A Mixture Model for Stock Prices

Bo Zhou *

November 22, 2013†

Abstract

Building on a Lucas tree asset pricing model, this paper relates the tail risk of asset prices to the component-density of a Normal-Laplace mixture distribution and proposes a new method to measure extreme event behavior in financial markets. The hidden state of the model represents the underlying state of the macroeconomy, which follows a two-state Markov regime switching process. Conditional on the state being "normal" or "extreme", the log dividend is subject to Normal or Laplace (fat-tailed) shocks respectively. The asset’s price is derived from discounted dividend values, where the stochastic discount factor is determined by the utility maximization of a representative agent who holds the asset. Finally, the identifiability of the model parameters, Maximum Likelihood estimation techniques and asymptotic properties of the MLE are discussed, and the estimation results are illustrated using S&P index returns.

*Department of Economics, University of Southern California
†The author thanks C. Hsiao, G. Ridder, S. Sakata, M. H. Pesaran (Department of Economics, USC), C. Jones (Marshall Scholl of Business, USC), T. W. Lee (Department of Economics, University of California, Riverside) and participants in USC econometrics seminar for comments. All errors are mine.
1 Introduction

Financial economists and econometricians have long been devoting their efforts to model tail events, both investigating their impact on financial asset prices theoretically and using real-world financial data to evaluate tail risk statistically. However, assessing tail risk is intrinsically difficult due to the infrequent nature of these extreme events. In the existing literature, selecting tail observations requires a “cutoff”, which is often arbitrarily taken since it is difficult to find an optimal rule to determine the cutoff. In this paper, we propose a new approach to model extreme events by using the Normal-Laplace mixture model to avoid the “cutoff” problem. Our work contributes to the asset pricing literature in two aspects. From a theoretical perspective, we build a foundation to capture the power-law property of the tail distribution of asset prices under an “extreme” state using the Lucas-tree model framework by assuming the Markov-Switching behavior of the underlying macroeconomic state. Statistically, we propose an easy-to-implement method to estimate the tail distribution by modeling stock returns as a mixture distribution of a Normal component and a fat-tailed component.

The asset pricing model in this paper follows the framework of Lucas (1978). We assume an endowment economy with a representative agent who holds assets producing non-storable dividends. In the basic model, there is a single asset in the economy; the representative-agent has the time-separable Constant Relative Risk Aversion (CRRA) utility; and shocks to dividend growth rates are a mixture of Log-Normal and Log-Laplace random variables. Under the Markov setting, stock prices have closed-form solutions given the assumptions on dividend growth rates, and the returns are described by a mixture of Normal-Laplace distributions. We then generalize the model to include multiple assets with correlated tail risks and the Epstein-Zin utility function, where the agent’s Intertemporal Elasticity of Substitution (IES) is
separated from the risk aversion.

We next discuss the estimation of the model (see Section 3). By using the time series returns data, we are able to identify the parameters in the density functions of the Normal and Laplace component, the transition probabilities of the Markov switching process and the difference between the expected returns in the next period conditional on the state being “normal” or “extreme”. We apply the forward-backward algorithm to estimate the transition matrix and the smoothed probabilities. After introducing the algorithm, we examine the indentifiability issues and asymptotic properties of the Maximum Likelihood Estimator (MLE).

In Section 4, we demonstrate the estimation results using S&P index returns. We first estimate the model parameters by MLE and then calculate the filtered probabilities of the Laplace component. Our results suggest that the probability of the “extreme” state is high during crises periods. The estimated transition matrix indicates persistant states and implies an expected duration of 88.3 months for the “normal” state and 13.3 months for the “extreme” state. Next we compare the Normal-Laplace mixture model with an alternative two-component Normal mixture model. The first model indicates a larger probability of transition into the “extreme” state, and its expected duration is longer.

Related Literature

This work is related to several branches of asset-pricing and financial econometrics literature. Following Lucas’s original work, the asset pricing model in this paper is related to a series of dividend-based multi-asset pricing models including Cochrane et al (2008) and Martin (2013). This paper differs from these works by working under a discrete-time setup and allowing time-dependence of the dividend process through the hidden Markov chain. Al-
though a closed form solution is generally not obtainable under the multi-
asset setting, we are able to discuss our model implications in the limit case
where one asset dominates the other in market share (intuitively, the orchard
consists of one large tree which is the market and another tiny tree).

This paper can be also viewed as an application of the Markov switching
specification to the “rare disaster” asset pricing theory. Since Barro (2006)
there has been a resurgence of interest in the rare-disaster explanation of
equity premium puzzle. Barro and Jin (2011) fitted the size distribution
of consumption disasters using a power law model to a larger database for
the GDP and consumption in 36 countries. Gourio (2008) extended the
model by allowing the probability of disasters to vary in time and take two
discrete values. Watcher (2013) also allowed time-varying disaster risk and
developed a model under a continuous-time framework to explain the equity
premium. This work differs from previous research by assuming that the
dynamics of the underlying state follow a “hidden” Markov chain, where the
state of the chain is unobservable by the agent. The agent thus perceives the
macroeconomy as a mixture of “normal” and “extreme” states.

This paper is also related to the literature in estimating tail risk. The
power-law decay rate of tail distributions has been documented frequently
in the literature, which can be traced back in time to Fama (1963), who
estimated the tail of financial returns in commodity markets by assum-
ing stable Paretian distribution. Some recent works include Bollerslev and
Todorov (2011), who assumed extreme-value distributions and developed a
non-parametric framework to estimate the “large” jumps from high-frequency
5-minute S&P 500 market data. Instead of using time-series data, Kelly
(2011) estimated tail risk from cross-sectional stock returns and stated that
this risk measure has strong predictive power for future market returns. Some
additional examples include modeling the distribution of jump sizes in jump-
diffusion models for option pricing and fitting tail innovations of the GARCH
model (McNeil and Frey (2003)). The above applications of power-law distribution focus on the estimation of the power-law exponent and often require a cutoff for tail observations. We adopt the same assumption of power-law behavior and assume Laplace distribution in extreme state. This paper differs from existing work by taking non-tail observations into account. We are able to avoid the problem of using an arbitrary cut-off for tail distribution by taking tail risk probability as the component-density of the mixture model and applying mixture models to estimate asset return distributions.

The statistical tool used in this paper belongs to finite mixture-models and Markov-switching models, which have been widely used in biostatistics, medicine and engineering due to their flexibility in fitting complex distributions. An introduction of estimation and inference of finite mixture models can be found in McLachlan and Peel (2001). In statistics literature the regime-switching models in which the states are unobservable are also documented as Hidden Markov Models (HMMs). An example of application of mixture-models in finance is Linden (2001), who estimated a mixture model of Normal-Laplace stock returns using maximum likelihood. Markov-Switching models have been used in modeling regimes of the business cycle (Hamilton (1989)), forecasting recession probabilities (Chauvet and Potter (2002)), modeling changes in policies, characterizing stock returns (Schaller and Van Norden (1997)) and forecasting volatility (Calvet and Fisher (2004)). Recent advances in inference in Hidden Markov models can be found in Cappe, Moulines and Ryden (2005), and applications of HMMs to finance can be found in Mamon and Elliott (2007).

