptotic covariance are available in an analytical form. For simplicity, we provide the explicit expression of the asymptotic covariance of $\hat{\theta}$ in the two-regime switching regression case where

$$\begin{aligned} y_i &= \beta_1+\beta_2x_i+u_{1i} && \text{with probability } p_1 \\ y_i &= \gamma_1+\gamma_2x_i+u_{2i} && \text{with probability } 1-p_1 \qquad (9) \end{aligned}$$

where for $i=1,2,...,n$, $u_{1i}\sim N(0,\sigma_1^2)$ and $u_{2i}\sim N(0,\sigma_2^2)$. x_i is $n\times 1$. θ is a 7×1 vector as $(p_1,\beta_1,\beta_2,\gamma_1,\gamma_2,\sigma_1^2,\sigma_2^2)'$.

Proposition 2 If a random sample y_i is generated from a 2-regime regression system defined in (9), the distance measure is defined as in (8) with K=2, and $\hat{\theta}=\text{argmin}[D(y;\theta,x)]$, then $\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{d}N(0,\Lambda^{-1}\Omega\Lambda^{-1})$ with the closed form expressions for Λ and Ω. The derivations are given in the proof.

Proof: See the Appendix.

Based on the closed form asymptotic covariance, the iterative estimation procedure can be implemented as the following steps:[5]

Step 1. Initiate the program with the starting bandwidth value of b, say b_0 ;
Step 2. With $b = b_0$ and the data of y and x, minimize the closed form distance function in (8) and get θ_0 , i.e. $\hat{\theta}_0 = \text{argmin}[D(y; \theta, x)]$;
Step 3. Plug $\hat{\theta}_0$ into the asymptotic covariance matrix M and get M_0. Construct a precision measure, such as the trace or determinant of M, which is a function of b. Update the bandwidth b via $b_1 = \text{argmin}[\text{trace}(M_0)]$ or $b_1 = \text{argmin}[\det(M_0)]$;
Step 4. Repeat the step 2 to 3 until a stopping criterion is met,[6] for example, $|b_t - b_{t-1}| < \epsilon$ and $\epsilon = 10^{-3}$.

Through the above procedure, the estimator's efficiency is improved from the updating process. For comparison, an asymptotic relative efficiency (ARE) measure is constructed, which is defined as the trace of the inverse of the information matrix over the trace of the asymptotic covariance matrix from the updating ISE estimator. In the mixture models, it is difficult to derive an analytical form of the information matrix, see Schimdt (1982). Consequently, the comparisons are conducted numerically in the Monte Carlo section.

3 Monte Carlo Experiments

A set of Monte Carlo experiments are carried out to examine the implentability, finite sample property and efficiency of the proposed procedure. We set up nine cases, in which the first four cases have been examined by Quandt and Ramsey (1978) and the last five cases have been studies in Goldfeld and Quandt (1972) and Quandt (1972). The experiment characteristics are presented in Table 1.[7]

[5]Xu and Knight (2008) apply a similar iterated procedure for estimating the MN parameters. See the paper for more details.

[6]Provided that the influence function of $\hat{\theta}(b)$ is bounded, the optimal b theoretically exists, which has been shown in Besbeas (1999).

[7]Some other experiments with different parameter designs have also been examined, for example, x is generated from a normal distribution or student-t distribution. We found similar results and patterns. To save space, those simulation results are not reported in this paper.

We first examine the asymptotic efficiency of the proposed iterative estimator. Table 2 presents the diagonal elements from the asymptotic covariance matrix of the proposed estimator under the optimal bandwidth value. The optimal value, b^*, for each case, is determined by two precision measures, namely the trace minimization (IT-ISE-TRC) and the determinant minimization (IT-ISE-DET). For comparison, the inverse of the information matrix (IF) is numerically evaluated via the approximation of the expectation of the outer-product of the score vector or the negative expectation of the Hessian matrix based on the log-likelihood function. Consequently, the corresponding diagonal elements from the IF matrix are reported in the last row of Table 2 for each case. Graphically, Figure 2 plots the trace and the determinant over a certain range of b values in each simulation case. For all the cases, the global optimal b values do exist. However, the optimal bandwidth varies from case to case, for instance, b^* changes from 0.3100 to 1.5120 under the trace minimization from Case 4 to Case 9. A more interesting finding is that optimal b values are different under the two precision criteria and in general, the trace minimization generates uniformly smaller asymptotic variances than the determinant minimization does for all the cases (see Table 2 for the detailed comparisons). In addition, the asymptotic variances from the IT-ISE-TRC are fairly close to those from the IF row. Therefore, for the rest of the Monte Carlo simulations, we adopt the trace minimization in the iterative procedure. To further examine the efficiency improvement by iteratively updating the bandwidth parameter value, Table 3 constructs an ARE comparison (using the trace measure). As noted from the table (bold numbers in the 5th row), the proposed estimator achieves the highest efficiency when b is updated to b^* for each individual case. The optimal b values for all the nine cases correspond to b^* in Table 2 by the trace minimization.

