ESSAYS IN DYNAMIC PORTFOLIO OPTIMIZATION
AND DIFFUSION ESTIMATIONS

By

Hua He

B.S. Math., Fudan University
(1982)

Submitted to the School of Management
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Management

at the

Massachusetts Institute of Technology

August 1989

©Hua He, 1989. All rights reserved

The author hereby grants to MIT permission to reproduce and to
distribute copies of this thesis document in whole or in part.

Signature of Author

Certified by

<

Accepted by

School of Management
August, 1989

Chi-fu Huang
Professor of Finance
Thesis Supervisor

James B. Orlin
Chairman, Ph.D's Program
ESSAYS IN DYNAMIC PORTFOLIO OPTIMIZATION
AND DIFFUSION ESTIMATIONS

By

Hua He

August 1989

Submitted to the School of Management
in August, 1989 in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
at the
Massachusetts Institute of Technology

Abstract

This dissertation consists of three independent essays. Essay One employs a martingale representation approach to study a dynamic consumption-portfolio problem in continuous time with incomplete markets and short-sale constraints. The main idea of the martingale approach is to transform the dynamic consumption-portfolio problem into a static variational problem using a duality technique. Sufficient conditions are provided for a dynamic consumption-portfolio problem to have a solution. To characterize the optimal solution, the optimal consumption and portfolio policies are related to the solution of a quasi-linear partial differential equation (PDE). The quasi-linear PDE is easier to solve numerically than the Bellman equation obtained in dynamic programming.

Essay Two generalizes the Cox, Ross and Rubinstein [1979] binomial model, and develops a theory of convergence from discrete time multivariate and multinomial models to continuous time multi-dimensional diffusion models in the context of contingent claim pricing and consumption-portfolio selection. The key to our approach is to approximate the \( N \)-dimensional diffusion price process by a sequence of \( N \)-variate and \( N + 1 \)-nomial processes. It is shown that contingent claim prices and optimal consumption-portfolio policies derived from discrete time models converge to the corresponding contingent claim price and optimal consumption-portfolio policy of the limiting continuous time model. This result generalizes the binomial model in that it allows us to price contingent claims whose payoffs may involve more than one asset. Moreover, instead of solving a partial differential equation (PDE) with possibly more than one state variable, this approach provides a simpler numerical procedure for computing contingent claim prices and optimal consumption-portfolio policies.

Essay Three develops a generalized method of moments (GMM) estimation procedure for estimating parameters of general diffusion processes using discretely sampled data. When the functional forms of the conditional moments are known, GMM estimation is straightforward. When the functional form of conditional moments are unknown, a numerical approximation procedure is introduced, and GMM estimation is then applied by replacing the actual condi-
tional moments by their numerical approximations. Numerical approximations for conditional moments are obtained through approximating the continuous time diffusion process by a sequence of discrete time multinomial processes. The conditional moments computed from the discrete time processes are shown to converge to the actual conditional moments of the limiting diffusion process. Large sample properties are investigated for the GMM estimators obtained by this method.

Thesis Committee:

Chi-fu Huang, Professor of Finance, MIT (Chairman)
Andrew Lo, Associate Professor of Finance, MIT
Robert Merton, George Fisher Baker Professor of Business Administration,
Harvard University
Robert Pindyck, Mitsubishi Bank Professor of Applied Economics, MIT
ACKNOWLEDGEMENT

I am deeply indebted to Chi-fu Huang, my principle advisor, for his stimulating guidance and absolute commitment through all stages of my dissertation. Without Chi-fu's constant and caring support, this thesis would have been truly impossible. I am also deeply indebted to Robert Merton, whose insightful comments and suggestions inspired this thesis in many ways. Bob's commitment has always been a great encouragement to me. I would like to thank Andrew Lo and Robert Pindyck for giving their time to serve on the thesis committee. Their contributions to this thesis are evident. To me, Bob has been a caring advisor as well as a friend.

I wish to thank many members of the MIT faculty who freely gave me their time and attention during the four years at Sloan: Ernst Berndt, John Cox, Ann Friedlaender, David Scharfstein, and Tom Stoker. I also wish to thank all of our doctoral students in EFA area for their friendship and humor, in particular, Annette Beatty, Caterina Nelson, Vivian Ellis, Enrico Perotti, Raghuram Rajan, Mark Showalter and Steve Umlauf. Special thanks are due to Neil Pearson for sharing many of his research interests with me and to Ayman Hindy for his friendship and help.

I wish to express my deep appreciation to Mr. C. Tsang for his generous financial support during my first year at the Operations Research Center, and to the late Professor Ed Kuh for bringing me into the Sloan Ph.D program.

Finally, I am truly indebted to my parents and my parent-in-laws for their constant love and encouragement during such a long period of separation. Of course, the most important contributions were made by my beloved wife, Ding Ming. Her extraordinary loving and generous character made this endeavor possible. To her, I dedicate this thesis.
# Contents

I. Consumption and Portfolio Policies with Incomplete Markets and Short-sale Constraints:
   the Infinite Dimensional Case
   1. Introduction ........................................ 6
   2. Formulation ........................................ 8
   3. From a Dynamic Problem to a Static Variational Problem .............. 11
   4. Existence Theorem .................................. 22
   5. Characterizations and Computations .......................... 27
   6. Conclusion ......................................... 36
   Appendix.A ........................................... 39

II. Convergence from Discrete to Continuous Time Financial Models
   1. Introduction ........................................ 44
   2. Preliminaries on Weak Convergence ........................ 47
   3. Multinomial Approximation ............................. 49
   4. Contingent Claim Pricing ................................ 54
   5. Optimal Consumption and Portfolio Policies .................. 56
   6. Concluding Remarks ................................... 63
   Appendix.B.1 .......................................... 67
   Appendix.B.2 .......................................... 72

III. Method of Moments Estimation for Diffusion Processes
   1. Introduction ........................................ 75
   2. GMM Estimation .................................... 78
   3. Numerical Approximations for Conditional Moments .............. 80
   4. Large Sample Properties ................................ 83
   5. Empirical Example: Square Root Process ..................... 88
   6. Concluding Comments .................................... 93
   Appendix.C ........................................... 96
ESSAY ONE:
CONSUMPTION AND PORTFOLIO POLICIES WITH INCOMPLETE MARKETS AND SHORT-SALE CONSTRAINTS:
THE INFINITE DIMENSION CASE

Abstract This essay employs a martingale approach to study a dynamic consumption-portfolio problem in continuous time with incomplete markets and short-sale constraints. The main idea of the martingale approach is to transform the dynamic consumption-portfolio problem into a static variational problem using a duality technique. Sufficient conditions are provided for a dynamic consumption-portfolio problem to have a solution. To characterize the optimal solution, the optimal consumption and portfolio policies are related to the solution of a quasi-linear partial differential equation (PDE). The quasi-linear PDE is easier to solve numerically than the Bellman equation obtained in dynamic programming.

1. Introduction

In this essay we employ a martingale representation approach to study optimal intertemporal consumption and portfolio policies in continuous time with incomplete markets and short-sale constraints. The classical approach to the analysis of the consumption-portfolio problem in continuous time is the stochastic dynamic programming, the use of which in this context was pioneered by Merton [1969, 1971]. Recently Pliska [1986], Cox and Huang [1987a, 1987b], Karatzas, Lehoczky and Shreve [1987], and Pagès [1987] have used a martingale representation technology in place of the dynamic programming to study optimal intertemporal consumption and portfolio policies, while Chamberlain [1987], Duffie and Huang [1985], and Huang [1987] have used the martingale approach in a general equilibrium setting. However, to date, little is known about how the martingale representation technology may be useful in continuous time economies when markets are dynamically incomplete. With the exception of the paper by Pagès, those mentioned above assume that markets are dynamically complete, while Pagès makes an assumption about the asset price process and the nature of the incompleteness that excludes the cases of real interest.

In the martingale approach one solves a dynamic consumption-portfolio problem by separating it into two parts. First, one transforms the dynamic consumption-portfolio problem into a static variational problem and solves the static problem to get the optimal consumption bundle. Then, one applies the martingale representation theorem to determine the portfolio trading strategy needed to generate the optimal consumption bundle. Pliska [1986] explicitly carries out these computations for a simple example with consumption at a single terminal date, and also provides sufficient conditions for optimal policies to exist in a general continuous time stochastic environment in which asset prices are semi-martingales and consumption can be either positive or negative.

Cox and Huang [1987a] allow intermediate consumption, impose non-negativity constraints on consumption and provide a set of explicit conditions for existence weaker than those required
in dynamic programming. They find a unique system of Arrow-Debreu state prices (or, after normalization by the bond price, a unique equivalent martingale measure) that are consistent with the absence of arbitrage and consider the static problem of maximizing utility subject to a single budget constraint formed by these state prices. They show that the solution to the static problem and the portfolio policy needed to implement it are identical to the optimal consumption and portfolio policies given by dynamic programming. Karatzas, Lehoczky, and Shreve [1987] study the same problem with non-negativity constraints. However, they do not provide explicit sufficient conditions on an investor's utility function and on parameters of the price processes for existence. Cox and Huang [1987b] characterize the optimal policies and explicitly compute the optimal consumption and portfolio policies in certain situations in which it is difficult if not impossible to use dynamic programming. They also find that the value of optimally invested wealth satisfies a linear partial differential equation, which is much easier to solve than the nonlinear partial differential equation obtained in dynamic programming.

A limitation of the analyses of Pliska and Cox and Huang is that these authors assume that markets are dynamically complete. In the papers by Cox and Huang the completeness of markets gives a unique system of Arrow-Debreu state prices. These state prices are then used to form the budget constraint in the static problem that comprises the first part of the martingale approach. When markets are dynamically incomplete there are infinitely many number of Arrow-Debreu state prices that are consistent with the absence of arbitrage. The static problem of maximizing utility subject to the requirement that consumption be feasible with respect to budget constraints formed using all the state prices consistent with no arbitrage involves infinitely many budget constraints. Hence it is not immediate how standard optimization techniques can be used to attack the problem.

In this essay we use the martingale approach to study a dynamic consumption-portfolio problem in continuous time where we allow the parameters of the security price processes to depend upon a set of state variables, which are needed to describe completely the investment opportunities. In this setup, the price processes and the state variable processes form a multi-dimensional diffusion process. Since the risks associated with the state variables can not be completely hedged, markets are dynamically incomplete. More specifically, there exist contingent claims written on the security prices which can not be dynamically replicated by trading in traded assets. We also require that consumption be non-negative, and allow the possibility of short-sale constraints, i.e., no short-selling of risky securities or the riskless bond.

The basic idea of our approach is to transform the dynamic consumption-portfolio problem into a static variational problem, where one maximizes the expected utility over consumption

---

1 A similar dynamic consumption-portfolio problem in a finite dimensional discrete time economy with incomplete markets and short-sale constraints is studied in He and Pearson [1988].

2 For the purpose of existence, we could allow the security price processes to be Itô processes.
bundles that are feasible with respect to a single budget constraint formed by some Arrow-Debreu state prices consistent with no arbitrage. To do so, we establish a duality between the dynamic consumption-portfolio problem and a dual Arrow-Debreu state price problem. The duality theorem shows that the dynamic problem has a solution if the dual problem has a solution. The dual problem transforms the dynamic consumption-portfolio problem into a static variational problem. We provide sufficient conditions for the dynamic consumption-portfolio problem to have a solution. To characterize the optimal solution, we relate the optimal consumption and portfolio policies to the solution of a quasi-linear partial differential equation (PDE). The quasi-linear PDE is easier to solve numerically than the Bellman equation obtained in dynamic programming.

The balance of this essay is organized as follows. In Section 2 we formulate a dynamic consumption and portfolio problem with incomplete markets. Section 3 establishes an equivalence between the dynamic consumption-portfolio problem and a static variational problem using a duality technique. In Section 4, we provide sufficient conditions on utility functions for the dynamic problem to have a solution. The characterizations and computations of optimal consumption and portfolio policies are discussed in Section 5. Section 6 contains a few concluding remarks and suggestions for possible future extensions.

2. Formulation

In this section, a model of incomplete securities markets in a continuous time economy with security prices and state variables following a multi-dimensional diffusion process is formulated.

(Information Structure) Taken as primitive is a complete probability space \((\Omega, \mathcal{F}, P)\) and a time span \([0, T]\), where \(T\) is a strictly positive number. An element of \(\Omega\), denoted by \(\omega\), is a state of nature, which completely describes the exogenous uncertain environment from time 0 to time \(T\). The sigma-field \(\mathcal{F}\) is a collection of distinguishable events at time \(T\), and \(P\) is a probability measure representing an individual's beliefs about the likelihood of distinguishable events.