The literature in likelihood based estimation methods of mixture models and Hidden Markov Models is extensive. For an introduction to estimation techniques of finite-state regime-switching models, see Hamilton (1994). Meanwhile, the theoretical study of asymptotic properties of the Maximum Likelihood Estimator has received increasing attention. Under the station-
arity assumption, the consistency of the Maximum Likelihood Estimator (MLE) for HMM was established by Leroux (1990), and the asymptotic normality of the MLE was first documented in Bickel, Ritov and Ryden (BRR) (1998). Le Gland and Mevel (LGM) (2000) developed an alternative approach to develop asymptotics of the likelihood function using the “exponential forgetting” property of the geometrically ergodic Hidden Markov Models. The LGM approach does not require the stationarity of the Hidden Markov Chain, however, the regularity conditions in LGM are more restrictive than those in BRR. We follow BRR in proving the asymptotic normality of the Maximum Likelihood estimator to avoid making the additional assumptions required in LGM.

2 Model

2.1 Basic Model Setup

In this section, we consider an endowment economy in which the asset is a tree that produces non-storable fruit.

2.1.1 The Agent’s Utility Maximization Problem

There is a representative agent who is endowed with 1 share of the tree at time 0 and trades the stock at each period after the dividend is distributed. She solves the utility maximization problem by choosing consumption $C_t$ and

\footnote{LGM requires that the density functions of the components converge to zero at the same rate at infinity, which is not satisfied in our model since the Normal distribution decays faster. We could overcome this by assuming the distribution functions to be truncated-Normal and truncated-Laplace, however, we avoid making this artificial assumption and follow BRR to prove the asymptotic normality.}
her share of the stock in the next period $\alpha_{t+1}$:

$$\max_{\{C_t, \alpha_{t+1}\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma}$$

subject to

$$\begin{cases} C_t + P_t \alpha_{t+1} = (P_t + D_t)\alpha_t \\ \alpha_0 = 1 \end{cases}$$

Under the normalization condition $\alpha_0 = 1$, the non-storable property of fruit implies that $C_t = D_t$ and $\alpha_t = 1$ in equilibrium.

The stochastic discount factor is the marginal rate of substitution between $t+1$ and $t$ contingent claims:

$$M_{t+1} = \frac{\beta u'(C_{t+1})}{u'(C_t)} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} = \beta \left( \frac{D_{t+1}}{D_t} \right)^{-\gamma}$$

### 2.1.2 The Underlying State of The Economy

At each period $t$, the economy may be in a normal state (1) or a “boom” or “disaster” state (2), which is unobservable. Let $S_t$ denote the state of the Markov chain, which follows a Markov Chain with the transition matrix

$$\Pi = \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{pmatrix}$$

where $\Pi_{11} + \Pi_{12} = \Pi_{21} + \Pi_{22} = 1$. $\Pi_{ij}$ is the transition probability from state $i$ to state $j$ if the current state is observable. If $\Pi_{22} > \Pi_{12}$, the states are persistent, i.e., if the economy is in State 2 today, then it has a higher probability to be in State 2 tomorrow.

We assume that the agent does not observe the state $S_t$, and her subjective probability of the economy in State 2 in the next period is $q_t$, which depends on her information set $\mathcal{F}_t$ at time $t$ and the underlying state ($S_t$) of the Markov-Switching process. If we assume that the agent’s information set is $\mathcal{F}_t = \{r_1, \ldots, r_t\}$, then $q_t = Pr(S_t = 2|r_1, \ldots, r_t)$ and can be calculated using
the forward-backward algorithm for Hidden Markov Models, which we will
discuss in detail in Section 3.

2.1.3 Dividend Process

The dividend \( \{D_t\} \) follows a hidden Markov Process, i.e., \( \{D_t\} \) is conditionally independent given \( \{S_t\} \). To capture the power-law behavior of stock returns documented in the literature, we assume that the distribution of \( D_t \) is Normal, conditional on \( S_t = 1 \), and is fat-tailed, conditional on \( S_t = 2 \):

\[
\ln \left( \frac{D_{t+1}}{D_t} \right) = g + \begin{cases} 
\epsilon_{t+1} & \text{in State 1} \\
v_{t+1} & \text{in State 2}
\end{cases},
\]

where \( \epsilon_t \sim N(0, \sigma^2) \), \( v_{t+1} \sim \text{Laplace}(0, b) \). We assume that the dividend growth rate \( g \) is constant (i.e. \( g \) does not depend on the state \( S_t \)) for the sake of simplicity as we are focusing on the fat-tail property of the asset prices under an economic “disaster”.

2.2 Asset Prices and Returns

2.2.1 Stock Prices

The stock price is determined by the first order condition of the agent’s utility maximization problem and the transversality condition:

\[
P_t = E_t [M_{t+1}(P_{t+1} + D_{t+1})],
\]

\[
\lim_{k \to \infty} E_t \beta^k u'(C_{t+k})P_{t+k} \alpha_{t+k} = 0.
\]

Under the hidden Markov assumption for the dividend process, we obtain the closed form solution for the stock price as:

**Proposition 1.** The price dividend ratio of the stock is

\[
P_t/D_t = (K_0 + K_1 q_t)
\]
where $K_0$ and $K_1$ are constants described as follows

$$K_0 = \frac{A[(\pi_{22} - \pi_{12})B - 1]}{\pi_{12}A(B - 1) + (1 - \pi_{22}B)(A - 1)}$$  \hspace{1cm} (2)$$

$$K_1 = \frac{A - B}{\pi_{12}A(B - 1) + (1 - \pi_{22}B)(A - 1)}$$  \hspace{1cm} (3)$$

and $A$ and $B$ are the expected dividend growth rates times the discount rate $\beta$ in state 1 and 2:

$$A = \beta \exp \left[-(\gamma - 1)g + \frac{(\gamma - 1)^2\sigma^2}{2}\right]$$  \hspace{1cm} (4)$$

$$B = \beta \exp \left[-(\gamma - 1)g\right] \frac{1}{1 - (\gamma - 1)^2\sigma^2}.$$  \hspace{1cm} (5)$$

We put the proof of this proposition and all the following proofs in Sections 2 and 3 of the Technical Appendix. Proposition 1 is derived from the intertemporal Euler equation and transversality condition, and the solution is found using the guess-and-verify method. The uniqueness of the solution was established in Lucas’s 1978 paper.

The price dividend ratio is a linear function of $q_t$, which is the agent’s estimation of State 2 probability in the next period $t+1$ given her information set at time $t$. One limiting case is when $q_t = 0$ and $\Pi_{12} = 0$, the price dividend ratio $P_t/D_t = A/(1 - A)$ is constant: we are back to the classic case where the dividend is i.i.d. Normally distributed. When $\gamma = 1$, the agent has the myopic log utility, and it is straightforward to verify that $P_t/D_t = \beta/(1 - \beta)$ no matter what the transition probabilities are.