To further demonstrate the implentability and finite sample property of the proposed updating procedure, we generate the random samples following the parameter set-up specified in Table 1. The iterative procedure in section 2 is applied for all the Monte Carlo estimation experiments. As mentioned, due to the availabilities of the closed forms in both the objective distance function and asymptotic covariance matrix, the computation cost is not a big constraint. Approximately, each iteration duration is about 2 seconds.[8]

[8]The duration of the iterative estimation may depend on other factors, such as the initial values, the characteristics of the sampling data, the stopping rules for the convergence both in the parameters and bandwidth, optimization software as well as computer speed and etc. Here, we just provide an average approximation time on a computer with CPU 6400 at 2.13G Hz (0.99 GB of RAM) using Matlab 7.1.

With this advantage, we perform 1000 replications for all the cases without much computational burden. The experiment results are reported in Table 4.

Several standard measures are constructed for evaluating the Monte Carlo estimates including mean, bias and the root of mean square error (RMSE). Another reliability measure for the estimator, namely the Kolmogorov-Smirnov (K-S) test statistics, is also shown in Table 4. The mean values of the estimates are, in general, fairly close to the true parameter values. Since for each simulation loop the iterative procedure generates an optimal b, 1000 b^*s are produced for each simulating case. We report the average value over the 1000 replications in the last column of Table 4. As shown, the values of b^*s are very close to the asymptotic values reported in Table 2. This is also consistent with the illustration of convergence for b in Figure 1. Inspecting on the bias and RMSE measures, we find that the proposed estimation procedure produces stable estimates. Noting that even though our sampling sizes are small (50, 60, 100 and 120), the bias and RMSE are generally small and behave well as expected. In Case 9, we detect a relatively large RMSE for σ_2^2, which is 11.0147. The reason is that with a large variance on one mixture component, the probability of producing outliers is high. Consequently, the outliers deteriorate the quality of the estimates by increasing the overall bias and variance. We also identify some consistent results as shown in Quandt and Ramsey (1978), Goldfeld and Quandt (1972) and Quandt (1972). Case 1 and 3 only differ in the sample sizes in each replication. We find that increasing sample size will improve the quality of the estimates. In other words, the bias and RMSE become uniformly smaller in a larger-sample simulation compared with a smaller-sample simulation. Similar argument can be also applied for the comparison of Case 5 and 6. We expected that when the probability of one mixture component increases (high asymmetry), the quality of the estimates gets worse. However, from the simulation results, it is ambiguous. For example, through the comparisons between Case 1 and 2 as well as Case 3 and 4, this happens only 8 out of 14 possible pair comparisons. We observe similar patterns in the comparison of Case 5 and 7. Case 5 and 8 differ only in the variance of the second regime. Bias and RMSE are generally smaller in the lower-variance case (Case 5). Lastly, the regressor, x , is generated from a wider-interval uniform distribution in Case 9 compared to Case 5. We find, as expected, the bias and RMSE are generally smaller in Case 9 (with high separability) than those in Case 5 (with low separability).

Furthermore, a K-S normality test is carried out to investigate the asymptotic normality property of the proposed estimator. The K-S statistics are presented for each experiment in Table 4. The results show that for 39 out of

56 cases, normality can not be rejected at either 5% or 1% confidence level. Increasing the sample size will improve the quality of the K-S measure, see Case 2 and 4. Figure 2 presents the distributions of the estimates over the 1000 replications via the qq-plots.[9] As shown, most of the estimates fit well with the 45-degree quantile line against the normality. This reinforces the K-S test results reported in Table 4. Overall, the proposed procedure seems well-implementable and produces good finite sample properties.

4 Conclusion

This paper uses an iterative bandwidth-updating procedure based on the ISE estimator to efficiently estimate the SWR models. Due to the availabilities of both the objective distance function and the asymptotic covariance matrix, the procedure is easily implemented without much computational burden. The improvement of the estimation efficiency has been shown asymptotically and experimentally. The Monte Carlo simulation shows that the estimator generates sensible results and has good finite sample properties.

Appendix

Proof of Proposition 1

Noting that $\exp(jtz) = \cos(tx) + j\sin(tx)$, $j = \sqrt{-1}$, (5) can be written as:

$$G_i(y_i; t, \theta, x_i) = \sum_{k=1}^{K} p_k \cos(x_i'\beta_k t)\exp(-\frac{1}{2}\sigma_k^2 t^2) + j\sum_{k=1}^{K} p_k \sin(x_i'\beta_k t)\exp(-\frac{1}{2}\sigma_k^2 t^2)$$

Correspondingly, (6) can be transformed in a similar way as:

$$g_i(y_i; t) = \cos(ty_i) + j\sin(ty_i)$$

Define $D_i(y; \theta, x) = \int_{-\infty}^{+\infty} |G_i(x_i, y_i; t, \theta) - g_i(y_i; t)|^2 \exp(-bt^2)dt$, then $|G_i(x_i, y_i; t, \theta) - g_i(y_i; t)|^2$ can be separated into the real and imaginary parts (R and I), i.e.,

$$\begin{aligned} |G_i(x_i, y_i; t, \theta) - g_i(y_i; t)|^2 &= \cos(ty_i) - \sum_{k=1}^{K} p_k \cos(x_i'\beta_k t)\exp(-\frac{1}{2}\sigma_k^2 t^2) \\ &+ j\left(\sin(ty_i) - \sum_{k=1}^{K} p_k \sin(x_i'\beta_k t)\exp(-\frac{1}{2}\sigma_k^2 t^2)\right) \end{aligned}$$

[9] To save space, the qq-plots for the first 5 parameters are reported for each simulation case. The variance estimates are generally found to be significantly deviated from the normality (except for Case 4). Those graphs are available upon request.