The uncertainty in the economy is generated by an \(N\)-dimensional standard Brownian motion, which is defined on \((\Omega, P, \mathcal{F})\) and denoted by \(w = \{w_n(t); t \in [0, T], n = 1, 2, ..., N\}\). Let \(\mathcal{F} = \{\mathcal{F}_t; t \in [0, T]\}\) be the natural filtration generated by \(w\). We assume that \(\mathcal{F}_T = \mathcal{F}\), that is, the true state of nature is determined completely by the sample path of \(w\) from time 0 to time \(T\).

(Consumption Space) There is a single consumption good (taken as numeraire) available at each

---

3 A probability space is said to be complete if \(P(A) = 0\) implies that \(A \in \mathcal{F}\).

4 Let \(\mathcal{F}_t\) be the smallest sigma-field containing all the \(P\)-measure zero set with respect to which \(\{w(s); 0 \leq s \leq t\}\) is measurable. The increasing family of sub-sigma fields of \(\mathcal{F}\), \(\mathcal{F} \equiv \{\mathcal{F}_t; t \in [0, T]\}\), is usually termed the filtration generated by \(w\).
date. The individual's consumption pattern can be described by a pair of non-negative consumption rate process and non-negative final wealth, \((c, W)\), where \(c = \{c(t); t \in [0, T]\}\), with \(c(t)\) denoting the consumption rate at time \(t\) and \(W\) denoting the final wealth. Naturally, we require an individual's consumption decision at any date to depend upon information available only at that date. This can be done by using the notion of adaptedness. A stochastic process \(X = \{X(t); t \in [0, T]\}\) is said to be adapted to \(\mathcal{F}\) if \(X(t)\) is measurable with respect to \(\mathcal{F}_t\) for all \(t \in [0, T]\). We require that the final wealth \(W\) be measurable with respect to \(\mathcal{F}_T\), and the consumption rate process \(c\) be measurable with respect to the optional sigma-field\(^5\). As a result, \(c\) is adapted to \(\mathcal{F}\) (see Chung and Williams [1983]). Henceforth, the consumption space will be denoted by \(\Sigma_+\).

(Securities Price System) We consider a frictionless securities market with \(M + 1\) long-lived securities available for trading between time \(0\) and time \(T\). Security \(0\) is locally riskless, has an instantaneous rate of return \(\tau(t)\) and pays no dividends. Security \(m\) (\(m = 1, 2, ..., M\)) is locally risky, at time \(t\) pays dividends at rate \(\iota_m(t)\), and sells for \(S_m(t)\). We use \(S(t)\) and \(u(t)\) to denote \((S_1(t), S_2(t), ..., S_M(t))^T\) and \((\iota_1(t), \iota_2(t), ..., \iota_M(t))^T\), respectively. The security price and state variable processes follow an \(N\)-dimensional diffusion process, whose movement can be described by the following system of stochastic differential equations:

\[
\begin{align*}
    dS(t) + \iota(S(t), Y(t), t)dt &= b(S(t), Y(t), t)dt + \sigma(S(t), Y(t), t)dw(t), \\
    dY(t) &= \mu(S(t), Y(t), t)dt + \rho(S(t), Y(t), t)dw(t), \\
    dB(t) &= B(t) \tau(S(t), Y(t), t)dt,
\end{align*}
\]

where \(Y(t)\) is an \((N - M)\)-dimensional vector process for the state variables. The assumption that the dimension of \(Y\), \(\text{dim}(Y)\), equals to \(N - M\) is without any loss of generality. If \(\text{dim}(Y)\) is less than \(N - M\), we can construct a new, \(M + \text{dim}(Y)\)-dimensional Brownian motion such that \(S\) and \(Y\) satisfy similar stochastic differential equations with respect to the new Brownian motion. If \(\text{dim}(Y)\) is greater than \(N - M\), then the local movements of \(S\) and \(Y\) are linearly dependent. We can handle this case by simply dropping some of the components of \(Y\). However, notationally it is cumbersome to do so, since it is likely that different components of \(Y\) may be dropped at different time. For ease of exposition, we adopt the above formulation for \(S\) and \(Y\). The functions

\[
\begin{align*}
    \iota(z, y, t), \quad b(z, y, t) &\quad : \mathbb{R}^M \times \mathbb{R}^{(N-M)} \times [0, T] \to \mathbb{R}^M, \\
    \sigma(z, y, t) &\quad : \mathbb{R}^M \times \mathbb{R}^{(N-M)} \times [0, T] \to \mathbb{R}^{M \times N}, \\
    \mu(z, y, t) &\quad : \mathbb{R}^M \times \mathbb{R}^{(N-M)} \times [0, T] \to \mathbb{R}^{(N-M)}, \\
    \rho(z, y, t) &\quad : \mathbb{R}^M \times \mathbb{R}^{(N-M)} \times [0, T] \to \mathbb{R}^{(N-M) \times N},
\end{align*}
\]

\(^5\)The smallest sigma-field of subsets of \(\Omega \times [0, T]\) with respect to which all processes adapted to \(\mathcal{F}\) having right-continuous sample paths are measurable mapping from \(\Omega \times [0, T]\) to \(\mathbb{R}\) is termed optional sigma-field.
\[ \tau(x, y, t) : \mathbb{R}^{M} \times \mathbb{R}^{(N-M)} \times [0, T] \rightarrow \mathbb{R}, \]

are assumed to be continuous with respect to all of their arguments. The \( N \times N \) matrix process formed by stacking the matrices \( \sigma \) and \( \rho \) is assumed to be non-singular for all \( x \in \mathbb{R}^{M}, y \in \mathbb{R}^{(N-M)}, \) and \( t \in [0, T]. \) This implies that the filtration generated by \( S \) and \( Y \) is identical to that generated by \( N \) Brownian motions \( w \) (see Harrison and Kreps [1979]). Thus the portfolio policies based on the information revealed from the prices and the state variables incorporate all of the uncertainty in the economy. Note that we allow the securities prices to take on negative values. This permits us to include financial assets in the set of traded assets. Without loss of generality, we also assume that \( B(0) = 1 \) and \( \tau(t) \) is non-negative, which implies \( B(t) \geq 1. \)

We emphasize the role of state variables and their connection with market incompleteness. Since there are a total of \( M + 1 \) securities traded and a total of \( N \) sources of uncertainty (generated by \( N \) independent Brownian motions), the risks associated with the \( N - M \) state variables \( Y \) can not be completely hedged by dynamic trading. Therefore, markets are dynamically incomplete in the sense that contingent claims written on traded assets may not be replicable through dynamic trading in existing assets. Our setup is similar to that of Cox, Ingersoll and Ross [1985], where they allow the production technologies to depend upon a set of state variables. On the other hand, our model is more general than the model of Pagès [1987], where he has eliminated the opportunity set \( Y(t) \) by restricting the parameters of the price processes to be functions of current prices. This specialization essentially brings the market structure back to the complete markets case, and therefore excludes the situations of real interest.

(Trading Strategies) A feasible dynamic trading strategy is an \( M + 1 \)-vector process, denoted generically by

\[ \{\alpha(t), \theta(t) \equiv (\theta_1(t), \ldots, \theta_M(t))^T ; t \in [0, T]\}, \]

where \( \alpha(t) \) and \( \theta_m(t) \) are the number of shares of the 0-th and the \( m \)-th security held at time \( t \), respectively, and they satisfy the following conditions:

\[ \int_0^T |\alpha(t)B(t)r(t) + \theta(t)^T b(t)| dt < \infty \quad P - a.s., \]

(1)

B2.

\[ \int_0^T |\theta(t)^T \sigma(t)|^2 dt < \infty \quad P - a.s., \]

(2)

B3. there exists a consumption-final wealth pair \( (c, W) \in \Sigma_+ \), such that, \( P - a.s., \)

\[ \alpha(t)B(t) + \theta(t)^T S(t) + \int_0^t c(s) ds = \alpha(0)B(0) + \theta(0)^T S(0) + \int_0^t (\alpha(s)B(s)r(s) + \theta(s)^T b(s)) ds + \int_0^t \theta(s)^T \sigma(s) dw(s) \quad \forall t \in [0, T], \]

(3)
B4. 

\[ \alpha(T)B(T) + \theta(T)^\top S(T) = W \quad P - a.s., \]  

(4)

B5. and 

\[ \alpha(t)B(t) + \theta(t)S(t) \geq 0 \quad \forall t \in [0, T] \quad P - a.s., \]  

(5)

where \( ^\top \) denotes matrix transpose and \( |A| \equiv \text{trace}(AA^\top)^{\frac{1}{2}} \) denotes the norm for matrix \( A \). Conditions B1 and B2 ensure that the stochastic integrals on the right-hand-side of (3) are well-defined (see Lipster and Shiryaev [1977], Ch4.). Consequently, the integral with respect to \( c \) on the left-hand-side of (3) is also well defined. Equation (2) and (4) are the natural budget constraint for intermediate consumption and final wealth, respectively, while (5) is the non-negative wealth constraint, whose implication will be discussed in the next section.

A consumption-final wealth pair \((c, W)\) is said to be marketed with initial wealth \( W_0 \), and financed by the trading strategy \((\alpha, \theta)\), if \((c, W)\) and \((\alpha, \theta)\) satisfy B1 \(\sim\) B5. We will use \( C(W_0) \) to denote the set of marketed consumption bundles with initial wealth \( W_0 \).

(Preference and a Dynamic Problem) An individual’s preference is represented by maximizing the expected utility with a time-additive utility function, \( u(x, t) \), for consumption rate at time \( t \) and a utility function, \( V(x) \), for final wealth. We assume that both \( u \) and \( V \) are continuous, strictly increasing, strictly concave in \( x \), while \( u \) is decreasing in \( t \).

The individual’s dynamic consumption-portfolio problem is to maximize the expected utility over the set of marketed consumption-final wealth pairs, \( C(W_0) \). That is, the individual solves the following program:

\[
\sup_{(c, W) \in C(W_0)} E \left[ \int_0^T u(c(t), t) dt + V(W) \right].
\]

(P)

Our objective is to provide a set of sufficient conditions to ensure that (P) has a solution and to obtain a complete characterization for the optimal solution.

3. From a Dynamic Problem to a Static Variational Problem

In this section, we will first characterize the set of implicit Arrow-Debreu state prices consistent with the absence of arbitrage opportunity. We will then formulate a dual Arrow-Debreu state price problem, which establishes an equivalence between the dynamic consumption-portfolio problem (P) and a static variational problem that maximizes the expected utility over consumption bundles satisfying a single budget constraint.

3.1. Equivalent Martingale Measures and Arrow-Debreu State Prices
We begin with a discussion on no arbitrage opportunity for our price system. For a consumption-portfolio problem to be well-posed, we certainly do not want our price system to admit any arbitrage opportunities. That is, we want to eliminate the possibility of something (with positive final payoffs) to be created from nothing. Harrison and Kreps [1979], Huang [1985] and Kreps [1981] have shown that a sufficient condition for the absence of arbitrage opportunities for simple strategies is that $S$ plus the cumulated dividends are related to martingales after a normalization by the bond price and a change of probability measure, or formally, there exists an equivalent martingale measure. A probability measure $Q$ on $(\Omega, \mathcal{F})$ is called an equivalent martingale measure if

i) $Q$ is equivalent to $P$,\(^6\)

ii) $G^*$ is a martingale under $Q$, where $G^*(t) \equiv S(t)/B(t) + \int_0^t \upsilon(s)/B(s)ds$. That is, for all $0 \leq t \leq T$, $E_Q[G^*(s)|\mathcal{F}_t] = G^*(t)$, where $E_Q$ denotes the expectation under $Q$.

One can interpret $G^*(t)$ as the discounted gains process since it is the sum of the discounted price process (capital gains) and the cumulative discounted dividends process. Applying Itô's lemma, we obtain

$$G^*(t) = S(0)/B(0) + \int_0^t [b(s) - \tau(s)S(s)]/B(s)ds + \int_0^t \sigma(s)/B(s)dw(s). \quad (6)$$

In general, $G^*$ is not a martingale under $P$, since it has a non-zero drift. But, after a change of probability measure, the drift of $G^*$ becomes zero, and $G^*$ becomes a martingale under the new measure. The following regularity conditions will ensure the existence of an equivalent martingale measure. Let

$$\kappa(t) \equiv -\sigma(t)^\top(\sigma(t)\sigma(t)^\top)^{-1}(b(t) - \tau(t)S(t)),$$

$$\xi(t) \equiv \exp\left(\int_0^t \kappa(s)^\top dw(s) - 1/2 \int_0^t |\kappa(s)|^2 ds\right),$$

and define $Q(A) \equiv \int_A \xi(\omega, T)P(d\omega)$, we assume

A1.

$$E\left[\exp\left(\frac{1}{2} \int_0^T |\kappa(t)|^2 dt\right)\right] < \infty,$$

A2.

$$E\left[\int_0^T |\sigma(t)|^4 dt\right] < \infty.$$

We show below that $Q$ is an equivalent martingale measure.