### 2.2.2 Return distribution

After obtaining the price dividend ratio, we derive the asset’s return distribution as a direct corollary of Proposition 1.
Corollary 1. The conditional gross return distribution is described as a mixture of LogNormal and Log-Laplace distributions:

\[ R_{t+1} | F_t \sim \text{LogNormal}(\mu^N_{t+1}, \sigma^2) \text{ with probability } 1 - q_t \]

\[ R_{t+1} | F_t \sim \text{LogLaplace}(\mu^L_{t+1}, b) \text{ with probability } q_t \]

where \( R_{t+1} = (P_{t+1} + D_{t+1})/P_t \) denote the gross return at time \( t \),

\[ \mu^N_{t+1} = \ln(D_t/P_t) + g + \ln(1 + K_0 + K_1 \pi_{12}) , \]

\[ \mu^L_{t+1} = \ln(D_t/P_t) + g + \ln(1 + K_0 + K_1 \pi_{22}) , \]

\( \sigma^2 \) and \( b \) are the parameters related to the variances of the LogNormal and Log-Laplace distributions.

Using returns data, we are able to estimate the parameters of the distribution functions (\( \sigma \), \( b \)) and the growth rate \( g \) of the log dividend. However, the dependence of conditional return on the agent’s preference parameters (the discount rate \( \beta \) and the coefficient of risk aversion \( \gamma \)) is reflected through the constants \( K_0 \) and \( K_1 \), therefore \( \beta \) and \( \gamma \) are not identified. We leave the detailed discussion to Section 3.

2.3 Extensions of the Model

2.3.1 Risk Free Asset

Consider the case where there is a risk-free bond that pays 1 unit of the fruit in any state in the next period. When \( t = 0 \), the agent is endowed with no bonds (\( z_0 = 0 \)). The agent’s utility maximization problem is:

\[ \max_{\{C_t, \alpha_{t+1}, \pi_{t+1}\}} E_0 \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma} \]
subject to
\[
\left\{ \begin{array}{l}
C_t + P_t \alpha_{t+1} + R_t^f z_{t+1} = (P_t + D_t) \alpha_t + z_t \\
\alpha_0 = 1 \\
z_0 = 0
\end{array} \right.
\]

In equilibrium, \( z_t = 0 \) and the risk-free rate is
\[
R_t^f = \frac{1}{E_t(M_{t+1})} = \frac{1}{(1 - q_t)e^{-\gamma g + \frac{\sigma^2}{2}} + q_t \frac{e^{-\gamma g}}{1 - \gamma \sigma^2}}.
\]

2.3.2 Multiple Stocks

Consider a Lucas “orchard” where there are multiple risky assets (“trees”). The dividend stream of stock \( i \) follows the process:
\[
\ln \left( \frac{D_{i,t+1}}{D_{i,t}} \right) = g_i + \begin{cases} 
\epsilon_{i,t+1} & \text{in State 1} \\
v_{i,t+1} & \text{in State 2}
\end{cases}
\]

In general, the closed form solution of stock prices is not obtainable since the stochastic discount factor is a function of \( \sum_i D_{i,t+1}/(\sum_i D_{i,t}) \) and discounted values of dividend streams are difficult to calculate in our discrete-time, Hidden-Markov setting.

For illustration, we consider the special case where there are two stocks, and the market share of one stock is negligible to the other. An intuitive explanation is that one large “tree” is the stock market, and we consider the pricing of a small stock whose market share is negligible.

In the limit \( D_{2,t}/(D_{1,t} + D_{2,t}) \to 0 \), \( C_t \cong D_{1,t} \), asset 2 does not affect the pricing of asset 1. How does the price of asset 2 behave? We illustrate by assuming that the two stocks have the same dividend growth rate \( g \), and \( \epsilon_{1,t+1} \) and \( \epsilon_{2,t+1} \) are independent in state 1. Consider the following two cases under state 2: (i) \( v_{1,t+1} \) and \( v_{2,t+1} \) are independent (dividends are correlated only through the underlying hidden Markov process); (ii) \( v_{1,t+1} \) and \( v_{2,t+1} \) are perfectly correlated, \( v_{2,t+1} = \lambda v_{1,t+1} \) (asset prices are more highly correlated.
during volatile periods). Under power utility, the price of asset 2 can be calculated using the discount-factor pricing formula as follows:

**Proposition 2.** The price dividend ratio of stock 2 is

\[ \frac{P_{2,t}}{D_{2,t}} = (K_{2,0} + K_{2,1}q_t) \]  

where

\[ K_{2,0} = \frac{A_2[(\pi_{22} - \pi_{12})B_2 - 1]}{\pi_{12}A_2(B_2 - 1) + (1 - \pi_{22}B_2)(A_2 - 1)} \]

\[ K_{2,1} = \frac{A_2 - B_2}{\pi_{12}A_2(B_2 - 1) + (1 - \pi_{22}B_2)(A_2 - 1)} \]

and

\[ A_2 = \beta \exp(-\gamma g + \frac{\gamma^2 \sigma_1^2 + \sigma_2^2}{2}) \]

\[ B_2 = \beta \exp(-\gamma g) \frac{1}{1 - \gamma^2 b_1^2} \frac{1}{1 - b_2^2} \text{ under (i)}, \]

\[ B_2 = \beta \exp(-\gamma g) \frac{1}{1 - (\lambda - \gamma)^2 b_1^2} \text{ under (ii)}. \]

### 2.3.3 Epstein-Zin Preferences

The standard time-separable constant relative risk aversion (CRRA) preference implies that risk aversion is equal to the reciprocal of intertemporal elasticity of substitution (IES). Epstein and Zin (1989) introduced a new class of utility functions to decouple IES from risk aversion. The Epstein-Zin utility function is defined recursively as:

\[ U_t = \left[ (1 - \beta)C_t^{1-\rho} + \beta \left( E_t[U_{t+1}^{1-\rho}] \right)^{\frac{1-\rho}{\gamma}} \right]^{\frac{1}{1-\rho}}, \]

where IES = 1/\rho.

When \( \rho = \gamma \), we have the standard time-separable case as in Section 1.1. When \( \rho = 1 \), the following limit is used:

\[ U_t = C_t^{1-\beta} \left( E_t[U_{t+1}^{1-\rho}] \right)^{\frac{\beta}{\gamma}} \]  

(7)
The stochastic discount factor with Epstein-Zin preference is:

\[ M_{t+1} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\rho} \left( \frac{U_{t+1}}{R_t(U_{t+1})} \right)^{\rho-\gamma} \]

where \( R_t(U_{t+1}) = \left[ E_t(U_{t+1}^{1-\gamma}) \right]^{1/(1-\gamma)} \).

Finding closed-form solutions of asset prices with the Epstein-Zin utility is difficult since the discount factor is a function of the recursive value function \( U_t \). Under special cases where \( \rho = 1 \), the value function can be exactly solved as a linear function of \( C_t \).