Hence,

$$
\begin{aligned}
|G_i - g_i|^2 &= R^2 + I^2 \\
&= 1 + 2\sum_{k=1}^{K}\sum_{h\neq k}^{K} p_k p_h \exp(-\frac{1}{2}t^2(\sigma_k^2+\sigma_h^2))\cos(t(x_i'\beta_k - x_i'\beta_h)) \\
&+ \sum_{k=1}^{K} p_k^2 \exp(-\sigma_k^2 t^2) - 2\sum_{k=1}^{K} p_k \exp(-\frac{1}{2}\sigma_k^2 t^2)\cos(t(y_i - x_i'\beta_k))
\end{aligned}
$$

Now ignoring the constant term, we evaluate integral of the above expression with $\exp(-bt^2)$. Note that $\cos(x) = \frac{\exp(jx)+exp(-jx)}{2}$ and $\sin(x) = \frac{\exp(jx)-exp(-jx)}{2j}$.

$$
\begin{aligned}
D_i &= \int_{-\infty}^{+\infty}(R^2+I^2)\exp(-bt^2)dt \\
&= 2\sum_{k=1}^{K}\sum_{h\neq k}^{K}\int p_k p_h \exp(-\frac{1}{2}t^2(\sigma_k^2+\sigma_h^2))\frac{\exp[jt(x_i'\beta_k - x_i'\beta_h) + \exp[-jt(x_i'\beta_k - x_i'\beta_h)]]}{2}\exp(-bt^2)dt \\
&+ \sum_{k=1}^{K}\int p_k^2\exp(-(\sigma_k^2+b)t^2)dt - 2\sum_{k=1}^{K}\int p_k\exp(-\frac{1}{2}\sigma_k^2t^2)\cos(t(y_i - x_i\beta_k))\exp(-bt^2)dt \\
&= \sum_{k=1}^{K}\sum_{h\neq k}^{K} 2p_kp_h\sqrt{\frac{\pi}{b+\frac{1}{2}(\sigma_k^2+\sigma_h^2)}}\exp(-\frac{(x_i'\beta_k - x_i'\beta_h)^2}{4b+2(\sigma_k^2+\sigma_h^2)}) \\
&+ \sum_{k=1}^{K} p_k^2\sqrt{\frac{\pi}{b+\sigma_k^2}} - 2\sum_{k=1}^{K} p_k\sqrt{\frac{\pi}{b+\frac{1}{2}\sigma_k^2}}\exp(-\frac{(y_i - x_i'\beta_k)^2}{4b+2\sigma_k^2})
\end{aligned}
$$

Therefore, substituting D_i into the distance measure in (8), defined as $D(y;\theta,x) = \frac{1}{n}\sum_{i=1}^{n} D_i$, completes the proof.

Proof of Proposition 2

In general, we define the prime as the differentiation w.r.t. the parameter vector. Let $Re[f]$ and $Im[f]$ be the real and imaginary part of f respectively. Then,

$$D'(\theta) = \frac{\partial D(\theta)}{\partial\theta} = -\frac{2}{n}\sum_{i=1}^{n} H_i(\theta)$$

With $H_i(\theta) = \int\left((\cos(ty_i) - Re[G_i])\frac{\partial Re[G_i]}{\partial\theta} + (\sin(ty_i) - Im[G_i])\frac{\partial Im[G_i]}{\partial\theta}\right)\exp(-bt^2)dt$.
Let $\Omega_i = \mathrm{var}(\frac{\partial D_i(\theta)}{\partial\theta})$. Then $\mathrm{var}\left(\frac{\partial D(\theta)}{\partial\theta}\right) = \frac{4}{n^2}\mathrm{var}(\sum_{i=1}^{n} H_i) = \frac{1}{n}\sum_{i=1}^{n}\Omega_i = \frac{1}{n}\Omega.$

From Heathcote (1977), H_i is bounded and by the Central Limit Theorem (CLT) and Laws of Large Numbers(LLN), at the true parameter value, θ^*, we have,

$$\sqrt{n}D'(\theta^*) \quad \xrightarrow{d} \quad N(0,\Omega)$$

Now, we take the expectation of the second derivative of $D(\theta)$ evaluated at θ^*, which is,

$$E(D''(\theta)) = E\left(\frac{\partial D^2(\theta^*)}{\partial\theta\partial\theta'}\right) = \Lambda$$

First-order Taylor expansion of $D'(\hat{\theta})$ around the true parameter value, θ^*, yields,

$$D'(\hat{\theta}) = D'(\theta^*) + (\hat{\theta} - \theta^*)D''(\theta^*)$$

Then, the Slutsky Theorem immediately implies,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Lambda^{-1}\Omega\Lambda^{-1})$$