**Lemma 1** Suppose $E \int_0^T |\xi(t)|^2 dt < \infty$, and assumptions A1 and A2 are satisfied. Then the probability measure $Q$ is an equivalent martingale measure. Furthermore, $w^*(t) \equiv w(t) - \int_0^t \kappa(s)ds$

\(^6\) A probability measure $Q$ is said to be equivalent to $P$ if they have the same measure zero sets. A necessary and sufficient condition for this is that the Radon-Nikodym derivative $dQ/dP$ is strictly positive, $P$- a.s.
defines a standard \( N \)-dimensional Brownian motion under \( Q \), and the discounted gains process can be written as
\[
G^*(t) = S(0)/B(0) + \int_0^t \sigma(s)/B(s) dw^*(s).
\]

PROOF. First, A1 implies that \( \int_0^T |\xi(t)|^2 dt < \infty \) a.s., hence \( \xi(T) > 0 \), \( P-a.s. \). Applying Girsanov's theorem (Liptser and Shiryaev [1977], Ch. 6.), we conclude that \( Q \) is a probability measure, is equivalent to \( P \), and \( w^*(t) \) defines an \( N \)-dimensional Brownian motion under \( Q \).

Next, substituting \( w^* \) into (6), we can rewrite the discounted gains process as
\[
G^*(t) = S(0)/B(0) + \int_0^t \sigma(s)/B(s) dw^*(s).
\]

What remains to be proved is that \( G^*(t) \) is a martingale under \( Q \). We can prove this by showing that the stochastic integral on the right-hand side satisfies an \( L^2(Q) \)-integrability condition:
\[
E_Q \int_0^T |\sigma(t)/B(t)|^2 dt = E \int_0^T \xi(t)|\sigma(t)/B(t)|^2 dt
\leq E \int_0^T \xi(t)|\sigma(t)|^2 dt 
\leq \left[ E \int_0^T |\xi(t)|^2 dt E \int_0^T |\sigma(t)|^4 dt \right]^{1/2} < \infty,
\]

where \( E_Q \) denotes the expectation taken under \( Q \). \( \blacksquare \)

The existence of an equivalent martingale measure ensures that there is no arbitrage opportunity for simple strategies. However, arbitrage opportunity can still exist for strategies defined in (3) \( \sim \) (4). There are two ways to impose restrictions on the trading strategies to preclude arbitrage opportunity for strategies other than simple strategies. One way is to impose an \( L^2(Q) \)-integrability condition on \( \theta \) as follows (see Cox and Huang [1987a]),
\[
E_Q \left[ \int_0^T \left| \frac{\theta(t)^T \sigma(t)}{B(t)} \right|^2 dt \right] < \infty.
\]

The second way is to impose a non-negative wealth constraint, such as (5) (see Dybvig and Huang [1988]). Consequently, the feasible trading strategies defined in Section 2 do not admit any arbitrage opportunities. It has been shown in Dybvig and Huang that the \( L^2(Q) \)-integrability condition is stronger than the non-negative wealth constraint.

An important feature of the equivalent martingale measure \( Q \) is that it allows us to compute the value of a consumption-final wealth pair \( (c, W) \) by taking the conditional expectation under \( Q \). Let us assume momentarily that \( (c, W) \in C(W_0) \) is financed by \( (\alpha, \theta) \), where \( \theta \) satisfies the \( L^2(Q) \)-integrability condition. Then, it is easy to show that the value of \( \{c(s); s \in [t, T]\} \) and \( W \)

13
at time $t$ can be computed as follows (see Proposition 3.2, Cox and Huang [1987a]):

$$
\alpha(t)B(t) + \theta(t)^T S(t) = B(t)E_Q \left[ \int_t^T c(s)/B(s)ds + W/B(T) | \mathcal{F}_t \right] \\
= \frac{B(t)}{\xi(t)} E \left[ \int_t^T c(s)\xi(s)/B(s)ds + W\xi(T)/B(T) | \mathcal{F}_t \right], \quad \text{a.s.} \quad (7)
$$

From (7), we can interpret $\int_A \xi(\omega, t)/B(\omega, t)P(d\omega)$ to be the time 0 Arrow-Debreu price of a security that pays one unit of consumption in event $A \in \mathcal{F}_t$ at time $t$ and nothing otherwise. Or put it differently, $\xi(t)$ forms an implicit system of Arrow-Debreu state price that values a consumption-final wealth pair by (7).

It is known that when markets are dynamically complete there exists a unique equivalent martingale measure (see Harrison and Kreps [1979] and Huang [1985]). We show below that when markets are incomplete there exist infinitely many equivalent martingale measures. The following proposition (due to Pagés [1988]) presents a complete characterization of these measures.

**Proposition 1** i) If $Q$ is an equivalent martingale measure with respect to $P$, then the Radon-Nikodym derivative, $dQ/dP$, equals to $\xi_\nu(T)$, where

$$
\xi_\nu(t) = \exp \left( \int_0^t (\kappa(s) + \nu(s))^T dw(s) - \frac{1}{2} \int_0^T (|\kappa(t)|^2 + |\nu(t)|^2)dt \right), \quad (8)
$$

with $\nu \in \ker(\sigma) \equiv \{\nu(t) \in \mathbb{R}^N : \nu \text{ is adapted, } \sigma(t)^T\nu(t) = 0, \forall t \in [0, T]\}$.

ii) Conversely, if $\xi_\nu$ can be represented as in (8) for some $\nu \in \ker(\sigma)$ such that $E[\xi_\nu(T)] = 1$ and $E[\int_0^T |\xi_\nu(t)|^2dt] < \infty$, then the measure defined by

$$
Q_\nu(A) = \int_A \xi_\nu(\omega, T)P(d\omega)
$$

is an equivalent martingale measure. Furthermore,

$$
\omega_\nu(t) = w(t) - \int_0^t (\kappa(s) + \nu(s))ds = w^*(t) - \int_0^t \nu(s)ds
$$

defines a standard $N$-dimensional Brownian motion under $Q$, and $G^*$ can be written as

$$
G^*(t) = S(0)/B(0) + \int_0^t \sigma(s)/B(s)d\omega_\nu(s).
$$

**Proof.** Our first claim follows directly from Proposition 2.2 of Pagés [1987]. To be complete, we sketch the proof here. Let $\xi(t) = E[dQ/dP|\mathcal{F}_t]$. Since $\xi(t)$ is an $L^1(P)$ martingale, we can apply $L^1(P)$-martingale representation theory (see Clark [1970]) to express $\xi(t)$ as an Itô integral:

$$
\xi(t) = 1 + \int_0^t \theta(t)^T dw(t), \quad (9)
$$
for some $\theta$ adapted to $\mathcal{F}_t$ and $\int_0^t |\theta(t)|^2 dt < \infty$, a.s. Now use Bayes rule to argue that $G^*(t)$ is a martingale under $Q$ if and only if $\xi(t)G^*(t)$ is a martingale under $P$ (cf. Dellacherie and Meyer [1982], Lemma VII.48). But,

$$d(\xi(t)G^*(t)) = \xi(t)dG^*(t) + G^*(t)d\xi(t) + d\xi(t)dG^*(t)$$

$$= (\xi(t)[b(t) - \tau(t)S(t)] + \sigma(t)\theta(t))/B(t)dt + (\xi(t)\sigma(t)/B(t) + G^*(t)\theta(t)^	op)dw(t).$$

A standard argument leads to:

$$\xi(t)[b(t) - \tau(t)S(t)] + \sigma(t)\theta(t) = 0.$$

The general solution of the above equation is

$$\theta(t) = \xi(t)\kappa(t) + \xi(t)\nu(t),$$

for some $\nu \in \ker(\sigma)$. Consequently, the solution for (9) has the form of (8).

ii) The claim for the change of measure is a direct application of Girsanov’s Theorem (see, e.g., Liptser and Shiryayev [1977], Ch. 6.), and the martingale property can be proved as we did in Lemma 1.

We emphasize the significance of this proposition. It characterizes the nature of market incompleteness. When markets are complete, $\sigma$ is a non-singular square matrix and the null space of $\sigma$ (i.e. $\ker(\sigma)$) is $\{0\}$. Therefore there exists a unique equivalent martingale measure or a unique system of implicit Arrow-Debreu state prices. When markets are incomplete, the space of equivalent martingale measures is determined completely by the null space of $\sigma$. As illustrated by Proposition 1, there exist infinitely many equivalent martingale measures.

Similar to $\xi(t)$, $\zeta(t)$ also can be interpreted as a system of implicit Arrow-Debreu state prices for admissible consumption-final wealth pairs. To see that, we apply the Ito’s lemma and rewrite the budget constraint (3) as follows:

$$W(t)/B(t) + \int_0^t c(s)/B(s)ds = W_0 + \int_0^t \theta(s)\sigma(s)/B(s)dw^*(s)$$

$$= W_0 + \int_0^t \theta(s)\sigma(s)/B(s)dw_\nu(s),$$

where $W(t) = \alpha(t)B(t) + \theta(t)S(t)$. Similar to (7), we have

$$\alpha(t)B(t) + \theta(t)S(t) = \frac{B(t)}{\zeta_\nu(t)} E \left[ \int_t^T c(s)\xi_\nu(s)/B(s)ds + W\xi_\nu(T)/B(T)|\mathcal{F}_t \right] \quad \text{a.s.,}$$

if $\theta$ satisfies an $L^2(Q_\nu)$-integrability condition. Following from the above equation, we can interpret $\int_A \xi_\nu(\omega, t)/B(\omega, t)P(d\omega)$ to be the time 0 Arrow-Debreu price of a security that pays one unit of consumption in event $A \in \mathcal{F}_t$ and nothing otherwise.
This interpretation of $\xi_\nu$ as Arrow-Debreu state prices can be generalized to all $\xi_\nu$ that has a form of (8). In fact, applying Ito's lemma, we obtain

$$
\xi_\nu(t)W(t)/B(t) + \int_0^t \xi_\nu(s)c(s)/B(s)ds = W_0 + \int_0^t \xi_\nu(s)\theta(s)^T\sigma(s)/B(s)dw(s)
$$

$$
+ \int_0^t \xi_\nu(s)W(s)/B(s)(\kappa(s) + \nu(s))^Td\omega(s).
$$

(10)

Since the left-hand-side of the above equation is non-negative, the right-hand-side must be a super-martingale (see Lipster and Shiryayev [1977]). Thus,

$$
E \left[ \int_0^T \xi_\nu(t)c(t)/B(t)dt + \xi_\nu(T)W/B(T) \right] \leq W_0.
$$

(11)

Equality holds if $\theta$ satisfies certain integrability condition. In words, the value of an admissible consumption-final wealth pair must be less than or equal to the initial wealth $W_0$ under all $\xi_\nu$.

Although $\xi_\nu$ may not define an equivalent martingale measure, we claim that $\xi_\nu$ can always be used to define an equivalent local martingale measure in the sense that it defines a probability measure such that $G^*$ becomes a local martingale. Let

$$
\tau_n = \inf\{t \leq T : \int_0^t |\kappa(s) + \nu(s)|^2ds + \int_0^t |\sigma(s)/B(s)|^2 \geq n\},
$$

and define

$$
Q^n_\nu(A) = \int_A \xi_\nu(\omega, \tau_n)P(d\omega),
$$

then $Q^n_\nu$ is a probability measure, and $G^*(t)$ becomes a local martingale under $Q^n_\nu$.

In fact, following Girsanov’s theorem, we know that $w^n_\nu(t) = w^*(t \wedge \tau_n) - \int_0^{t\wedge \tau_n} \nu(s)ds$ is a standard Brownian motion under $Q^n_\nu$. Therefore,

$$
G^n(t) = G(t \wedge \tau_n) = S(0)/B(0) + \int_0^{t\wedge \tau_n} \sigma(s)/B(s)dw^n_\nu(s)
$$

is a martingale under $Q^n_\nu$.