**Lemma 1.** Given the Epstein-Zin utility function with \( \rho = 1 \), and the consumption process as \( C_t = D_t \), the value function is

\[ U_t = f(q_t)C_t \]  

(8)

where \( f(q_t) \) is described by the functional equation:

\[ f(q_t) = e^{\beta g \left[ \frac{1}{1-q_t} f(\pi_{12})^{1-\gamma} \frac{1}{e^{(1-\gamma)^2b^2}} + q_t f(\pi_{22})^{1-\gamma} \frac{1}{1-(1-\gamma)^2b^2} \right]} \]

We thus have the following form of stock prices:

**Proposition 3.** The price of the stock with Epstein-Zin preferences (\( \rho = 1 \)) is

\[ P^{EZ}_t = D_t \cdot h(q_t) \]

where \( h(q_t) \) satisfies the following equation:

\[ h(q_t) = \frac{\beta \left\{ (1-q_t)f(\pi_{12})^{1-\gamma}h(\pi_{12})e^{\frac{(1-\gamma)^2b^2}{2}} + q_t f(\pi_{22})^{1-\gamma}h(\pi_{22})e^{\frac{(1-\gamma)^2b^2}{2}} \right\}}{(1-q_t)f(\pi_{12})^{1-\gamma}h(\pi_{12})e^{\frac{(1-\gamma)^2b^2}{2}} + q_t f(\pi_{22})^{1-\gamma}h(\pi_{22})e^{\frac{(1-\gamma)^2b^2}{2}}} \]  

(9)

**Corollary 2.** The conditional gross return \( R^{EZ} \) described by a mixture of LogNormal and Log-Laplace distributions:

\[ R^{EZ}_{t+1} | F_t \sim LogNormal(\mu^{N,EZ}_{t+1}, \sigma^2) \]  

with probability \( 1 - q_t \)

\[ R^{EZ}_{t+1} | F_t \sim LogLaplace(\mu^{L,EZ}_{t+1}, b) \]  

with probability \( q_t \)
3 Estimation

We discuss the estimation methodology of the model parameters using time-series data of returns in Section 3. We start by a brief description of the statistical tool (Hidden Markov Model) in Section 3.1. Identifiability of the model is discussed in Section 3.2. Section 3.3 introduces the forward-backward algorithm, one of the most commonly used methodology to fit Hidden Markov Models. Section 3.4 discusses the asymptotic properties of the Maximum Likelihood Estimator (MLE).

3.1 Hidden Markov Process and its Stationary Distribution

The starting point of our statistical analysis is equation (1), and the parameter of interest is \( \theta_i = (\sigma_i, b_i, g_i, K_0, K_1, \Pi) \). We omit the subscript \( i \) henceforth for simplicity of notation. The aim of the estimation is to infer model parameter \( \theta \), including distributional parameters and the transition probabilities of the underlying Markov process, from the observed stock prices.

Taking the logarithm of equation (1) and calculating the difference in time, we get

\[
\Delta \ln(P_{t+1} + D_{t+1}) = \ln \left( \frac{1 + K_0 + K_1 q_{t+1}}{1 + K_0 + K_1 q_t} \right) + g + \begin{cases} 
\epsilon_{t+1} & \text{in state 1} \\
v_{t+1} & \text{in state 2}
\end{cases}
\]

The distribution of \( r_t \) depends on the state variables \( S_t \) and \( S_{t-1} \). We can define a new state variable

\[
\tilde{S}_t = \begin{cases} 
1 & \text{if } S_t = 1, S_{t-1} = 1 \\
2 & \text{if } S_t = 1, S_{t-1} = 2 \\
3 & \text{if } S_t = 2, S_{t-1} = 1 \\
4 & \text{if } S_t = 2, S_{t-1} = 2
\end{cases}
\]
and \( \tilde{S}_t \) is governed by the following Markov switching process:

\[
\tilde{\Pi} = \begin{pmatrix}
\Pi_{11} & 0 & \Pi_{12} & 0 \\
\Pi_{11} & 0 & \Pi_{12} & 0 \\
0 & \Pi_{21} & 0 & \Pi_{22} \\
0 & \Pi_{21} & 0 & \Pi_{22}
\end{pmatrix}.
\]

The process \( \{\tilde{S}_t, r_t\} \), in which the state of the Markov process \( \{\tilde{S}_t\} \) is not observed, and inference is made using observable \( \{r_t\} \), belongs to the Hidden Markov Model (HMMs) in the statistical literature: \( \{\tilde{S}_t\} \) is a Markov chain on a 4-state space, \( \{r_t\} \) is not a Markov chain, but is conditionally independent given \( \{\tilde{S}_t\} \) so the joint chain \( \{\tilde{S}_t, r_t\} \) is Markov.

Throughout this section we assume:

**Assumption 1.** The Markov chain \( \{\tilde{S}_t\} \) is irreducible and aperiodic.

**Assumption 2.** The process \( \{(\tilde{S}_t, r_t)\} \) is stationary.

Assumption 1 ensures that \( \{\tilde{S}_t\} \) has a stationary distribution \( \pi^* \), where

\[
\pi^* = (\pi_1^*, \pi_2^*, \pi_3^*, \pi_4^*) = \left( \frac{\Pi_{11}\Pi_{22}}{\Pi_{12} + \Pi_{22}}, \frac{\Pi_{12}\Pi_{22}}{\Pi_{12} + \Pi_{22}}, \frac{\Pi_{12}\Pi_{21}}{\Pi_{12} + \Pi_{22}}, \frac{\Pi_{12}\Pi_{21}}{\Pi_{12} + \Pi_{22}} \right)
\]

is the stationary probability density of \( \tilde{S} \).

The joint chain has a stationary distribution \( \tilde{\Pi} \)

\[Pr(\tilde{S}_t = s, r_t \in dr; \theta) \equiv f(s, dr; \theta) = \pi^*_s f_s(r; \theta) dr,\]

where \( s = 1, \ldots, 4 \), \( f_1(r; \theta) = n(r|\mu_1, \sigma) \), \( f_2(r; \theta) = n(r|\mu_2, \sigma) \), \( f_3(r; \theta) = l(r|\mu_3, b) \), \( f_4(r; \theta) = l(r|\mu_4, b) \).

Assumption 2 can be relaxed if the process \( \{\tilde{S}_t, r_t\} \), the prediction filter and its derivative converge to a stationary distribution at a geometric rate.

\[\text{For a rigorous definition of stationary distribution of the Hidden Markov process, see Cappe, Moulines and Ryden 2005, Page 560.}\]
(Le Gland and Mevel 1997). This requires that the ratio of the component densities satisfy certain integrability conditions, which are not satisfied in our setup due to the different convergence rates of the Normal and Laplace density functions at infinity. Technically, we could assume an upper bound for $r_i$ to avoid this violation. To keep the proof simple we avoid making additional assumptions, instead, we tested different probability distributions of the initial state $S_0$ as a robustness check. The numerical estimation results suggest that the estimated parameter values are reasonably robust to changes in the initial distribution.

### 3.2 Identifiability of Parameters

Identifiability issues of mixture models often arise because of the label-switching problem if the components are from the same family of distributions. For example, for the mixture of two normal densities, if we interchange the labels of the two components, the models are equivalent. In equation (10) there is no label-switching problem as the underlying distributions are from different families.

Equation (12) and (13) suggest that the parameters $K_0$, $K_1$ can not be identified simultaneously since the likelihood function depends on $\ln \frac{1+K_0+K_1\Pi_{12}}{1+K_0+K_1\Pi_{22}}$. Let $\delta = -\ln \frac{1+K_0+K_1\Pi_{12}}{1+K_0+K_1\Pi_{22}}$, then $\mu_2 = g - \delta$ and $\mu_3 = g + \delta$; $\mu_1 = \mu_4 = g$, which implies that the conditional expected return at $t$ is equal if the economy is at the same state at $t - 1$ and $t$, but will differ if the state switches at $t$. So the identified parameters are $(\sigma, b, g, \delta, \Pi_{12}, \Pi_{22})$.