In the two regime-switching case (when $K = 2$) defined in (9), the distance function, (8), can be written as the following closed form:

$$\begin{aligned} D(y;\theta,x) &= \frac{1}{n}\sum_{i=1}^{n}\left(2p_1(1-p_1)\sqrt{\frac{\pi}{b+\frac{1}{2}(\sigma_1^2+\sigma_2^2)}}\exp(-\frac{(\beta_1+\beta_2x_i-\gamma_1-\gamma_2x_i)^2}{4b+2(\sigma_1^2+\sigma_2^2)})\right. \\ &+ p_1^2\sqrt{\frac{\pi}{b+\sigma_1^2}} + (1-p_1)^2\sqrt{\frac{\pi}{b+\sigma_2^2}} - 2p_1\sqrt{\frac{\pi}{b+\frac{1}{2}\sigma_1^2}}\exp(-\frac{(y_i-\beta_1-\beta_2x_i)^2}{4b+2\sigma_1^2}) \\ &- \left. 2(1-p_1)\sqrt{\frac{\pi}{b+\frac{1}{2}\sigma_2^2}}\exp(-\frac{(y_i-\gamma_1-\gamma_2x_i)^2}{4b+2\sigma_2^2})\right) \end{aligned}$$

There are 7 unknown parameters to be estimated, which are $(p_1, \beta_1, \beta_2, \gamma_1, \gamma_2, \sigma_1^2, \sigma_2^2)$. Following the construction of Ω and Λ, we essentially need to solve for the two expectations in the Ω and Λ expressions. With an analytical form of D_i, the software Maple or Mathematica can be applied for solving the closed form solutions of each corresponding elements in the asymptotic covariance matrix. Alternatively, in this paper, we develop the closed form solution of Ω and Λ based on the solutions for asymptotic covariance in the mixtures of two normal case. The following results are established in Xu and Knight (2008) and Xu (2007).

In the mixtures of two normal case, $\theta^m = (p_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ and $\sqrt{n}(\hat{\theta}^m - \theta) \xrightarrow{d} N(0, (\Lambda^m)^{-1}\Omega^m(\Lambda^m)^{-1})$, where

$$\Lambda_{ij}^m = U_{ij}^m + V_{ij}^m \times \exp\left(-\frac{(\mu_1-\mu_2)^2}{4b+2\sigma_1^2+2\sigma_2^2}\right) \qquad i,j = 1,2,...,5$$

with

$$U^m = \begin{bmatrix} \frac{\sqrt{\pi}}{\sqrt{b+\sigma_1^2}} + \frac{\sqrt{\pi}}{\sqrt{b+\sigma_1^2}} & 0 & 0 & -\frac{p_1\sqrt{\pi}}{4(b+\sigma_1^2)^{\frac{3}{2}}} & \frac{(1-p_1)\sqrt{\pi}}{4(b+\sigma_1^2)^{\frac{3}{2}}} \\ & \frac{p_1^2\sqrt{\pi}}{2(b+\sigma_1^2)^{\frac{3}{2}}} & 0 & 0 & 0 \\ & & \frac{(1-p_1)^2\sqrt{\pi}}{2(b+\sigma_2^2)^{\frac{3}{2}}} & 0 & 0 \\ & & & \frac{3p_1^2\sqrt{\pi}}{16(b+\sigma_1^2)^{\frac{5}{2}}} & 0 \\ & & & & \frac{3(1-p_1)^2\sqrt{\pi}}{16(b+\sigma_2^2)^{\frac{5}{2}}} \end{bmatrix}$$

$$V^m = \begin{bmatrix} V11 & V12 & V13 & V14 & V15 \\ & V22 & V23 & V24 & V25 \\ & & V33 & V34 & V35 \\ & & & V44 & V45 \\ & & & & V55 \end{bmatrix}$$

The expressions for V^m matrix are available in Xu and Knight (2008) and Xu (2007).

In our two regime switching regression case, only the mean of each mixture component changes across the observations, particularly, $\mu_1 = \beta_1 + \beta_2 x_i$ and $\mu_2 = \gamma_1 + \gamma_2 x_i$. The following relations can be easily derived through the Chain rule of the derivatives, for example,

$$\begin{aligned} E(\frac{\partial^2 D}{\partial p_1 \partial \beta_1}) &= E(\frac{\partial^2 D}{\partial p_1 \partial \mu_1}\frac{\partial \mu_1}{\partial \beta_1}) = E(\frac{\partial^2 D}{\partial p_1 \partial \mu_1}); \\ E(\frac{\partial^2 D}{\partial \beta_1 \partial \beta_2}) &= E(\frac{\partial^2 D}{\partial \mu_1 \partial \mu_1}\frac{\partial \mu_1}{\partial \beta_1}\frac{\partial \mu_1}{\partial \beta_2}) = E(\frac{\partial^2 D}{\partial \mu_1 \partial \mu_1})x_i; \\ E(\frac{\partial^2 D}{\partial \beta_1 \partial \gamma_2}) &= E(\frac{\partial^2 D}{\partial \mu_1 \partial \mu_2}\frac{\partial \mu_1}{\partial \beta_1}\frac{\partial \mu_2}{\partial \gamma_2}) = E(\frac{\partial^2 D}{\partial \mu_1 \partial \mu_2})x_i; \\ E(\frac{\partial^2 D}{\partial \beta_2 \partial \gamma_2}) &= E(\frac{\partial^2 D}{\partial \mu_1 \partial \mu_2}\frac{\partial \mu_1}{\partial \beta_2}\frac{\partial \mu_2}{\partial \gamma_2}) = E(\frac{\partial^2 D}{\partial \mu_1 \partial \mu_2})x_i^2; \\ E(\frac{\partial^2 D}{\partial \beta_2 \partial \sigma_1^2}) &= E(\frac{\partial^2 D}{\partial \mu_1 \partial \sigma_2^2}\frac{\partial \mu_1}{\partial \beta_2}) = E(\frac{\partial^2 D}{\partial \mu_1 \partial \sigma_2^2})x_i \end{aligned}$$