We now see from (11) that an admissible consumption-final wealth pair under incomplete markets must satisfy infinitely many budget constraints formed by all $\xi_\nu$, where $\xi_\nu$ defines an equivalent local martingale measure. Consequently, standard optimization techniques such as Lagrangian theory can not be applied for solving (P). In the sequel, we will use $\Pi$ to denote the set of such $\xi_\nu$. To abuse the terminology a little bit, we will call $\xi_\nu$ the implicit Arrow-Debreu state price. In

---

7A stochastic process $X$ is said to be a super-martingale under $P$ if for all $0 \leq s < t \leq T$, $E[X(t)|F_s] \leq X(s)$, $P$-a.s.

8A stochastic process $X$ is said to be a local martingale under $P$ if there exists a sequence of stopping times $\tau_n$, such that $\tau_n \to T$ almost surely as $n \to \infty$ and $X^n(t) = X(t \wedge \tau_n)$ is a martingale under $P$. We are grateful to Steven Shreve for pointing out that $\xi_\nu$ may define an equivalent local martingale measure rather than an equivalent martingale measure.
the next section, we will make use of the set of implicit Arrow-Debreu state prices and establish a duality between the dynamic consumption-portfolio problem and a dual Arrow-Debreu state price problem.

3.2. A Duality between the Dynamic Problem and a Dual Problem

In this section, we will establish a duality theorem between the prime dynamic consumption-portfolio problem \((P)\) and a dual Arrow-Debreu state price problem. The objective of this duality theorem is to transform the original dynamic problem into a static variational problem, where the individual maximizes the expected utility over consumption bundles satisfying a single budget constraint. Standard optimization techniques will be used later to tackle the static variational problem subject to a single budget constraint.

We begin with a definition of minimax local martingale measure. First, let us define a static variational problem \((P_\nu)\) as follows:

\[
\sup_{(c, W) \in \Sigma_+} E \left[ \int_0^T u(c(t), t) dt + V(W) \right],
\]

s.t. \( E \left[ \int_0^T \xi_\nu(t)c(t)/B(t)dt + \xi_\nu(T)W/B(T) \right] \leq W_0. \quad (P_\nu)
\]

**Definition 1** \( \xi_\nu \in \Pi \) is said to define a minimax local martingale measure if the solution to \((P_\nu)\) coincides with the solution to \((P)\).

When the minimax local martingale measure exists, the original dynamic consumption-portfolio problem becomes equivalent to a static variational problem subject to a single budget constraint. We will demonstrate in this section as well as in later sections that the existence of minimax local martingale measure gives rise to a fundamental characterization of the dynamic consumption-portfolio problem with incomplete markets. To establish an existence as well as a characterization of the minimax local martingale measure, we introduce a dual Arrow-Debreu state price problem.

Let \(\text{val}(P_\nu)\) denote the supremum expected utility level of \((P_\nu)\). The dual problem for \((P)\) is defined as follows:

\[
\inf_{\{\nu: \xi_\nu \in \Pi\}} \text{val}(P_\nu). \quad (D)
\]

In words, the objective of the dual problem is to search for a system of implicit Arrow-Debreu state prices \(\{\xi_\nu\}\) such that it minimizes the maximum expected utility level of \((P_\nu)\). The following lemma provides essentially a complete characterization of the minimax local martingale measure.

**Lemma 2** If \(\xi_{\nu^*}\) defines a minimax local martingale measure, then \(\nu^*\) is a solution to \((D)\).
PROOF. If it is not true, then there exists \( \nu' \), where \( \xi_{\nu'} \in \Pi \), such that \( \text{val}(P_{\nu'}) < \text{val}(P_{\nu}) \). Thus

\[
\text{val}(P) \leq \text{val}(P_{\nu'}) < \text{val}(P_{\nu}) = \text{val}(P).
\]

The first inequality follows from the observation that

\[
C(W_0) \subset \left\{ (c, W) \in \Sigma_+: E \left[ \int_0^T \xi_{\nu'}(t)c(t)/B(t)dt + \xi_{\nu'}(T)W/B(T) \right] \leq W_0 \right\}.
\]

The last equality follows from the definition of minimax local martingale measure. We get a clear contradiction. \( \blacksquare \)

We intend to show below by a sequence of lemmas that the converse of Lemma 2 is also true. We need some notations to begin with. Define

\[
\hat{f}(y, t) = \inf\{ x \geq 0 | u_{x+}(x, t) \leq y \},
\]

\[
\hat{g}(y) = \inf\{ x \geq 0 | V_{x+}(x) \leq y \},
\]

where \( u_{x+}(x, t) \) and \( V_{x+}(x) \) denote the right-hand derivatives, which always exist since \( u \) and \( V \) are concave. Also, define the convex conjugate of \( u \) and \( V \) to be:

\[
\tilde{u}(y, t) = \sup_{x \geq 0} (u(x, t) - yx),
\]

\[
\tilde{V}(y) = \sup_{x \geq 0} (V(x) - yx).
\]

By definition, for any \( x > 0 \) and \( y > 0 \), we have \( \tilde{u}(y, t) = u(\hat{f}(y, t), t) - y\hat{f}(y, t) \) and \( \tilde{V}(y) = V(\hat{g}(y)) - y\hat{g}(y) \). Furthermore, \( u(x, t) \leq \tilde{u}(y, t) + xy \) and \( V(x) \leq \tilde{V}(y) + xy \). Rockafellar [1974] has shown that \( \tilde{u}(y, t) \) and \( \tilde{V}(y) \) are strictly convex and strictly decreasing if \( u \) and \( V \) are strictly concave and strictly increasing, and \( \tilde{u}'(y, t) = -\hat{f}(y, t) \) and \( \tilde{V}'(y) = -\hat{g}(y) \). Furthermore, \( -\tilde{u}(x, t) = -u(x, t) \) and \( -\tilde{V}(x) = -V(x) \).

Lemma 3 If \( \nu^* \) is a solution to \( (D) \), then for any \( \xi_{\nu} \in \Pi \) such that

\[
E \left[ \int_0^T \tilde{u}(\lambda_0 \xi_{\nu}(t)/B(t), t)dt + \tilde{V}(\lambda_0 \xi_{\nu}(T)/B(T)) \right] < \infty,
\]

we have

\[
E \left[ \int_0^T \xi_{\nu}(t)c^*(t)/B(t)dt + \xi_{\nu}(T)W^*/B(T) \right] \leq W_0,
\]

where \( c^*(t) = \hat{f}(\lambda_0 \xi_{\nu^*}(t)/B(t), t) \), \( W^* = \hat{g}(\lambda_0 \xi_{\nu^*}(T)/B(T)) \) and \( \lambda_0 \) is such that the equality holds for the budget constraint in \( (P_{\nu^*}) \).
PROOF. Assume for simplicity that \( V = 0 \) and \( W^* = 0 \). For \( \epsilon \in [0, 1] \), define \( \xi_\epsilon(t) = \epsilon \xi_\nu(t) + (1 - \epsilon) \xi_\nu^*(t) \). Clearly, \( \xi_\epsilon \in II \), which implies that
\[
E \left[ \int_0^T \frac{\bar{u}(\lambda_0 \xi_\epsilon(t)/B(t), t) - \bar{u}(\lambda_0 \xi_\nu^*(t)/B(t), t)}{\epsilon} \, dt \right] \geq 0.
\]
Since \( \bar{u} \) is convex, we have
\[
\frac{\bar{u}(\lambda_0 \xi_\epsilon(t)/B(t), t) - \bar{u}(\lambda_0 \xi_\nu^*(t)/B(t), t)}{\epsilon} \leq \bar{u}(\lambda_0 \xi_\nu(t)/B(t), t) - \bar{u}(\lambda_0 \xi_\nu^*(t)/B(t), t).
\] (12)
The right-hand-side of (12) is integrable. Fatou’s lemma implies that
\[
E \left[ \int_0^T \limsup_{\epsilon \downarrow 0} \frac{\bar{u}(\lambda_0 \xi_\epsilon(t)/B(t), t) - \bar{u}(\lambda_0 \xi_\nu^*(t)/B(t), t)}{\epsilon} \, dt \right] \geq 0.
\]
Hence,
\[
E \left[ \int_0^T \lambda_0 \bar{u}'(\lambda_0 \xi_\nu^*(t)/B(t), t)(\xi_\nu(t) - \xi_\nu^*(t))/B(t) \, dt \right] \geq 0.
\]
Since \( \bar{u}' = -\tilde{f} \), we obtain
\[
E \left[ \int_0^T \xi_\nu(t)\xi_\nu^*(t)/B(t) \, dt \right] \leq E \left[ \int_0^T \xi_\nu^*(t)\xi_\nu^*(t)/B(t) \, dt \right] = W_0.
\]

We need to make a technical assumption for the rest of this essay. For given \( \xi_\nu \in II \), define
\[
H_\nu(\lambda) = E \left[ \int_0^T \bar{u}(\lambda \xi_\nu(t)/B(t), t) \, dt + \tilde{V}(\lambda \xi_\nu(T)/B(T)) \right].
\]

Assumption T For any \( \tilde{\lambda} > 0 \), if \( H_\nu(\tilde{\lambda}) < \infty \), then there exists a constant \( \delta_{\tilde{\lambda}} \in (0, \tilde{\lambda}) \), such that
\( H_\nu(\tilde{\lambda} - \delta_{\tilde{\lambda}}) < \infty \).

The purpose of this assumption is to allow \( H_\nu(\lambda) \) to be differentiable with respect to \( \lambda \) for \( \lambda \in (\tilde{\lambda} - \delta_{\tilde{\lambda}}, +\infty) \). In fact, since \( \bar{u} \) and \( \tilde{V} \) are concave and decreasing, we have
\[
\left| \frac{\bar{u}(\lambda \xi_\nu(t)/B(t), t) - \bar{u}((\lambda - \epsilon) \xi_\nu(t)/B(t), t)}{\epsilon} \right| \leq \left| \frac{\bar{u}(\lambda \xi_\nu(t)/B(t), t) - \bar{u}(\tilde{\lambda} \xi_\nu(t)/B(t), t)}{\lambda - \tilde{\lambda} + \delta_{\tilde{\lambda}}} \right|,
\]
for \( \epsilon < \lambda - \tilde{\lambda} + \delta_{\tilde{\lambda}} \), and \( V \) satisfies a similar inequality. The right-hand-side of the above equation is integrable, therefore, the derivative of \( H \) can pass through the integral sign, which gives
\[
H_\nu'(\lambda) = E \left[ \int_0^T \tilde{u}'(\lambda \xi_\nu(t)/B(t), t) \xi_\nu(t)/B(t) \, dt + \tilde{V}'(\lambda \xi_\nu(T)/B(T)) \xi_\nu(T)/B(T) \right] < \infty.
\]
A sufficient condition for Assumption T to be satisfied is that \( u \) and \( V \) are asymptotically power functions. That is, there exist constants \( b, b', B_1, B'_1, B_2, B'_2 > 0 \) and \( A_1, A'_1, A_2, A'_2 \) such that
\[
A_1 + B_1 \frac{x^{1-b}}{1-b} \leq u(x,t) \leq A_2 + B_2 \frac{x^{1-b}}{1-b},
\]
\[
A'_1 + B'_1 \frac{x^{1-b'}}{1-b'} \leq V(z) \leq A'_2 + B'_2 \frac{x^{1-b'}}{1-b'}.
\]

The following lemma provides an alternative formulation of the dual problem.

**Lemma 4** Let \((\lambda_0, \nu^*)\) be defined as in Lemma 3, then \( \nu^* \) is a solution to \((D)\) if and only if \((\lambda_0, \nu^*)\) is a solution to
\[
\inf_{\lambda > 0} \left\{ \inf_{\xi \in \Pi} E \left[ \int_0^T \bar{u}(\lambda \xi(t)/B(t), t)dt + \bar{V}(\lambda \xi(T)/B(T)) \right] + \lambda W_0 \right\}. \quad (D')
\]

**Proof.** Necessity: For any \( \lambda > 0, \xi \in \Pi \), we have
\[
u(c^*(t), t) \leq \bar{u}(\lambda \xi(t)/B(t), t) + \lambda \xi(t)c^*(t)/B(t),
\]
\[
V(W^*) \leq \bar{V}(\lambda \xi(t)/B(T)) + \lambda \xi(T)W^*/B(T).
\]

Following Lemma 3, this implies that
\[
\text{val}(P_{\nu^*}) \leq E \left[ \int_0^T \bar{u}(\lambda \xi(t)/B(t), t)dt + \bar{V}(\lambda \xi(T)/B(T)) \right] + \lambda W_0.
\]

However, the right-hand-side of the above equation equals to \( \text{val}(P_{\nu^*}) \) when \( \lambda = \lambda_0 \) and \( \nu = \nu^* \). Hence, \((\lambda_0, \nu^*)\) is a solution to \((D')\).