We first discuss identifiability of the Normal-Laplace mixture model in one dimension. The class of distribution functions $\mathcal{H}$ is restricted to a mixture of Normal and Laplace distributions:

$$\mathcal{H} = \{H_\pi(x) : H_\pi(x) = \pi n(x|\mu, \sigma) + (1 - \pi) l(x|\mu', b)\} \quad (11)$$

We define identifiability of $\mathcal{H}$ as:
Definition $\mathcal{H}$ is identifiable if $H_{\pi_1}(x) = H_{\pi_2}(x)$ $m$-$a.e.$ $x$ if and only if $\pi_1 = \pi_2$ and $\phi_1 = \phi_2$, where $\phi = (\mu, \sigma, \mu', b) \in \Phi$ and $m$ is the Lebesgue measure.

Following the procedures in Teicher (1963), it is verified that the class of models defined in (11) satisfies the above condition. Teicher's work in 1967 generalizes identifiability to the product densities and showed that the class $H_n(\phi)(x) = H_{\phi_1}(x_1) \cdots H_{\phi_n}(x_n)$ with parameter $\phi \in \Phi^n$ is identifiable. In the hidden Markov Model, $\phi$ is now a function of the state variable $\phi(S_t)$, and identifiability in the product measure implies

$$Pr_{\theta^*} \{ (\phi_*(S_1), \ldots, \phi_*(S_T)) \in A \} = Pr_\theta \{ (\phi(S_1), \ldots, \phi(S_T) \in A) \}$$

where $\theta = (\sigma, b, g, \delta, \Pi_{12}, \Pi_{22})$. This leads to $\theta = \theta_*$ in the hidden Markov model.

3.3 Maximum Likelihood Estimator of the Markov-Switching Mixture Model

The Markov assumption has made it possible for the econometrician to write down the closed form of the likelihood function, since her estimate of $q_t$ is the transition probability conditional on the underlying state, i.e. $q_t = \Pi_{ss'}$ if $S_t = s$ for $s = 1, 2$. The conditional likelihood of $r_t$ can be written as:

$$L(r_t|r_1, \ldots, r_{t-1}) = \sum_{s=1}^{2} \sum_{s'='1}^{2} L(r_t|S_t = s, S_{t-1} = s', r_1, \ldots, r_{t-1}) Pr(S_t = s, S_{t-1} = s'|r_1, \ldots, r_{t-1})$$

$$= \sum_{s=1}^{2} \sum_{s'=1}^{2} L(r_t|S_t, S_{t-1}) Pr(S_t = s, S_{t-1} = s'|r_1, \ldots, r_{t-1})$$

The log likelihood of the model given $r_1, \ldots, r_T$ is written as a summation
of the conditional probabilities:

\[ \ln L(r_1, \ldots, r_T; \theta) = \sum_{t=1}^{T} \ln L(r_t | r_1, \ldots, r_{t-1}; \theta) \]

\[ = \sum_{t=1}^{T} \ln \sum_{s=1}^{4} L(r_t | \tilde{S}_t = s, \theta) Pr(\tilde{S}_t = s | r_1, \ldots, r_{t-1}, \theta) \]

\[ = \sum_{t=1}^{T} \ln \left[ Pr(\tilde{S}_t = 1 | r_1, \ldots, r_{t-1}) n(r_t | \mu_1, \sigma) + Pr(\tilde{S}_t = 2 | r_1, \ldots, r_{t-1}) n(r_t | \mu_2, \sigma) \right. \]
\[ \left. + Pr(\tilde{S}_t = 3 | r_1, \ldots, r_{t-1}) l(r_t | \mu_3, b) + Pr(\tilde{S}_t = 4 | r_1, \ldots, r_{t-1}) l(r_t | \mu_4, b) \right] , \]

where \( n(r | \mu, \sigma) \) and \( l(r | \mu, b) \) are the density function of the Normal and Laplace distributions and

\[ \begin{cases} \mu_1 = \mu_4 = g \\ \mu_2 = g + \ln \frac{1 + K_0 + K_1 \Pi_{12}}{1 + K_0 + K_1 \Pi_{22}} \\ \mu_3 = g + \ln \frac{1 + K_0 + K_1 \Pi_{21}}{1 + K_0 + K_1 \Pi_{11}} \end{cases} \] (13)

We apply Hamilton’s (1989) filtering technique to estimate \( Pr(\tilde{S}_t = s) \) in the likelihood function. The algorithm is as follows.

1. Start with a guess for the initial probability \( Pr(\tilde{S}_0 = s) \) for \( s = 1, \ldots, 4 \) and a guess for initial parameter values. One would expect that the asymptotic properties of the Maximum Likelihood estimator do not depend on the initial probability under certain regularity conditions, as we will discuss in Section 3.4.

In practice we use the steady-state probability as the initial guess, Normally fit the log returns to obtain initial value \( g_0 \) of log dividend growth rate and the initial \( \sigma_0^2 \) for the Normal density function, and exponentially fit the absolute value of log returns obtain the initial value \( b_0 \) in the Laplace density function. We select \( \delta_0 = 0 \) as the initial
value for \( \delta \). We choose \( \Pi_0 = \begin{pmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{pmatrix} \) as the initial values for the transition matrix.

2. For \( t = 1, \ldots, T \), calculate \( Pr(\tilde{S}_t = s) \) iteratively using the following formula:

\[
Pr(\tilde{S}_t = s| r_1, \ldots, r_{t-1}) = \sum_{s'=1}^{4} \tilde{\Pi}_{s's} Pr(\tilde{S}_{t-1} = s'| r_1, \ldots, r_{t-1})
\]  
\[ (14) \]

\[
Pr(\tilde{S}_t = s| r_1, \ldots, r_t) = \frac{L(r_t|\tilde{S}_t = s) Pr(\tilde{S}_t = s| r_1, \ldots, r_{t-1})}{\sum_{s'=1}^{4} L(r_t|\tilde{S}_t = s') Pr(\tilde{S}_t = s'| r_1, \ldots, r_{t-1})}
\]  
\[ (15) \]

The log likelihood in equation (12) is maximized to obtain the estimated parameter value \( \hat{\theta}_T \). After finding the estimated parameter values, we apply the forward-backward algorithm to infer the smoothed probabilities at each time based on full-sample information. The steps are as following:

1. For \( t = 1, \ldots, T \), calculate the filtered probabilities \( \hat{Pr}(\tilde{S}_t = s| r_1, \ldots, r_t, \hat{\theta}) \) using equations (14) and (15).