Other corresponding elements in the Λ matrix can be easily derived in similar way as shown above. In other words, given the closed form of Λ^m, we can fill in the Λ matrix (7×7) using the above transformations. Specifically,

$$\Lambda_i = \begin{bmatrix} \Lambda_{11}^m & \Lambda_{12}^m & \Lambda_{12}^m x_i & \Lambda_{13}^m & \Lambda_{13}^m x_i & \Lambda_{14}^m & \Lambda_{15}^m \\ & \Lambda_{22}^m & \Lambda_{22}^m x_i & \Lambda_{23}^m & \Lambda_{23}^m x_i & \Lambda_{24}^m & \Lambda_{25}^m \\ & & \Lambda_{22}^m x_i^2 & \Lambda_{23}^m x_i & \Lambda_{23}^m x_i^2 & \Lambda_{24}^m x_i & \Lambda_{25}^m x_i \\ & & & \Lambda_{33}^m & \Lambda_{33}^m x_i & \Lambda_{34}^m & \Lambda_{35}^m \\ & & & & \Lambda_{33}^m x_i^2 & \Lambda_{34}^m x_i & \Lambda_{35}^m x_i \\ & & & & & \Lambda_{44}^m & \Lambda_{45}^m \\ & & & & & & \Lambda_{55}^m \end{bmatrix}$$

Similarly, the Ω matrix in SWR case can be expressed using the Ω^m matrix from the mixtures of two normal case using the following transformations, for example,

$$\begin{aligned} E(\frac{\partial D}{\partial p_1}\frac{\partial D}{\partial \beta_1}) &= E(\frac{\partial D}{\partial p_1}\frac{\partial D}{\partial \mu_1}\frac{\partial \mu_1}{\partial \beta_1}) = E(\frac{\partial D}{\partial p_1}\frac{\partial D}{\partial \mu_1}); \\ E(\frac{\partial D}{\partial \beta_1}\frac{\partial D}{\partial \beta_2}) &= E(\frac{\partial D}{\partial \mu_1}\frac{\partial D}{\partial \mu_1}\frac{\partial \mu_1}{\partial \beta_1}\frac{\partial \mu_1}{\partial \beta_2}) = E(\frac{\partial D}{\partial \mu_1}\frac{\partial D}{\partial \mu_1})x_i; \\ E(\frac{\partial D}{\partial \beta_1}\frac{\partial D}{\partial \gamma_2}) &= E(\frac{\partial D}{\partial \mu_1}\frac{\partial D}{\partial \mu_2}\frac{\partial \mu_1}{\partial \beta_1}\frac{\partial \mu_2}{\partial \gamma_2}) = E(\frac{\partial D}{\partial \mu_1}\frac{\partial D}{\partial \mu_2})x_i; \\ E(\frac{\partial D}{\partial \beta_2}\frac{\partial D}{\partial \gamma_2}) &= E(\frac{\partial D}{\partial \mu_1}\frac{\partial D}{\partial \mu_2}\frac{\partial \mu_1}{\partial \beta_2}\frac{\partial \mu_2}{\partial \gamma_2}) = E(\frac{\partial D}{\partial \mu_1}\frac{\partial D}{\partial \mu_2})x_i^2; \\ E(\frac{\partial D}{\partial \beta_2}\frac{\partial D}{\partial \sigma_1^2}) &= E(\frac{\partial D}{\partial \mu_1}\frac{\partial D}{\partial \sigma_1^2}\frac{\partial \mu_1}{\partial \beta_2}) = E(\frac{\partial D}{\partial \mu_1}\frac{\partial D}{\partial \sigma_1^2})x_i \end{aligned}$$

Consequently,

$$\Omega_i = \begin{bmatrix} \Omega_{11}^m & \Omega_{12}^m & \Omega_{12}^m x_i & \Omega_{13}^m & \Omega_{13}^m x_i & \Omega_{14}^m & \Omega_{15}^m \\ & \Omega_{22}^m & \Omega_{22}^m x_i & \Omega_{23}^m & \Omega_{23}^m x_i & \Omega_{24}^m & \Omega_{25}^m \\ & & \Omega_{22}^m x_i^2 & \Omega_{23}^m x_i & \Omega_{23}^m x_i^2 & \Omega_{24}^m x_i & \Omega_{25}^m x_i \\ & & & \Omega_{33}^m & \Omega_{33}^m x_i & \Omega_{34}^m & \Omega_{35}^m \\ & & & & \Omega_{33}^m x_i^2 & \Omega_{34}^m x_i & \Omega_{35}^m x_i \\ & & & & & \Omega_{44}^m & \Omega_{45}^m \\ & & & & & & \Omega_{55}^m \end{bmatrix}$$

References

[1] Aitchison. J and Aitken C.G.G. (1976), Multivariate Binary Discrimination by the Kernel Method. *Biometrika*, 63, pp. 413-421.

[2] Besbeas, P. and Morgan, B.J.T. (2002), Integrated Squared Error Estimation of Normal Mixtures. *Computational Statistics and Data Analysis*, 44 (3), 517-526.

[3] Besbeas, P., (1999), Parameter Estimation Based on Empirical Transforms. Ph.D. Thesis, University of Kent, UK.