Sufficiency: Let \((\lambda_0, \nu^*)\) be a solution to \((D')\). Consider the following function
\[
h(\lambda) = E \left[ \int_0^T \bar{u}(\lambda \xi^*(t)/B(t), t)dt + \bar{V}(\lambda \xi^*(T)/B(T)) \right] + \lambda W_0.
\]
According to Assumption T, \( h \) is differentiable with respect to \( \lambda \) for \( \lambda \in (\lambda_0 - \delta_{\lambda_0}, +\infty) \) and \( h(\lambda) \) attains its minimum at \( \lambda = \lambda_0 \). Thus, \( h'(\lambda_0) = 0 \), i.e.
\[
E \left[ \int_0^T \xi^*(t)\bar{u}'(\lambda_0 \xi^*(t)/B(t), t)dt + \xi^*(t)\bar{V}'(\lambda_0 \xi^*(T)/B(T))/B(T) \right] + W_0 = 0.
\]

Since \( \bar{u}'(x,t) = -\bar{f}(x,t) \) and \( \bar{V}'(x) = -\bar{g}(x) \), we infer that \((c^*, W^*)\) is a solution to \((P_{\nu^*})\). Therefore,
\[
h(\lambda_0) = \text{val}(P_{\nu^*}) = \inf_{\nu, \xi \in \Pi} \text{val}(P_{\nu}), \text{ i.e. } \nu^* \text{ is a solution to } (D).
\]

Although the dual problem \((D)\) is originated from the idea of transforming the dynamic problem to a static variational problem subject to a single budget constraint, the alternative formulation of \((D')\) is analytically more convenient to work with. We will make use of this fact later. The following theorem establishes a duality between the prime problem \((P)\) and the dual problem \((D)\).
Theorem 1 If \( \nu^* \) is a solution to \((D)\), then \((P)\) has a solution and \( \xi_{\nu^*} \) defines a minimax local martingale measure. More specifically, \((c^*, W^*)\) is a solution to \((P)\), where \((c^*, W^*)\) is defined as in Lemma 3.

Proof. See Appendix A. \( \square \)

We have obtained a duality between the prime dynamic consumption-portfolio problem and the dual Arrow-Debreu state price problem. Since the existence of the dual problem guarantees the existence of the prime problem, we can alternatively impose conditions to ensure that the dual problem have a solution. When the dual problem has a solution, the dynamic problem is equivalent to a static variational problem subject to a single budget constraint formed by the minimax local martingale measure. This characterization plays an important role in the later sections when we characterize the optimal solution.

3.3. Extension: Short-sale Constraints

We now extend our results to the case when there are short-sale restrictions on trading strategies. Mathematically, if an individual is forbidden from selling short the \( m \)-th security, \( \theta_m \) must be non-negative. For ease of exposition, we assume in the following discussion that the individual is only restricted from selling short the first security.

For purpose of establishing a duality between the dynamic consumption-portfolio problem subject to short-sale constraints and a static variational problem, we also need to find all the Arrow-Debreu state prices consistent with the absence of arbitrage opportunities. It turns out to that such Arrow-Debreu state prices are related to equivalent local super-martingale measures, i.e. there exists a sequence of stopping times \( \tau_n \to T \) as \( n \to \infty \), such that \( G^*(t \wedge \tau_n) \) becomes a super-martingale.\(^9\) We will briefly describe the results.

First, we claim that the Arrow-Debreu state prices consistent with the absence of arbitrage will have the following form:

\[
\xi_{\nu}(t) = \exp \left( \int_0^t (\kappa(s) + \nu(s))^T dw(s) - \frac{1}{2} \int_0^T |\kappa(s) + \nu(s)|^2 ds \right),
\]

where \( \nu \) is adapted, \( \sigma_1(t)\nu(t) \leq 0, \sigma_m(t)\nu(t) = 0, m = 2, \cdots, M \), and \( \sigma_l(t) \) is the \( l \)-th row of \( \sigma(t) \) (for \( l = 1, \cdots, M \)). More specifically, if a positive consumption-final wealth pair satisfies the budget constraint formed by such a \( \xi_{\nu} \), then this consumption plan can not be created from nothing. A little algebra shows that

\[
\xi_{\nu}(t) = \xi(t) \exp \left( \int_0^T \nu(t)^T dw^*(t) - \frac{1}{2} \int_0^T |\nu(t)|^2 dt \right).
\]

\(^9\)Similar results have been established for the finite dimensional case in He and Pearson [1988].
One can easily verify that under \( Q_\nu \), the price process for the first security becomes a local supermartingale, while all of the remaining price processes are local martingales. In fact, applying Girsanov’s theorem, we get

\[
G^n(t) = S(0)/B(0) + \int_0^{t\tau_n} \sigma(s)/B(s)dw^n_\nu(s) + \int_0^{t\tau_n} \sigma_1(s)^T\nu(s)/B(s)e_1ds,
\]

where \( e_1 = (1,0,\cdots,0)^T \) and \( \tau_n \) is the same as defined before. The first integral on the right-hand-side of the above equation satisfies an \( L^2(Q^n_\nu) \)-integrability condition, and it is therefore a martingale under \( Q^n_\nu \). The first row of the second integral is obviously a super-martingale under \( Q^n_\nu \) since \( \sigma_1(t)\nu(t) \leq 0 \), while the rest of it is zero. Hence, \( G^* \) is a local super-martingale. We will use \( \Pi_1 \) to denote the set of such \( \xi_\nu \). Obviously, \( \Pi \subset \Pi_1 \). Now, define the dynamic consumption-portfolio problem subject to short-sale constraints as follows:

\[
\sup_{(c,W)\in C(W_0),\nu(\cdot)\geq 0} \quad E \left[ \int_0^T u(c(t),t)dt + V(W) \right], \quad (P_1)
\]

and define the dual Arrow-Debreu state price problem for \((P_1)\) as follows:

\[
\inf_{\{\nu: \xi_\nu \in \Pi_1\}} \text{val}(P_\nu). \quad (D_1)
\]

We have the following duality theorem, whose proof is omitted.

**Theorem 2** If \( \nu^* \) is a solution to \((D_1)\), then the solution to \((P_1)\) exists and coincides with the solution to \((P_{\nu^*})\).

4. **Existence Theorem**

In this section, we focus on establishing sufficient conditions for \((P)\) and \((D)\) to have a solution. Our sufficient conditions involve with conditions on the measure of relative risk aversion for the utility functions, which are defined as

\[
R_u(x,t) = -xu''(x,t)/u'(x,t), \quad R_V(x) = -xV''(x)/V'(x).
\]

Since \( d^2V(e^x)/dz^2 = V''(e^x)e^{2x} + V'(e^x)e^x \), \( V' > 0 \) and \( V'' < 0 \), \( u(e^x,t) \) and \( V(e^x) \) are concave (or convex) in \( x \) if and only if \( R_u(x,t) \) and \( R_V(x) \geq 1(\forall r \leq 1) \), respectively. Moreover, since \( \tilde{V}'(y) = -\tilde{g}(y) < 0 \), and \( \tilde{V}''(y) = -1/V''(\tilde{g}(y)) > 0 \), we obtain

\[
R_{\tilde{u}}(y,t) = -y\tilde{f}(y,t)u''(\tilde{f}(y,t),t)/R_u(\tilde{f}(y,t),t),
\]

\[
R_{\tilde{V}}(y) = -\tilde{g}(y)V''(\tilde{g}(y))/R_V(\tilde{g}(y)).
\]
Hence, \( \tilde{u}(e^z, t) \) and \( \tilde{V}(e^z) \) is concave (or convex) in \( z \) if and only if \( R_u(z) \) and \( R_v(z) \geq 1 \) (or \( \leq 1 \)), respectively.

We assume for now that the individual's consumption occurs only at final date, i.e. \( u \equiv 0, \ c \equiv 0 \) (no intermediate consumptions). We need a lemma before we present the basic existence theorem.

Let
\[
C_1(W_0) = \{ W \in C(W_0) : \forall \omega, \exists U_\omega > 0 \text{ such that } W(\omega, t) \leq U_\omega \ \forall t \in [0, T] \},
\]
\[
C_2(W_0) = \{ W \in C(W_0) : \forall \omega, \exists L_\omega > 0 \text{ such that } W(\omega, t) \geq L_\omega \ \forall t \in [0, T] \},
\]
where \( W(\omega, t) \) is the wealth process for the final wealth \( W \). In words, \( C_1(W_0) \) contains those final wealth bundles whose wealth process has a uniformly bounded sample path. \( C_2(W_0) \) contains those final wealth bundles for which the sample path of the wealth process is uniformly bounded from below by some strictly positive number as well as uniformly bounded from above.

Lemma 5
\[
\sup_{W \in C(W_0)} EV(W) = \sup_{W \in C_1(W_0)} EV(W) = \sup_{W \in C_2(W_0)} EV(W).
\]

Proof. We prove this argument by showing that for any \( W \in C(W_0) \) and \( \epsilon > 0 \), there exist \( W' \in C_1(W_0) \) and \( W'' \in C_2(W_0) \) such that
\[
EV(W'') > EV(W') - \epsilon > EV(W) - 2\epsilon.
\]
This can be done as follows. By continuity, there exists a constant \( K > W_0 > 0 \), such that
\[
EV(W \wedge K) > EV(W) - \epsilon.
\]
Let us define a stopping time \( \tau \) to be \( \inf_t \{ t \leq T : W(t) \geq K \} \), where \( W(t) \) is the wealth process for \( W \). Let \( W' = W(T \wedge \tau) \), then \( W' \in C_1(W_0) \), \( W' \geq W \wedge K \), and
\[
EV(W') \geq EV(W \wedge K) > \epsilon.
\]
Now, let \( W'(t) \) be wealth process for \( W' \) and define stopping time \( \tau_n \) as follows,
\[
\tau_n = \inf_t \{ t : W'(t) \leq \frac{1}{n} \}, \quad \tau_\infty = \inf_t \{ t : W'(t) \leq 0 \}.
\]
Then \( \tau_n \to \tau_\infty \). Let \( W'_n = W'(\tau_n \wedge T) \), it is clear that \( W'_n \in C_2(W_0) \) and \( W_n \downarrow W' \). By monotone convergence theorem, we have
\[
EV(W'_n) \to EV(W').
\]
Therefore, there exists a \( n_0 \) such that if \( W'' = W'_{n_0} \), \( EV(W'') > EV(W') - \epsilon \).

We now present one of the main existence theorems for (P), while assuming that there is no intermediate consumptions. We define \( \text{span}(\sigma) = \{ \psi(t) : \exists \theta, \text{adapted, s.t.} \ \psi(t) = \sigma(t)^T \theta(t) \} \).

23
Theorem 3 Suppose \( u \equiv 0 \). If the measure of relative risk aversion \( R_v(x) \) is greater than 1, and \( V(x) \) is bounded from above, then there exists a solution to \( (P) \).

PROOF. See Appendix A. \( \square \)

Remark 1 1) Note that this theorem only applies to portfolio problems with no intermediate consumptions. We will present a more general existence theorem for \( (P) \) using a dual approach, i.e. through proving the existence to the dual problem \( (D) \).

2) If \( V(x) = A + B x^{1-b} \) for \( B > 0 \) and \( b > 1 \), then \( V \) satisfies the sufficient conditions of this theorem.

We now investigate the existence of the dual problem \( (D) \). We need the following lemma to begin with.

Lemma 6 Suppose that \( u \) and \( V \) are bounded from below, but unbounded from above as \( x \to \infty \). If for any \( \lambda > 0 \), there exists a solution to

\[
\inf_{\xi \in \Pi} \mathbb{E} \left[ \int_0^T \tilde{u}(\lambda \xi(t)/B(t), t)dt + \tilde{V}(\lambda \xi(t)/B(T)) \right],
\]

then there exists a solution to \( (D) \).

PROOF. Define

\[
f(\lambda) = \inf_{\xi \in \Pi} \mathbb{E} \left[ \int_0^T \tilde{u}(\lambda \xi(t)/B(t), t)dt + \tilde{V}(\lambda \xi(t)/B(T)) \right] + \lambda W_0.
\]

We claim that \( f(\lambda) \) is convex in \( \lambda \) for \( \lambda > 0 \) and goes to infinity as \( \lambda \to 0 \) or \( \infty \). Therefore \( f \) attains its infimum at some \( \lambda_0 > 0 \).