2. For \( t = T, T-1, \ldots, 1 \), the smoothed probabilities can be obtained by backward iteration starting from observation at \( T \):

\[
Pr(\tilde{S}_t = s| r_1, \ldots, r_T) = \sum_{s'=1}^{4} Pr(\tilde{S}_{t+1} = s'| r_1, \ldots, r_T) \hat{Pr}(\tilde{S}_t = s| r_1, \ldots, r_t, \hat{\theta}) \tilde{\Pi}_{s's}
\]  
\[ (16) \]

The smoothed probabilities \( Pr(\tilde{S}_t = s| r_1, \ldots, r_T) \) can be explained as the econometrician’s posterior knowledge of the economy’s state given information at time \( T \).
3.4 Asymptotic Properties of Maximum Likelihood Estimator

3.4.1 Regularity Conditions

Throughout the proof we assume the following regularity conditions hold:

**Assumption 3.** (i) The parameter set $\Theta$ is compact, and the true parameter $\theta_0$ lies in the interior of $\Theta$. (ii) $\sigma, b$ is bounded away from zero.

The compactness assumption (i) assumes there are known bounds of parameters. It is also assumed that $\theta_0$ is away from the boundary, so that the likelihood function can be Taylor expanded in the neighborhood of $\theta_0$. This condition also ensures that the transition matrix $\tilde{\Pi}$ is primitive (with index of primitivity 2), thus the process $\{(\tilde{S}_t, r_t)\}$ is ergodic. (ii) rules out singularities of the density function and ensures that the mapping $\theta \rightarrow f_s(r; \theta)$ is analytic.\[3\]

3.4.2 Asymptotic Normality

The proof of asymptotic normality takes the standard approach of establishing (i) almost sure convergence of the log-likelihood to a limit function, followed by (ii) central limit theorem (CLT) for the score function and (iii) law of large numbers (LLN) of the observed information matrix.

The identifiability condition, condition (i) and compactness of the parameter set $\Theta$ lead to strong consistency of the ML estimator, which is proved in [Leroux (1990)] by applying Kingman’s theorems for subergodic sequences to the generalized Kullback-Leibler divergence. Convergence of the score function and the information matrix to a limit in $L^2$-norm is obtained by writing the likelihood into summations using an identity for missing data in computation, NaN and Inf appear due to rounding errors. We remove the likelihood values with NaN and Inf in the Matlab code.\[3\]
The summation then is approximated by the sum of a martingale difference sequence to apply CLT and LLN. Finally, the asymptotic distribution of $\hat{\theta}_T - \theta_0$ is derived using the Slustky’s theorem.

**Theorem 1.** Under Assumptions 1, 2 and 3 the MLE is asymptotically normal:

$$\hat{\theta}_T - \theta_0 \to N(0, I_0^{-1}) \text{ under } P_0.$$

where $I_0$ is the Fisher information matrix under the true parameter value $\theta_0$.

**Proof.** See appendix.

4 **Analysis of S&P Index**

In this section, we apply the methodology described above to monthly S&P index data.

4.1 **Data**

Monthly S&P index data from Dec 1925 to Dec 2011 (1033 observations) was obtained from the Center for Research in Security Prices (CRSP) database. A summary of the data is in the following table.

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P index level</th>
<th>Returns on S&amp;P index</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Obs.</td>
<td>1,033</td>
<td>1,032</td>
</tr>
<tr>
<td>min</td>
<td>4.43</td>
<td>-0.299</td>
</tr>
<tr>
<td>max</td>
<td>1549.38</td>
<td>0.422</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.00601</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td>0.00306</td>
</tr>
<tr>
<td>Skewness</td>
<td></td>
<td>0.3034</td>
</tr>
<tr>
<td>Kurtosis</td>
<td></td>
<td>12.16</td>
</tr>
</tbody>
</table>
4.2 Results

4.2.1 Summary

The filtered probabilities $\hat{P}(S_t = 2)$ are plotted in Figure 1. The probability of State 2 is high during the great depression, especially in time periods 1929-1933 and 1937-1940. The figure also captures events such as the “flash crash” in May 1962, the bear market in 1973-1974, the Russian financial crisis in 1998, the Oct. 1987 market crash, the market downturn in Sep 2001 and Oct 2002, and also the Oct. 2008 crash during the recent financial crisis.

The estimated transition probability matrix is

$$\Pi = \begin{pmatrix} 0.9887 & 0.0113 \\ 0.0755 & 0.9245 \end{pmatrix}.$$

As expected, if the economy is at state 1 (or 2) today then it has a higher probability to remain in state 1 (or 2) tomorrow. The expected duration of a “normal” state is $1/\Pi_{12} = 88.3$ months and the expected duration of a “extreme” state is $1/\Pi_{21} = 13.3$ months.

The other estimated parameter values are $\hat{\sigma} = 0.0381$, $\hat{b} = 0.0841$, $\hat{g} = 0.0074$ and $\hat{\delta} = -0.0689$. $\delta < 0$ implies that $K_1 < 0$, i.e., the price-dividend ratio is low when $q_t$ is high.

4.2.2 Comparison with the Normal Mixture Model

One key assumption in our model is that dividend distribution is non-Normal under the “extreme” state. Instead of testing the distributional assumption, we estimate the model parameters assuming the distribution of dividend under state 2 is also Normal with higher variance and compare with the results of the Normal-Laplace mixture model. The estimated transition probabilities are:

$$\Pi_{\text{normmix}} = \begin{pmatrix} 0.9820 & 0.0180 \\ 0.3550 & 0.6450 \end{pmatrix}.$$
The estimated standard deviation of the estimated components are \( \hat{\sigma}_1 = 0.0413 \) and \( \hat{\sigma}_2 = 0.1858 \). The estimated \( \hat{g} = 0.0075 \) is close to the Normal-Laplace mixture model result, however, \( \hat{\delta} = 0.07 \) leads to the counterintuitive explanation: the price-dividend ratio is higher when the probability of the “extreme” state is high.

Figure 2 plots the stationary density function of returns to S&P 500 of Normal-Laplace mixture model, two-component Normal mixture model and the kernel density. The Normal-Laplaces fits the distribution better than the mixture of Normal distributions. To capture the asymmetric property of returns data, a possible extension of our model is to incorporate an asymmetric distribution for “boom” and “disaster” states respectively.

Although mixtures of Normal distribution are sufficient to generate the fat-tail property, in our model the assumption of the Normal-Laplace mixture distribution has non-trivial implications. Compared with the Normal-Laplace mixture model, the Normal mixture model predicts a lower probability and a shorter duration of state 2. The expected length of a “normal” state is 55.6 months and the expected length of an “extreme” state is 2.8 months. The NBER recession records of average contraction length are 18.2 months (1919-1945) and 11.1 months (1945-2009), which are closer to the Normal-Laplace mixture model results. For a comparison of the filtered probabilities of state 2 for both models, see figure 3.

We evaluate the probability forecasting of recession by calculating the Brier Score:

\[
BS = \frac{1}{T} \sum_{t=1}^{T} (REC_{NBER,t} - \hat{q}_t)^2,
\]

where \( REC_{NBER,t} \) is a monthly dummy variable indicating the NBER recession periods. The calculated Brier Score for the Normal-Laplace mixture model is 0.155, and for the Normal mixture model, the corresponding score is 0.184.
4.2.3 Estimating Preference Parameters Using Price-Dividend Ratios

After finding the transition matrix and the parameters in the Normal and Laplace distribution functions, we use the price-dividend ratio data to estimate the agent’s preference parameters: discount rate $\beta$ and risk aversion coefficient $\gamma$.