[4] Cohen, A.C. (1967), Estimation In Mixtures of Two Normal Distributions, *Biometrika*, 56, 15-28.

[5] Day, N.E. (1969), Estimating the Components of a Mixture of Normal Distributions, *Biometrika*, 56, 463-474.

[6] Everitt, B. S. and Hand, D. J. (1981), *Finite Mixture Distributions*, Chapman and Hall.

[7] Goldfeld, S. M., and Quandt, R. E. (1972), Nonlinear Methods in Econometrics, Amsterdam, North-Holland Publishing Co.

[8] Heathcote, C.R. (1977), Integrated Mean Square Error Estimation of Parameters, *Biometrika*, 64, 255-264.

[9] McLachlan, G. and Peel, D. (2000), *Finite Mixture Models*, John Wiley & Sons, INC.

[10] Knight, J. L. and Yu, J. (2002), Empirical Characteristic Function in Time Series Estimation, *Econometric Theory*, 18, 691 - 721.

[11] Kon, S. (1984), Models of Stock Returns - A Comparison, *The Journal of Finance*, 39, 147 - 165.

[12] Paulson, A. S. , Holcomb, E. W. and Leitch, R.A. (1975), The Estimation of the Parameters of the Stable Laws, *Biometrika*, 62, 163-170.

[13] Quandt, R. E. (1972), A New Approach to Estimating Switching Regressions, *Journal of the American Statistical Association*, 67, 306-310.

[14] Quandt, R. E. and Ramsey, J. B. (1978), Estimating Mixtures of Normal Distributions and Switching Regressions, *Journal of the American Statistical Association*, 73, 730 - 738.

[15] Quandt, R. E. (1988), *The Econometrics of Disequilibrium*, Basil Blackwell.

[16] Schmidt, P. (1982), An Improved Version of the Quandt-Ramsey MGF Estimator for Mixtures of Normal Distributions and Switching Regressions, *Econometrica*, 50, 501 - 516.

[17] Titterington, D. M. , Smith, A. F. M. and Makov, U. E. (1985), *Statistical Analysis of Finite of Mixture Distributions*, John Willey & Son Ltd.

[18] Tran, K. (1998), Estimating Mixtures of Normal Distributions via Empirical Characteristic Function, *Econometric Reviews*, 17(2), 167-183.

[19] Wong, C. and Li, W. (2000), On a Mixture Autoregressive Model, *Journal of the Royal Statistical Society: Series B*, 62(1), 95-115.

[20] Xu, D. (2007), Asset Returns, Volatility and Value at Risk, Ph.D Thesis, The University of Western Ontario.

[21] Xu, D. and Knight, J. L. (2008), Continuous Empirical Characteristic Function Estimation of Mixtures of Normal Parameters, Working Paper # 08-006, University of Waterloo.

[22] Yu, J. (2004), Empirical Characteristic Function Estimation and Its Applications, *Econometric Reviews*, 23, 93-123.

Table 1 Monte Carlo Simulation Design

Case	p_1	β_1	β_2	γ_1	γ_2	σ_1^2	σ_2^2	x_i	n
1	0.5	-1.0	0.5	-3.08333	1.0	3.0	2.0	$U(0,10)$	50
2	0.8	-1.0	0.5	-3.08333	1.0	3.0	2.0	$U(0,10)$	50
3	0.5	-1.0	0.5	-3.08333	1.0	3.0	2.0	$U(0,10)$	100
4	0.8	-1.0	0.5	-3.08333	1.0	3.0	2.0	$U(0,10)$	100
5	0.5	1.0	1.0	0.5	1.5	2.0	2.5	$U(10,20)$	60
6	0.5	1.0	1.0	0.5	1.5	2.0	2.5	$U(10,20)$	120
7	0.75	1.0	1.0	0.5	1.5	2.0	2.5	$U(10,20)$	60
8	0.5	1.0	1.0	0.5	1.5	2.0	25.0	$U(10,20)$	60
9	0.5	1.0	1.0	0.5	1.5	2.0	2.5	$U(0,40)$	60

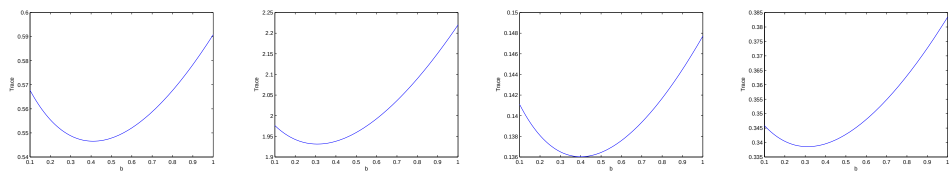
Table 2 Comparisons of Asymptotic Variances from ISE under Optimal b Values and Information Matrix