For simplicity, we assume that \( B(t) \equiv 1 \) and \( u \equiv 0 \). Let \( \xi_i \), be such that \( f(\lambda_i) = \mathbb{E}[\tilde{V}(\lambda_i \xi_i(t))] + \lambda_i W_0 \) (\( i = 1, 2 \)). Then, for any \( \mu \in [0, 1] \),

\[
f(\mu \lambda_1 + (1 - \mu) \lambda_2) \leq \mathbb{E}[\tilde{V}(\mu \lambda_1 \xi_1(T) + (1 - \mu) \lambda_2 \xi_2(T))] + (\mu \lambda_1 + (1 - \mu) \lambda_2) W_0
\leq \mu \{\mathbb{E}[\tilde{V}(\lambda_1 \xi_1(T))] + \lambda_1 W_0\} + (1 - \mu) \{\mathbb{E}[\tilde{V}(\lambda_2 \xi_2(T))] + \lambda_2 W_0\}
= \mu f(\lambda_1) + (1 - \mu) f(\lambda_2).
\]

The first inequality follows from the fact that \( \Pi \) is convex, while the second inequality follows from the fact that \( \tilde{V} \) is convex.

Since \( V \) is bounded from below, \( \tilde{V} \) is also bounded from below, which implies \( f(+\infty) = +\infty \). Furthermore, let \( \xi \) be such that \( f(\lambda) \) is attained at \( \xi \). By the convexity of \( \tilde{V} \) and \( \mathbb{E}[\xi(T)] \leq 1 \), we have

\[
f(\lambda) > \tilde{V}(\lambda \xi(T)) + \lambda W_0 > \tilde{V}(\lambda) + \lambda W_0 > V \left( \frac{1}{\lambda} \right) - 1 + \lambda W_0 \to \infty,
\]

24
as \( \lambda \downarrow 0 \). These properties together imply that \( \inf_{\lambda > 0} f(\lambda) \) is attained at some \( \lambda_0 > 0 \). According to Lemma 4, there is a solution to \((D)\). □

The following theorem provides sufficient conditions for the dual problem \((D)\) to have a solution.

**Theorem 4** Suppose that \( u \) and \( V \) are bounded from below. A sufficient condition for \((D)\) to have a solution is that the measures of relative risk aversion \( R_u(x) \) and \( R_V(x) \) are less than or equal to one.

**Proof.** see Appendix A. □

**Remark 2** This theorem allows us to claim existence for the dynamic portfolio problem with
\[
u(x,t) = e^{t \frac{1-b}{1-b'}}, \quad V(x) = \frac{e^{b'}-1}{b'},
\]
where \( b, b' < 1 \).

So far, we have provided sufficient conditions for \((P)\) and \((D)\) to have a solution. Although the direct existence theorem for the prime problem requires \( u \equiv 0 \), the existence theorem for the dual problem allows the existence of the prime problem with intermediate consumptions for preferences whose measure of relative risk aversion is less than one. Moreover, when the dual problem has a solution, the solution to the prime problem can be characterized through the minimax local martingale measure.

Naturally, we want to know whether the minimax local martingale property still hold if we only have the existence to the prime problem. Our next theorem shows this is indeed true.

**Theorem 5** Suppose that the conditions for Theorem 3 are satisfied, and suppose further that there exists constants \( B, B' > 0 \), \( b' > b > 1 \) and \( C \) such that
\[
C + B' \frac{e^{1-b'}}{1-b'} \leq V(x) \leq C + B \frac{e^{1-b}}{1-b'},
\]
then there exists a solution to \((D)\).

**Proof.** See Appendix A. □

Similar existence theorems (such as Theorem 3, 4 and 5) can be established for portfolio problems with short-sale constraints. We will not repeat these arguments. Instead, we present an example which demonstrates how the duality technique can be applied to log-utility preferences.

**Example 1** (Optimal solution for log-utility preferences with short-sale constraints)
Consider an individual with initial wealth \( W_0 = 1 \) whose preferences can be represented by \( u(x,t) = 0 \) and \( V(W) = \ln W \). The individual’s problem is to maximize the expected utility of final wealth at date \( T = 1 \) subject to short-sale constraints on all \( M \) securities.
We consider the following maximization problem:

$$\max_W E[\ln(W)]$$

s.t. \(E[\xi_\nu(1)W/B(1)] = 1,\) \hspace{1cm} (P_\nu)

where \(\xi_\nu(t) = \exp(\int_0^t (\kappa(s) + \nu(s))dw(s) - \frac{1}{2} \int_0^t |\kappa(s) + \nu(s)|^2 ds)\) and \(\sigma(s)\nu(s) \leq 0\). The first order condition for this problem is

\[ W = B(1)/\xi_\nu(1). \]

Substituting into the objective function, we obtain

\[ \text{val}(P_\nu) = E[\ln B(1)] + \frac{1}{2} E \left[ \int_0^1 |\kappa(t)|^2 dt \right] + E \left[ \int_0^1 \kappa(t)^T \nu(t) + \frac{1}{2} |\nu(t)|^2 dt \right]. \]

In order to minimize \(\text{val}(P_\nu)\), we consider the quadratic program

\[ \min_{\kappa(t)\nu(t) \leq 0} \kappa(t)^T \nu(t) + \frac{1}{2} \nu(t)^T \nu(t). \]

The solution to this problem always exists, so the infimum is attained. Standard methods can be used to find the optimal point \(\nu^*\). We use the Kuhn-Tucker theorem to illustrate that \(\xi_\nu\) defines a minimax local martingale measure. By this theorem, there exists an adapted process, \(\lambda(t) \geq 0\), such that

\[ \kappa(t) + \nu(t) + \sigma(t)^T \lambda(t) = 0 \]

and

\[ \lambda_i(t) = 0 \text{ if } \sigma_i(t)\nu(t) < 0, \quad \sigma_i(t)\nu(t) = 0 \text{ if } \lambda_i(t) > 0. \]

These imply \(\nu(t)^T \kappa(t) + \nu(t)^T \nu(t) = 0\). We can verify that the resulting consumption bundle is marketed. Since \(dw^*(t) = dw(t) - \kappa(t)dt\),

\[ \frac{W}{B(1)} = \exp(\int_0^1 -\kappa(t) + \nu(t))^T dw^*(t) - \frac{1}{2} \int_0^1 (|\kappa(t)|^2 - |\nu(t)|^2) dt) \]

\[ = \exp(\int_0^1 -\kappa(t) - \nu(t))^T dw^*(t) - \frac{1}{2} \int_0^1 |\kappa(t) + \nu(t)|^2 dt), \]

where in the second equality we used the fact that \(\nu(t)^T \kappa(t) + \nu(t)^T \nu(t) = 0\). If we define

\[ \psi(t) = \exp(\int_0^t -\kappa(s) - \nu(s)dw^*(s) - \frac{1}{2} \int_0^t |\kappa(s) + \nu(s)|^2 ds), \]

then \(\psi(t)\) is the optimal discounted wealth process, and

\[ \frac{W}{B(1)} = 1 + \int_0^1 -\psi(t)(\kappa(t) + \nu(t))^T dw^*(t) \]

\[ = 1 + \int_0^1 \psi(t)\lambda(t)^T \sigma(t)dw^*(t) \]

\[ = 1 + \int_0^1 \psi(t)B(t)\lambda(t)^T dG^*(t). \]
Clearly, $W$ is marketed and the corresponding trading strategy is $\theta = \psi(t)B(t)\lambda(t)$. Furthermore, the complementarity condition suggests that the constraint $\sigma(t)\nu(t) = 0$ will be binding whenever $\lambda(t) > 0$. On the other hand, whenever the constraint is not binding, the individual will not hold that risky asset. This is expected from the structure of the minimax local martingale measure (see Lemma 7 below).

5. Characterizations and Computations

Though we have proved the existence of an optimal solution to the dynamic consumption-portfolio problem $(P)$ under certain conditions on the shape of the utility functions, we do not yet know how to compute optimal solutions. What makes the problem so difficult is exactly the fact that the feasible consumption bundles must satisfy infinitely many budget constraints. Moreover, finding a solution to the dual state price problem $(D)$ is also not a trivial exercise, although this can be done when preferences can be represented by log-utility.

The objective of this section is to characterize the optimal solution. We present a method of computing optimal policies through solving a quasi-linear partial differential equation. Specifically, we relate the optimal wealth process, the optimal trading strategy, and the minimax local martingale measure to the solution of a quasi-linear partial differential equation. Similar to the dynamic programming approach, we provide a verification theorem, which guarantees that the solution obtained by solving the PDE be optimal. The advantage of our approach is that we must only solve a quasi-linear partial differential equation. Numerical solution of this equation is easier than solution of the strongly non-linear Bellman equation.

We first deal with the case where short-sales are permitted. We need to introduce a new state variable for the rest of this section. Let $\xi(t) \in \Pi$ and define $Z(t) = Z_0 B(t)/\xi(t)$. Then $Z(t)$ satisfies the following stochastic differential equation:

$$dZ(t) = Z(t)(r(t) + |\kappa(t) + \nu(t)|^2)dt - Z(t)(\kappa(t) + \nu(t))^T dw(t),$$  \hspace{1cm} (13)

with $Z(0) = Z_0$. Note that $(Z(t), S(t), Y(t))$ forms a diffusion process.

The following theorem provides a characterization of the optimal solution and demonstrates a one-to-one correspondence between the minimax local martingale measure and a quasi-linear partial differential equation.

**Theorem 6** Suppose that $\xi(t)$ defines a minimax local martingale measure with the Lagrangian multiplier $\lambda_0 > 0$ being defined as in Lemma 3. Let $Z(t)$ be defined as in (13) using $\xi(t)$ with $Z_0 = 1/\lambda_0$, and define the function:

$$F(Z(t), S(t), Y(t), t)$$
\[ Z(t)E \left[ \int_t^T Z(s)^{-1} \tilde{f}(Z(s)^{-1}, s)ds + Z(T)^{-1} \tilde{g}(Z(T)^{-1}) \mid Z(t), S(t), Y(t) \right]. \] (14)

If \( F(Z, S, Y, t) \) has a continuous first order derivative w.r.t. to \( t \) and continuous second order derivatives w.r.t. to \( Z \), \( S \), and \( Y \), then \( F \) satisfies the following partial differential equation

\[ \mathcal{L}F + F_t = F_Z \mathbf{Z} \mathbf{Z}^T (\kappa + \nu)^T (\kappa + \nu) - F_S \mathbf{S} \mathbf{S}^T (\kappa + \nu) - F_Y \mathbf{Y} \mathbf{Y}^T (\kappa + \nu) + \rho F - \tilde{f}(Z^{-1}, t), \] (15)

with boundary conditions

\[ F(Z, Y, S, T) = \tilde{g}(Z^{-1}), \] (16)
\[ F(Z(0), S(0), Y(0), 0) = W_0, \] (17)

where \( \mathcal{L} \) is the differential generator of \((Z, S, Y, t)\), and \( \nu \) and \( F \) have the relationship

\[ \nu^T = F_Y^T \rho(I_n - \sigma^T (\sigma \sigma^T)^{-1} \sigma) \chi_{\{F_Z > 0\}} / F_Z Z. \]

Furthermore, \( F \) defines exactly the optimal wealth process and the optimal policies can be determined completely from \( F \) as follows:

\[ \theta(t) = \left( F_Z(t) + (\sigma(t) \sigma(t)^T)^{-1}(b(t) - \rho(t)S(t))Z(t)F_Z(t) + (\sigma(t) \sigma(t)^T)^{-1} \sigma(t) \rho(t)^T F_Y(t) \right) / B(t), \]
\[ \alpha(t) = \left( F(Z(t), S(t), Y(t), t) - \theta(t)^T S(t) \right) / B(t). \] (18)