Equation (1) states that $P_t/D_t$ ratio is a linear function of $q_t$. We run the weighted least squares regression of the historical S&P price-dividend ratio data (downloaded from R. Shiller’s website\footnote{www.econ.yale.edu/~shiller/data.htm}) on the estimated $\hat{q}_t$ to find $\hat{K}_0$ and $\hat{K}_1$:

\[ \begin{array}{ccc}
\hat{K}_0 & 28.0 & 0.36 \\
\hat{K}_1 & -10.1 & 1.22 \\
\end{array} \]

Next we substitute $\hat{K}_0$, $\hat{K}_1$, and the estimated $\hat{\Pi}$ and $\hat{g}$, $\hat{\sigma}$, $\hat{b}$ from returns data into equations (2), (3), (4), (5) to solve for the constants $A$, $B$, and the parameters $\beta$, $\gamma$ in Proposition 1. The estimated values are:

\[ \begin{array}{cccc}
A & B & \hat{\beta} & \hat{\gamma} \\
0.969 & 0.910 & 0.984 & 3.809 \\
\end{array} \]

Using the simplified two-step analysis above, we found that the estimated risk aversion parameter is in a reasonable range under the power utility setting.

4.2.4 Volatility Forecasting

The Markov-switching property of the model implies volatility clustering property since the underlying states are persistent. In this section, we com-
pare 1-peroid volatility forecasting of the model with the prevalent GARCH(1,1) model.

Corollary[1] implies that the 1-period prediction of the mean and variance of return is:

$$\mu_{t+1} \equiv E(r_{t+1}|\mathcal{F}_t) = (1 - \hat{q}_t)\mu_{t+1}^N + \hat{q}_t\mu_{t+1}^L$$

$$\sigma_{t+1}^2 \equiv E((r_{t+1} - \mu_{t+1})^2|\mathcal{F}_t)$$

$$= (1 - \hat{q}_t)((\mu_{t+1}^N - \mu_{t+1})^2 + \sigma^2) + \hat{q}_t[(\mu_{t+1}^L - \mu_{t+1})^2 + 2b^2]$$

$$= (1 - \hat{q}_t)(\hat{q}_t^2\delta^2 + \sigma^2) + \hat{q}_t[(1 - \hat{q}_t)^2\delta^2 + 2b^2]$$

We use the Dec. 1925 to Dec. 2011 S&P index data to estimate the parameters of the Normal-Laplace mixture model and the GARCH(1,1) model, and use the Dec. 2011-Dec. 2012 data for forecasting. We use the returns data from time 1 to t to calculate 1-period out-of-sample prediction of $\hat{\sigma}_{t+1}$. The mean squared error is $2.02 \times 10^{-6}$ for the mixture model and $2.88 \times 10^{-6}$ for the GARCH(1,1) model.

5 Conclusion

We propose a mixture model to capture the fat-tail property of stock prices in this paper and estimated the model using S&P index in this paper. Our work can be extended in several directions.

One possible line of future research is to investigate the pricing of individual stocks by looking at both return and dividend data. The estimation technique in this paper can be directly applied to individual stocks to find out the transition matrix and model parameters separately; alternatively, transition probabilities estimated from the market can be used to price individual stocks.
Our model can also be extended by incorporating asymmetry and combining with the GARCH model to capture volatility clustering. Another direction is to explore the model’s implications on option pricing, especially far-out-of-the-money options, and compare with alternative methods such as jump-diffusion models.

The unsolved inference issues involved in this paper include testing the Normal-Laplace mixture versus various alternatives such as Normal mixtures and testing the number of components in the model. Another question is how the ML estimator behaves asymptotically when we relax Assumption and allow the true parameter values to be close to the boundary.

\footnote{The author thanks M. H. Pesaran for pointing out and elaborating on this issue.}
References


Figure 1: Laplace Component Density
Figure 2: Stationary distribution of the Normal-Laplace mixture model, two-component Normal mixture model vs. Kernel estimation of return distribution
Probably of the Pareto Component: Monthly data, Dec1925–Dec2010

Probably of Component 2 in the Normal–Mixture model: Monthly data, Dec1925–Dec2010

Figure 3: Probability of State 2, Normal-Laplace mixture model vs. Normal mixtures. NBER recession dates are plotted in the shaded areas.
6 Technical Appendix

6.1 Proof of Theorems in Section 2

6.1.1 Proof of Proposition 1

Proof. The first order condition and transversality condition imply that

\[
\frac{P_t}{D_t} = E_t \left[ \beta \frac{P_{t+1} + D_{t+1}}{D_t} \left( \frac{D_{t+1}}{D_t} \right)^{-\gamma} \right] = E_t \left[ \sum_{j=1}^{\infty} \beta^j \left( \frac{D_{t+j}}{D_t} \right)^{1-\gamma} \right].
\]  

(17)

Note that (17) is a linear function of \( q_t \):

\[
\frac{P_t}{D_t} = (1-q_t)E_t \left[ \sum_{j=1}^{\infty} \beta^j \left( \frac{D_{t+j}}{D_t} \right)^{1-\gamma} | S_{t+1} = 1 \right] + q_t E_t \left[ \sum_{j=1}^{\infty} \beta^j \left( \frac{D_{t+j}}{D_t} \right)^{1-\gamma} | S_{t+1} = 2 \right].
\]

Let \( P_t = K_0 + K_1 q_t \) and substitute into (17) to find the coefficients \( K_0, K_1 \):

\[
D_t * (K_0 + K_1 q_t) = E_t[M_{t+1}D_{t+1}(1 + K_0 + K_1 q_{t+1})]
\]

\[
= D_t \left\{ (1 - q_t)(1 + K_0 + K_1 \pi_{12})E [\beta e^{-\gamma (g + \epsilon_{t+1})} e^{g + \epsilon_{t+1}}] \right. \\
+ q_t(1 + K_0 + K_1 \pi_{22}) E [\beta e^{-\gamma (g + \nu_{t+1})} e^{g + \nu_{t+1}}] \right\}
\]

\[
= D_t \left\{ (1 - q_t)(1 + K_0 + K_1 \pi_{12}) / (1 - (\gamma - 1)^2 \sigma^2) \right. \\
+ q_t(1 + K_0 + K_1 \pi_{22}) \right\}
\]

This must hold for all \( q_t \) so \( K_0, K_1 \) satisfy the equations in Proposition 1. \( \square \)

6.1.2 Proof of Corollary 1

Proof. The gross return \( R_{t+1} \) is:

\[
R_{t+1} = (P_{t+1} + D_{t+1})/P_t = D_{t+1}(1 + K_0 + K_1 q_{t+1})/P_t
\]

\[
R_{t+1} | \mathcal{F}_t \sim \frac{D_t}{P_t} \left\{ \begin{array}{ll}
(1 + K_0 + K_1 \pi_{12}) \exp (g + \epsilon_{t+1}) & \text{w.p. } 1 - q_t \\
(1 + K_0 + K_1 \pi_{22}) \exp (g + \nu_{t+1}) & \text{w.p. } q_t
\end{array} \right. \]