	var(p_1)	var(β_1)	var(β_2)	var(γ_1)	var(γ_2)	var(σ_1^2)	var(σ_2^2)	b^*
				Case 1.				
IT-ISE-TRC	0.0265	0.2056	0.0104	0.0940	0.0047	0.0853	0.1310	0.4152
IT-ISE-DET	0.0394	0.2586	0.0137	0.1371	0.0074	0.0703	0.1963	1.6737
IF	0.0279	0.2427	0.0120	0.1028	0.0050	0.1161	0.1368	-
				Case 2.				
IT-ISE-TRC	0.0269	0.0838	0.0042	0.5892	0.0296	0.0353	0.8098	0.3121
IT-ISE-DET	0.0378	0.0990	0.0052	0.8487	0.0456	0.0274	1.1848	1.6582
IF	0.0291	0.0959	0.0048	0.6410	0.0313	0.0435	0.8778	-
				Case 3.				
IT-ISE-TRC	0.0064	0.0492	0.0025	0.0225	0.0011	0.0211	0.0323	0.4129
IT-ISE-DET	0.0098	0.0640	0.0034	0.0338	0.0018	0.0175	0.0489	1.6761
IF	0.0053	0.0449	0.0022	0.0207	0.0010	0.0186	0.0255	-
				Case 4.				
IT-ISE-TRC	0.0065	0.0203	0.0010	0.1427	0.0072	0.0087	0.2009	0.3100
IT-ISE-DET	0.0094	0.0244	0.0013	0.2081	0.0113	0.0068	0.2948	1.6622
IF	0.0056	0.0178	0.0009	0.1293	0.0063	0.0074	0.1631	-
				Case 5.				
IT-ISE-TRC	0.0001	0.0486	0.0002	0.0628	0.0003	0.0092	0.0151	0.7136
IT-ISE-DET	0.0001	0.0493	0.0002	0.0638	0.0003	0.0097	0.0157	0.9503
IF	0.0001	0.0546	0.0002	0.0698	0.0003	0.0104	0.0164	-
				Case 6.				
IT-ISE-TRC	2.16e-5	0.0120	4.91e-5	0.0155	0.0001	0.0023	0.0038	0.7103
IT-ISE-DET	2.24e-5	0.0121	4.93e-5	0.0157	0.0001	0.0024	0.0039	0.9434
IF	2.17e-5	0.0121	4.98e-5	0.0145	0.0001	0.0022	0.0032	-
				Case 7.				
IT-ISE-TRC	0.0001	0.0299	0.0001	0.1431	0.0006	0.0055	0.0349	0.6603
IT-ISE-DET	0.0001	0.0295	0.0001	0.1473	0.0006	0.0054	0.0381	0.9723
IF	0.0001	0.0301	0.0001	0.2172	0.0009	0.0056	0.0599	-
				Case 8.				
IT-ISE-TRC	0.0003	0.0538	0.0002	1.1286	0.0042	0.0175	2.6414	0.9587
IT-ISE-DET	0.0003	0.0534	0.0002	1.1279	0.0042	0.0181	2.6173	1.0909
IF	0.0004	0.0607	0.0003	0.9944	0.0039	0.0184	3.7729	-
				Case 9.				
IT-ISE-TRC	0.0001	0.0086	1.36e-5	0.0108	1.72e-5	0.0089	0.0137	1.5120
IT-ISE-DET	0.0001	0.0087	1.36e-5	0.0108	1.72e-5	0.0094	0.0139	2.0753
IF	0.0001	0.0094	1.62e-5	0.0118	1.98e-5	0.0104	0.0152	-

Figure 1 Optimal Bandwidth Values

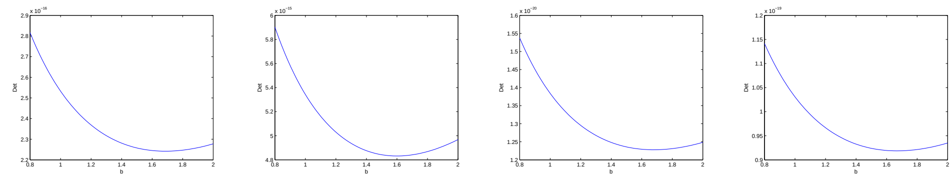
By Trace Minimization:

Case 1	Case 2	Case 3	Case 4



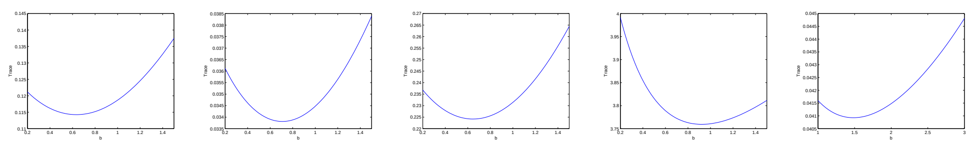
By Determinant Minimization:

Case 1	Case 2	Case 3	Case 4



By Trace Minimization:

Case 5	Case 6	Case 7	Case 8	Case 9



By Determinant Minimization:

Case 5	Case 6	Case 7	Case 8	Case 9

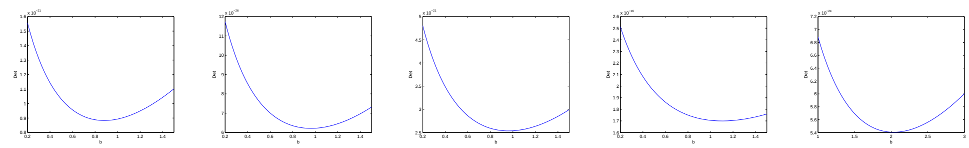


Table 3 Asymptotic Relative Efficiency Comparisons

b values	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
$b = 0.005$	1.0580	1.1581	0.8214	0.8080	0.8512	0.8391	1.2759	1.1857	0.8720
$b = 0.01$	1.0602	1.1602	0.8230	0.8095	0.8530	0.8408	1.2783	1.1885	0.8742
$b = 0.1$	1.0934	1.1910	0.8476	0.8313	0.8818	0.8697	1.3192	1.2327	0.9116
$b = b^*$	**1.1391**	**1.2194**	**0.8794**	**0.8531**	**0.9641**	**0.9627**	**1.4386**	**1.3524**	**1.0895**
$b = 1$	1.0597	1.0844	0.8138	0.7490	0.9391	0.9477	1.3960	1.3521	1.0856
$b = 5$	0.3190	0.2799	0.2483	0.1751	0.3727	0.3560	0.4788	1.0581	0.7153
$b = 10$	0.1075	0.0893	0.0902	0.0421	0.2810	0.2645	0.3606	0.7440	0.3456