**Proof.** By assumption, \((P)\) is equivalent to \((P_\nu)\). Thus from the first order condition of \((P_\nu)\), we have that \( c(t) = \tilde{f}(Z(t)^{-1}, t) \) and \( W = \tilde{g}(Z(T)^{-1}) \) are the solution to \((P)\) and in particular \((c, W)\) is market-ed. Let \((\alpha, \theta)\) be the trading strategy associated with \((c, W)\), then

\[ W(t)/B(t) + \int_0^t c(s)/B(s)ds = W_0 + \int_0^t \theta(s)^T \sigma(s)/B(s)dw^*(s), \] (19)

\[ E \left[ \xi_\nu(T)W/B(T) + \int_0^T \xi_\nu(t) c(t)/B(t)dt \right] = W_0, \] (20)

where \( W(t) = \alpha(t)B(t) + \theta(t)^T S(t) \). Applying Ito's lemma, we obtain

\[ \xi_\nu(t)W(t)/B(t) + \int_0^t \xi_\nu(s)c(s)/B(s)ds = W_0 + \int_0^t \xi_\nu(s)\theta(s)^T \sigma(s)/B(s)dw(s) \]

\[ + \int_0^t \xi_\nu(s)W(s)/B(s)(\kappa(s) + \nu(s)) / dw(s). \] (21)

Equations (19), (20) and (21) imply that

\[ \xi_\nu(t)W(t)/B(t) = E \left[ \xi_\nu(T)W/B(T) + \int_t^T \xi_\nu(s)c(s)/B(s)ds \mid Z(t), S(t), Y(t) \right]. \]
Thus,
\[
W(t) = \frac{B(t)}{\xi(t)} E \left[ \int_t^T \xi(s)c(s)/B(s)ds + \xi(T)W/B(T) \mid Z(t), S(t), Y(t) \right] \\
= Z(t)E \left[ \int_t^T Z(s)^{-1}c(s)ds + Z(T)^{-1}W \mid Z(t), S(t), Y(t) \right] \\
= Z(t)E \left[ \int_t^T Z(s)^{-1} \hat{\xi}(Z(s)^{-1}, s)ds + Z(T)^{-1} \hat{\xi}(Z(T)^{-1}) \mid Z(t), S(t), Y(t) \right] \\
= F(Z(t), S(t), Y(t), t).
\]

This proves that \( F \) defines exactly the optimal wealth process. Since \( F \) has the desired differential property by assumption, Ito's lemma implies that
\[
\xi(t)F(Z(t), Y(t), S(t))/B(t) + \int_0^t \xi(s)\hat{\xi}(Z(s)^{-1}, s)/B(s)ds \\
= F(0)/B(0) + \int_0^t \xi(s)(F^\top_S \sigma - F_Z Z(\kappa + \nu)^\top + F^\top_Y \rho)/B(s) dw(s) \\
+ \int_0^t \xi(s)(\mathcal{L}F + F_S - F_Z Z|\kappa + \nu|^2 + F^\top_S \sigma(\kappa + \nu) + F^\top_Y \rho(\kappa + \nu) - \gamma F \\
+ \hat{\xi}(Z(s)^{-1}, s))/B(s)ds.
\]

Equations (20) and (21) imply that the right-hand side of (22) is a martingale under \( P \). Therefore the drift term must be zero, which gives (15). Now, applying Ito's lemma again and using the fact that (15) holds, we obtain
\[
F(Z(t), Y(t), S(t))/B(t) + \int_0^t \hat{\xi}(Z(s)^{-1}, s)/B(s)ds \\
= F(0)/B(0) + \int_0^t (F^\top_S \sigma - F_Z Z(\kappa + \nu)^\top + F^\top_Y \rho)/B(s) dw^\ast(s).
\]

Since \((c, W)\) is marketed, the coefficients of the diffusion term must lie in \( \text{span}(\sigma) \). This implies
\[
-F_Z Z \nu^\top + F^\top_Y \rho(I_n - \sigma^\top(\sigma \sigma^\top)^{-1} \sigma) = 0.
\]

That is, if we decompose \( \rho \) into two parts, the part that is orthogonal to \( \text{span}(\sigma) \) must be canceled out by \( \nu \). In general, \( F_Z > 0 \) (see Cox and Huang [1987b]), and becomes zero if and only if consumption after that point becomes zero and \( F \equiv 0 \) thereafter. Thus we can solve for \( \nu \) from (23) to get
\[
\nu = F^\top_Y \rho(I_n - \sigma^\top(\sigma \sigma^\top)^{-1} \sigma)\chi_{\{F_Z > 0\}}/F_Z Z.
\]

The coefficients on the diffusion term become exactly the right-hand side of (18) for \( \theta \). A standard argument shows that the optimal portfolio policies are as specified in (18). The boundary conditions are satisfied trivially, since \( F(t) = W(t) \).
Remark 3 Rather than trying to give enough conditions to ensure the smoothness of $F$, we start by assuming that $F$ is smooth. The standard sufficient conditions for smoothness involve growth conditions and Lipschitz continuity conditions on the parameters of the price and state variable processes, and growth conditions on $\hat{f}$ and $\hat{g}$. (See Cox and Huang [1987b] for details.)

The portfolio strategy obtained through (18) is consistent with the results of Merton [1973] and Cox and Huang [1987b]. The additional term, $(\sigma\sigma^T)^{-1}\sigma\rho^TF_Y/B(t)$, represents the hedging demand arising from the changing investment opportunity sets. In general, there are two sets of conditions that need to be satisfied in order to apply Theorem 6. The first one is the existence of a minimax local martingale measure, which is guaranteed by Theorem 3 or 4. The second one is related to the smoothness of $F$. This is very hard to verify in general.

For practical purpose, one would prefer to have a procedure to construct a consumption-portfolio policy and then use some criterion to verify that the constructed policy is indeed an optimal policy. This is the idea of the verification theorem which we come to next. We need some definitions.

Definition 2 A function $f(x, t): \mathbb{R}^N \times [0, T] \to \mathbb{R}$ is said to satisfy a polynomial growth condition if there exist constants $K, \beta > 0$ such that

$$|f(x, t)| \leq K(1 + |x|^{\beta}), \quad \forall x \in \mathbb{R}^N, \ t \in [0, T].$$

If $\beta = 1$, we say the function satisfies a linear growth condition.

Definition 3 A function $f(x, t): \mathbb{R}^N \times [0, T] \to \mathbb{R}$ is said to satisfy a local Lipschitz condition if for any compact set $E \subset \mathbb{R}^N$, there exists some constant $L_E$ such that

$$|f(x, t) - f(y, t)| \leq L_E|x - y|, \quad \forall x, y \in E.$$

If $L_E$ can be chosen to be independent of $E$, we say that $f$ satisfies a uniform Lipschitz condition.

The following theorem is a counterpart of the verification theorem in dynamic programming. The main idea is to manufacture a minimax local martingale measure in order to verify the optimality of a policy.

Theorem 7 (Verification Theorem) Suppose the following:

(a) $F: \mathbb{R}^{N+1} \times [0, T] \to \mathbb{R}$ satisfies (15) with

$$\nu^T = F_Y^T\rho(I_n - \sigma^T(\sigma\sigma^T)^{-1}\sigma)\chi_{\{F_Z > 0\}}/F_ZZ$$
and boundary conditions (16) and (17) for some constant $Z_0 > 0$;

b) There exist $K, \beta > 0$ such that, for all $(y, z, t) \in \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}_+$,

$$|e^{-y} F(e^y, z, t)| \leq K(1 + |y|^{\beta} + |z|^\beta), \quad |e^{-y} \hat{f}(e^{-y}, t)| \leq K(1 + |y|^{\beta}),$$

and $F(z, y, t) \to F(z, y, T)$ as $t \to T$;

c) For the $\nu$ defined in a), there exists a solution $Z(t)$ to (13);

d) $b(z, t), \sigma(z, t), \rho(z, t), r(z, t), |\kappa(z, t) + \nu(y, z, t)b|^2$ and $\kappa(z, t) + \nu(y, z, t)$ satisfy linear growth conditions and local Lipschitz condition for $(y, z) \in (0, +\infty) \times \mathbb{R}^N$.

Then there exists a solution to $(P)$ and $(P_\nu)$, $\nu \in \mathcal{P}$ is equivalent to $P_\nu$. Therefore, $\xi_{\nu}$ defines the minimax local martingale measure. The optimal policies can be computed from $F$ as in (18).

**Proof.** Firstly, $Z(t)$ is well defined by assumption c). Second, $F$ is continuously differentiable. So applying Ito's lemma and using the definition of $\nu$, we obtain

$$F(Z(t), Y(t), S(t))/B(t) + \int_0^t \hat{f}(Z(s)^{-1}, s)/B(s)ds$$

$$= F(0)/B(0) + \int_0^t (F_S^T \sigma - F_Z \kappa + \nu^T + F_Y^T \rho)/B(s)dw^*(t)$$

$$= W_0 + \int_0^t (F_S^T \sigma - F_Z \kappa^T + F_Y^T \rho \sigma^T (\sigma \sigma^T)^{-1}\sigma)/B(s)dw^*(t)$$

$$= W_0 + \int_0^t (F_S^T + F_Z (b - \tau S)^T (\sigma \sigma^T)^{-1} + F_Y^T \rho \sigma^T (\sigma \sigma^T)^{-1})dG^*(t).$$

Therefore, the consumption-final wealth pair $(\hat{f}(Z(t)^{-1}, t), \hat{g}(Z(t)^{-1})$ can be dynamically replicated. If we can show that

$$E \left[ \xi_{\nu}(T)W/B(T) + \int_0^T \xi_{\nu}(t)c(t)/B(t)dt \right] = W_0,$$

then we can conclude from the Lagrangian theory that $c(t) = \hat{f}(Z(t)^{-1}, t), W = \hat{g}(Z(T)^{-1})$ is an optimal solution to $P_\nu$ with $\lambda = 1/Z_0$. Hence, this bundle is indeed an optimal solution to $(P)$. However, the above equality is guaranteed by assumptions b) and d). A detailed proof is omitted. We refer the interested readers to Theorem 2.3 of Cox and Huang [1987b].

**Remark 4** We obtain a quasi-linear partial differential equation (15) rather than a linear PDE. The coefficients that involve $\nu$ are related to the first order partial derivatives of the unknown function $F$ through the relationship between $\nu$ and $F$ of (23). One can rewrite (15) by substituting the formula for $\nu$ into the equation to obtain explicitly the PDE. This quasi-linear partial differential equation is easier to solve numerically than the non-linear Bellman equation typically obtained from dynamic programming, see Ames [1977].
Parallel to Theorems 6 and 7 are the following theorems for the case with short-sale constraints. To be precise, we only impose short-sale constraints on the first risky security. The basic idea in this situation is almost identical to the previous case. We characterize the minimax local martingale measure through a quasi-linear partial differential equation, but, with a free boundary. Before we move on to the short-sale case, we present the following lemma, which gives a more precise statement of the relationship between the minimax local super-martingale measure and the optimal trading strategy when there are short-sale constraints.

Lemma 7 Suppose that \( \xi_\nu \) defines a minimax local super-martingale measure to \( (P) \) and \((c, W)\) is the corresponding optimal solution. Letting \((\alpha, \theta)\) be the trading strategy that implements \((c, W)\), then \( \theta_1(t)\sigma_1(t)\nu(t) = 0 \) a.s.  

**Proof.** Applying Ito's lemma, we have

\[
\xi_\nu(t)W(t)/B(t) + \int_0^t \xi_\nu(s)c(s)/B(s)ds = W_0 + \int_0^t \xi_\nu(s)\theta(s)\sigma(s)/B(s)dw(s) \\
+ \int_0^t \xi_\nu(s)W(s)/B(s)(\kappa(s) + \nu(s))^Tdw(s) \\
+ \int_0^t \xi_\nu(s)\theta_1(s)\sigma_1(s)\nu(s)/B(s)ds.
\]

Since the left-hand-side of the above equation is non-negative and the last integral on the right-hand-side is negative, the sum of the first two integrals on the right-hand-side must be a super-martingale and has an expectation less than or equal to zero. However, according to the budget constraint, the left-hand-side has expectation \(W_0\) under \(P\) for \(t = T\). We conclude that \(\theta_1(t)\sigma_1(t)\nu(t) = 0\) a.s.  

The implication of this lemma is quite interesting. If the individual's optimal portfolio position on the securities with short-sale constraints is not binding, then the minimax "direction" \(\nu\) will be orthogonal to the corresponding vectors in the stochastic term \((\sigma)\). On the other hand, if the minimax direction is not orthogonal to that vector, e.g. \(\nu\sigma_1 < 0\), then the individual will not hold that particular portfolio. In other words, \(\nu\) will deviate from the no short-sale condition \((\sigma\nu = 0)\) if and only if the short-sale constraint is binding. We have observed this fact in Example 1.

Now we are ready to present our main results for the short-sale case.