\( \square \)
6.1.3 Proof of Lemma

Proof. From eq. (7),
\[ U_t = f(q_t)C_t = C_t^{1-\beta} \left( E_t(U_t^{1-\gamma}) \right)^{1/\gamma} = C_t^{1-\beta} \left( E_t(f(q_t^{1-\gamma}C_t^{1-\gamma})) \right)^{1/\gamma} \]
\[ = C_t e^{\beta g} \left[ (1 - q_t)f(\pi_{12})^{1-\gamma} e^{\frac{1-\gamma \pi_{12}^2 b^2}{2}} + q_t f(\pi_{22})^{1-\gamma} \frac{1}{1 - (1 - \gamma)^2 b^2} \right]^{1/\gamma} \]

6.1.4 Proof of Proposition

Proof. The Euler condition \( P_t = E_t[M_{t+1}(P_{t+1} + D_{t+1})] \) implies
\[ D_t h(q_t) = E_t[M_{t+1}D_{t+1}[1 + h(q_{t+1})]] \]
where
\[ M_{t+1} = \frac{\beta C_t}{C_{t+1}} \left[ \frac{f(q_{t+1})C_{t+1}}{R_t(U_{t+1})} \right]^{1-\gamma} \]
Substitute (8) into the above equation to obtain equation (9).

6.1.5 Proof of Corollary

Proof. The gross return \( R_{t+1} \) is:
\[ R_{t+1}^{EZ} = \frac{D_t}{P_t^{EZ}} \frac{(1 + h(\pi_{12})) \exp(g + \epsilon_{t+1})}{P_t^{EZ}} \]
\[ R_{t+1}^{EZ} | F_t \sim \frac{D_t}{P_t^{EZ}} \left\{ \begin{array}{ll}
(1 + h(\pi_{12})) \exp(g + \epsilon_{t+1}) & \text{w.p. } 1 - q_t \\
(1 + h(\pi_{22})) \exp(g + \epsilon_{t+1}) & \text{w.p. } q_t
\end{array} \right. \]
This is a mixture of Lognormal and Log-Laplace distributions with coefficients:
\[ \mu_t^{EZ} = \ln(D_t/P_t^{EZ}) + g + \ln(1 + h(\pi_{12})) \]
\[ k_t^{EZ} = \ln(D_t/P_t^{EZ}) + g + \ln(1 + h(\pi_{22})) \]
6.2 Theorems in Section

In the following section, we sketch the proof the theorem on asymptotic normality of the Maximum Likelihood Estimator. For a detailed reference, see Bickel, Ritov and Ryden (1998).

The proof consists of two parts: a central limit theorem for the score function and a law of large numbers for the observed information matrix in the following sections 6.2.2 and 6.2.3.

6.2.1 Notation

In this appendix the following short-hand notation is used to avoid lengthy equations.

\[ l_{1:t}(\theta) = \ln L(r_1, \ldots, r_t; \theta), \]
\[ l_{0:1:t-1}(\theta) = \ln L(r_t | r_1, \ldots, r_{t-1}; \theta). \]

We assume that the process \{(\tilde{S}_t, r_t)\} is stationary, therefore it can be extended to a doubly infinite sequence \{(\tilde{S}_t, r_t)\}_{t=-\infty}^{\infty}. It is also convenient to view \( \tilde{S}_t \) as missing data in Hidden Markov Models. In the following proof, we denote the complete information likelihood as:

\[ l_{1:3}(\tilde{S}; \theta) = \ln L(r_1, \ldots, r_t, \tilde{S}_1, \ldots, \tilde{S}_t; \theta). \]

6.2.2 Central Limit Theorem for the Score Function

This section proves the following lemma of CLT for the score function:

**Lemma 2.** Under Assumption 1, 2, and 3

\[
\frac{1}{\sqrt{T}} \nabla \theta l_{1:T}(\theta_0) \rightarrow N(0, I_0) \text{ as } T \rightarrow \infty.
\]

The main idea of the proof is to write the log of likelihood \( l_{1:T}(\theta_0) \) into the summation of conditional likelihood \( l_{0:1:t-1}(\theta) \) for each \( t \), and compare \( \{ \nabla \theta l_{0:1:t-1}(\theta) \} \) with a stationary sequence \( \{ \nabla \theta l_{0:1:t-1}(\theta) \} \).
Proposition 4.

\[ ||\nabla \theta_{l|1:t-1}(\theta) - \nabla \theta_{l|\infty:t-1}(\theta)||_2 \leq C \rho^t, \]

where \( C \) is a constant and \( 0 < \rho < 1 \) and \( || \cdot ||_2 \) denotes \( L_2 \)-norm under \( P_0 \).


Finally we prove Lemma 2 below:

\[ \sum_{t=1}^{T} \nabla \theta_{l|1:t-1}(\theta_0) \rightarrow N(0, I_0). \]

Proof. Since \( \{r_t\} \) is stationary \( \{\nabla \theta_{l|1:t-1}(\theta)\} \) is a stationary and ergodic sequence. Also \( E_0 \left[ \nabla \theta_{l|1:t-1}(\theta) \right] = 0 \), so \( \{\nabla \theta_{l|\infty:t-1}(\theta)\} \) is a martingale difference sequence. By central limit theorem for martingales,

\[ \frac{1}{\sqrt{T}} \nabla \theta_{l|\infty:t-1}(\theta) \rightarrow N(0, I_0). \]

By proposition 4,

\[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla \theta_{l|1:t-1}(\theta_0) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla \theta_{l|\infty:t-1}(\theta) \leq \frac{1}{\sqrt{T}} \sum_{t=1}^{T} ||\nabla \theta_{l|1:t-1}(\theta_0) - \nabla \theta_{l|\infty:t-1}(\theta_0)||_2 \rightarrow 0, \]

therefore

\[ \frac{1}{\sqrt{T}} \nabla \theta_{l|1:t-1}(\theta_0) \frac{1}{\sqrt{T}} \nabla \theta_{l|1:t-1}(\theta_0) \rightarrow N(0, I_0). \]

6.2.3 Law of Large Numbers for Observed Information Matrix

Lemma 3. Under Assumptions 1, 2 and 3

\[ \frac{1}{T} \nabla^2 \theta_{1:T}(\theta_T) \rightarrow -I_0 \text{ in probability under } P_0 \text{ as } T \rightarrow \infty \]

if \( \theta_T \rightarrow \theta_0 \) almost surely.

6.2.4 Proof of Theorem 1

Proof. Since \( \hat{\theta}_T \to \theta_0 \) almost surely, \( \hat{\theta}_T \) is an interior point of \( \Theta \) for large enough \( T \). Take the Taylor expansion of \( \nabla_\theta l_{1:T}(\hat{\theta}_T) \) and invert the equation:

\[
0 = \nabla_\theta l_{1:T}(\hat{\theta}_T) = \nabla_\theta l_{1:T}(\theta_0) + \nabla^2_\theta l_{1:T}(\hat{\theta}_T)(\hat{\theta}_T - \theta_0)
\]

\[
\sqrt{T}(\hat{\theta}_T - \theta_0) = \left[-\frac{1}{T} \nabla^2_\theta l_{1:T}(\hat{\theta}_T)\right]^{-1} \frac{1}{\sqrt{T}} \nabla_\theta l_{1:T}(\theta_0)
\]

By Lemmas 2 and 3 and the Slutsky theorem, Theorem 1 holds.