Table 4 Monte Carlo Simulation Statistics Summary

* Normality is not rejected at 5% significance level (cut-off value is 0.0428);
** Normality is not rejected at 1% significance level (cut-off value is 0.0513).

	p_1	β_1	β_2	γ_1	γ_2	σ_1^2	σ_2^2	b^*
	Case 1.							
Mean	0.5029	-0.8921	0.4887	-3.2915	1.0284	1.8732	1.2613	0.4139
Bias	0.0029	0.1079	-0.0113	-0.2082	0.0284	-1.1268	-0.7387	-
RMSE	0.1883	1.0541	0.2067	1.2777	0.2216	1.7929	1.1237	-
K-S	0.0704	0.0290**	0.0436*	0.0390**	0.0433*	0.1176	0.1060	-
	Case 2.							
Mean	0.6299	-0.6020	0.4259	-3.0489	0.9517	2.0281	1.0556	0.3108
Bias	-0.1701	0.3980	-0.0741	0.0344	-0.0483	-0.9719	-0.9444	-
RMSE	0.2358	1.2016	0.1865	1.4306	0.2426	1.5987	1.3292	-
K-S	0.1007	0.0313**	0.0545	0.0251**	0.0436*	0.0822	0.1805	-
	Case 3.							
Mean	0.5202	-1.0565	0.5066	-3.0992	1.0125	2.2628	1.4440	0.4121
Bias	0.0202	-0.0565	0.0066	-0.0158	0.0125	-0.7372	-0.5560	-
RMSE	0.1783	1.0376	0.1814	0.9780	0.1684	1.4027	1.0025	-
K-S	0.0757	0.0304**	0.0801	0.0508*	0.0221**	0.0679	0.0790	-
	Case 4.							
Mean	0.7178	-0.8979	0.4710	-2.9912	0.9640	2.5088	1.0031	0.3096
Bias	-0.0822	0.1021	-0.0290	0.0921	-0.0360	-0.4912	-0.9969	-
RMSE	0.1659	0.8005	0.1279	1.4358	0.2523	1.0877	1.2900	-
K-S	0.1140	0.0503*	0.0425**	0.0579	0.0276**	0.0300**	0.1640	-
	Case 5.							
Mean	0.4998	1.0601	0.9975	0.3886	1.5080	1.8432	2.3257	0.7138
Bias	-0.0002	0.0601	-0.0025	-0.1114	0.0080	-0.1568	-0.1743	-
RMSE	0.0751	1.6036	0.1027	1.9707	0.1244	0.7229	1.0072	-
K-S	0.0177**	0.0279**	0.0234**	0.0407**	0.0341**	0.0738	0.0961	-
	Case 6.							
Mean	0.4996	1.1761	0.9896	0.5390	1.4979	1.9653	2.3950	0.7114
Bias	-0.0004	0.1761	-0.0104	0.0390	-0.0021	-0.0347	-0.1050	-
RMSE	0.0519	1.2503	0.0788	1.1967	0.0764	0.5445	0.6664	-
K-S	0.0267**	0.0286**	0.0312**	0.0317**	0.0277**	0.0669	0.0638	-
	Case 7.							
Mean	0.7434	0.9739	1.0019	0.4123	1.5071	1.9167	1.9796	0.6630
Bias	-0.0066	-0.0261	0.0019	-0.0877	0.0071	-0.0833	-0.5204	-
RMSE	0.0600	1.3336	0.0859	2.3752	0.1508	0.5809	1.2609	-
K-S	0.0427*	0.0171**	0.0193**	0.0445*	0.0464*	0.0614	0.1095	-
	Case 8.							
Mean	0.5159	0.9444	1.0047	0.8845	1.4905	1.8933	18.6509	0.9606
Bias	0.0159	-0.0556	0.0047	0.3845	-0.0095	-0.1067	-6.3491	-
RMSE	0.1079	1.7107	0.1120	6.2877	0.4071	0.8952	11.0147	-
K-S	0.0431*	0.0288**	0.0326**	0.0507*	0.0474 *	0.0708	0.0651	-
	Case 9.							
Mean	0.5020	1.0741	0.9969	0.4801	1.5001	1.8376	2.2657	1.5032
Bias	0.0020	0.0741	-0.0031	-0.0199	0.0001	-0.1624	-0.2343	-
RMSE	0.0760	0.6837	0.0283	0.6742	0.0285	0.7107	0.8799	-
K-S	0.0161**	0.0286**	0.0232**	0.0204**	0.0263**	0.0721	0.0625	-

Figure 2 QQ-Plot of the Estimates