**Theorem 8** Suppose that \(\xi_\nu\) defines a minimax local super-martingale measure to the program \((P_1)\) with a Lagrangian multiplier \(\lambda_0 > 0\) as defined in Lemma 3. Let \(Z(t)\) be defined as in (13) using
$\xi$, with $Z_0 = 1/\lambda$, and define the function:

$$F(Z(t), S(t), Y(t), t)$$

$$= Z(t)E \left[ \int_t^T Z(s)^{-1} \dot{g}(Z(s)^{-1}) ds + \dot{g}(Z(T)^{-1})Z(T)^{-1} \mid Z(t), S(t), Y(t) \right].$$  \hspace{1cm} (24)$$

If $F(Z, S, Y, t)$ has a continuous first order derivative w.r.t. to $t$ and continuous second order derivatives w.r.t. to $Z$, $S$, and $Y$, then $F$ satisfies the following partial differential equation:

$$\mathcal{L}F + F_t = F_Z \kappa \nu^T (\kappa + \nu) - F_S \sigma (\kappa + \nu) - F_Y \rho (\kappa + \nu) + rF - \dot{f}(Z^{-1}, t)$$  \hspace{1cm} (25)$$

with boundary conditions

$$F(Z, Y, S, T) = \dot{g}(Z^{-1}),$$  \hspace{1cm} (26)$$

$$c(Z(u), S(0), Y(0), 0) = W_0,$$  \hspace{1cm} (27)$$

where $\mathcal{L}$ is the differential generator of $(Z, S, Y, t)$, and $\nu$ and $F$ satisfy the following relationship:

$$\nu^T = F_Y T \rho(I_n - \sigma^T (\sigma^T)^{-1} \sigma) \chi_{\{F_Z > 0\}} / ZF_Z,$$  \hspace{1cm} if $(Z, S, Y) \in \Omega_1,$

$$\nu^T = F_Y T \rho(I_n - \sigma^T (\sigma^T)^{-1} \sigma) \chi_{\{F_Z > 0\}} / ZF_Z + \gamma \eta^T \sigma$$  \hspace{1cm} if $(Z, S, Y) \in \Omega_2,$ \hspace{1cm} (28)$$

where

$$\gamma = (F_S + e_1^T (\sigma^T)^{-1} (b - rS)ZF_Z + e_1^T (\sigma^T)^{-1} \sigma \rho^T F_Y) \chi_{\{F_Z > 0\}} \sigma / F_Z Z,$$

$$\Omega_1 = \{(Z, S, Y) \mid F_S + e_1^T (\sigma^T)^{-1} (b - rS)ZF_Z + e_1^T (\sigma^T)^{-1} \sigma \rho^T F_Y \geq 0\},$$

$$\Omega_2 = \{(Z, S, Y) \mid F_S + e_1^T (\sigma^T)^{-1} (b - rS)ZF_Z + e_1^T (\sigma^T)^{-1} \sigma \rho^T F_Y < 0\},$$

$$\eta = (1, -\tau_2, \cdots, -\tau_M)^T$$

with

$$\begin{pmatrix}
\tau_2 \\
\vdots \\
\tau_M
\end{pmatrix} = \begin{pmatrix}
\sigma_2 \sigma_2^T \\
\vdots \\
\sigma_M \sigma_M^T
\end{pmatrix}^{-1} \begin{pmatrix}
\sigma_2 \sigma_1^T \\
\vdots \\
\sigma_M \sigma_1^T
\end{pmatrix},$$

and $e_1^T = (1, 0, \cdots, 0)$, an $N \times 1$ vector. Furthermore, $F$ defines exactly the optimal wealth process and the optimal policies can be determined completely from $F$ as follows:

$$\theta(t) = (F_S(t) + (\sigma(t) \sigma(t)^T)^{-1} (b(t) - \tau(t) S(t))Z(t) F_Z(t))$$

$$+ (\sigma(t) \sigma(t)^T)^{-1} \sigma(t) \rho(t)^T F_Y(t) - \gamma(t) \eta(t) Z(t) F_Z(t)) / B(t),$$

$$\alpha(t) = (F(Z(t), S(t), Y(t), t) - \theta(t)^T S(t)) / B(t).$$  \hspace{1cm} (29)$$

33
PROOF. We mimic the prove of Theorem 6. By assumption, \((P)\) is equivalent to \((P_\nu)\). Thus \(c(t) = \hat{f}(Z^{-1}, t)\) and \(W = \hat{\theta}(Z(T)^{-1})\) is the solution to \((P)\), and in particular, \((c, W)\) is marketable. Letting \((\alpha, \theta)\) be the trading strategy associated with \((c, W)\), then
\[
W(t)/B(T) + \int_0^t c(s)/B(s)ds = W_0 + \int_0^t \theta(s)^T \frac{\sigma(s)}{B(s)} dw^*(s),
\]
(30)
\[
E \left[ \xi_\nu(T) W/B(T) + \int_0^T \xi_\nu(t)c(t)/B(t)dt \right] = W_0.
\]
(31)
Similar \(\omega\) Theorem 6, equations (30) and (31) imply that
\[
W(t)/B(t) = E \left[ \xi_\nu(T) W/B(T) + \int_0^T \xi_\nu(s)c(s)/B(s)ds|Z(t), S(t), Y(t) \right]
\]
\[
= F(Z(t), S(t), Y(t), t)/B(t),
\]
where \(W(t) = \alpha(t)B(t) + \theta(t)^T S(t)\). That is, \(F\) defines exactly the wealth process. Since \(F\) has the desired differential property by assumption, Ito’s lemma implies that
\[
\xi_\nu(t) F(Z(t), Y(t), S(t))/B(t) + \int_0^t \xi_\nu(s) \hat{f}(Z(s)^{-1}, s)/B(s)ds
\]
\[
= F(0)/B(0) + \int_0^t \xi_\nu(s) (F_Z^T \sigma - F_Z Z(\kappa + \nu)^T + F_Y^T \rho)/B(s) dw(s)
\]
\[
+ \int_0^t \xi_\nu(s) (\mathcal{L} F + F_s - F_Z Z|\kappa + \nu|^2 + F_S^T \sigma(\kappa + \nu) + F_Y^T \rho(\kappa + \nu) - \tau F)
\]
\[
+ \hat{f}(Z^{-1}, t))/B(s)ds.
\]
(32)
As in Theorem 6, the right-hand-side is a martingale under \(P\). Therefore, the drift term on the right-hand side of (32) must be zero, which gives (25). Now, applying Ito’s lemma again and use the fact that (25) holds, we obtain
\[
F(Z(t), Y(t), S(t))/B(t) + \int_0^t \hat{f}(Z(s)^{-1}, s)/B(s)ds
\]
\[
= F(0)/B(0) + \int_0^t (F_Z^T \sigma - F_Z Z(\kappa + \nu)^T + F_Y^T \rho)/B(s) dw^*(s).
\]
Since \((c, W)\) is marketed, the diffusion term must be in \(\text{span}(\sigma)\), and the position in the first security, \(\theta_1\), must be non-negative. Next, decompose \(\nu\) into the sum of two orthogonal vectors \(\nu_1\) and \(\nu_2\), where \(\nu_1\) lies in \(\ker(\sigma)\), and \(\nu_2\) lies in \(\text{span}(\sigma)\) and \(\sigma_1 \nu_2 \leq 0\) but is orthogonal to \(\sigma_2, \ldots, \sigma_M\). One sees immediately that \(\nu_1^T = F_Y^T \rho(I_n - \sigma^T (\sigma \sigma^T)^{-1} \sigma) \chi_{\{F_Z > 0\}} / F_Z Z\). Also, simple linear algebra shows that \(\nu_2\) has the following representation:
\[
\nu_2^T = \gamma(1, -\tau_2, \cdots, -\tau_M) \sigma = \gamma \eta^T \sigma,
\]
where $\gamma$ is a non-positive function of $Z, S, Y$ and $t$. Substituting $\nu$ into (32), the integrand of $dw^*$ term in (32) becomes

$$F_S^T \sigma - F_Z \kappa^T - F_Z Z \nu_2^T + F_Y^T \rho \sigma^T (\sigma^T)^{-1} \sigma.$$  

If $\gamma = 0$, then $\nu = \nu_1$. We require that

$$F_{S_1} + e_1^T (\sigma \sigma^T)^{-1} (b - rS)ZF_Z + e_1^T (\sigma \sigma^T)^{-1} \sigma \rho^T F_Y \geq 0.$$  

If $\gamma < 0$, then according to Lemma 7, the position in the first security is zero. We conclude that

$$[F_{S_1} + e_1^T (\sigma \sigma^T)^{-1} (b - rS)ZF_Z + e_1^T (\sigma \sigma^T)^{-1} \sigma \rho^T F_Y] - F_Z Z \gamma = \dot{v}.$$  

Since $F_Z > 0$, this can happen only when

$$[F_{S_1} + e_1^T (\sigma \sigma^T)^{-1} (b - rS)ZF_Z + e_1^T (\sigma \sigma^T)^{-1} \sigma \rho^T F_Y] < 0.$$  

Solving for $\gamma$, we get (28). The argument for (29) should be obvious.  

Comparing (18) with (29), we obtain an additional term for the trading strategy 9. Obviously, this term corresponds to the short-sale constraints on the first security. Our next theorem is the verification theorem parallel to Theorem 7.

**Theorem 9 (Verification Theorem) Suppose the following:**

a) $F : \mathbb{R}^{N+1} \times [0, T] \to \mathbb{R}$ satisfies (25) with $\nu$ being defined as in (28) and boundary conditions (26) and (27) for some constant $Z_0 > 0$;

b) There exist $K, \beta > 0$ such that for all $(y, z, t) \in \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}$,

$$|e^{-y} F(e^y, z, t)| \leq K (1 + |y|^\beta + |z|^\beta), \quad |e^{-y} \dot{f}(e^{-y}, t)| \leq K (1 + |y|^\beta),$$

and $F(y, z, t) \to F(y, z, T)$ as $T \to \infty$;

c) For the $\nu$ defined in a), there exists a solution $Z(t)$ to (13);

d) $b(z, t), \sigma(z, t), \mu(z, t), \rho(z, t), r(z, t), |\kappa(z, t) + \nu(y, z, t)|^2$ and $\kappa(z, t) + \nu(y, z, t)$ satisfy linear growth conditions and local Lipschitz condition for $(y, z) \in (0, +\infty) \times \mathbb{R}^N$.

Then there exists a solution to $(P)$ and $(P_\nu)$, and $(P)$ is equivalent to $(P_\nu)$, i.e. $\nu$ defines a minimax local super-martingale measure. The optimal policies can be computed from $F$ as in (29).

**Proof.** First, $Z(t)$ is well defined from the assumption c). Second, $F$ is continuously differentiable, so applying Ito's lemma and using the definition of $\nu$, we obtain

$$F(Z(t), Y(t), S(t))/B(t) + \int_0^t \dot{f}(Z(s)^{-1}, s)/B(s) ds$$

$$= F(0)/B(0) + \int_0^t (F_S^T \sigma - F_Z \kappa^T + F_Y^T \rho)/B(s) dw^*(s).$$

35
From our construction of \( \nu \), the integrand of the \( dw^* \) term must lie in \( \text{span}(\sigma) \), i.e. it can be written as \( \psi \sigma \), and furthermore \( \psi_1 \geq 0 \). If we can show that

\[
E \left[ \xi_\nu(T)W/B(T) + \int_0^T \xi_\nu(t)c(t)/B(t)dt \right] = W_0,
\]

then we can conclude from the Lagrangian theory that \( c(t) = \dot{f}(Z(t)^{-1}, t), W = \dot{g}(Z(T)^{-1}) \) is an optimal solution to \( (P_\nu) \) with \( \lambda = 1/Z_0 \). However, the above equation is guaranteed by assumptions b) and d). A detailed derivation can be found in Cox and Huang [1987b]. □

**Remark 5** In general, the PDE involved in the verification Theorem 9 is a quasi-linear parabolic PDE, typically with a free boundary defined by the boundary of \( \Omega_1 \) or \( \Omega_2 \). We still find that this approach is advantageous compared to dynamic programming, where the Bellman equation is highly non-linear and also has a free boundary arising from the short-sale constraints.

In summary, the exact portfolio policies can be obtained by solving a quasi-linear partial differential equation (Theorem 7 and 9). In some special cases such as log-utility preferences, one may find an explicit solution by solving the dual problem. The PDE with free boundary arising from the short-sale constraints is non-trivial. However, we expect that the quasi-linear PDE is easier to implement numerically than the Bellman equation obtained in dynamic programming.

6. Conclusion

We conclude this paper by pointing out some possible future extensions. First, we expect that our results can be extended to include stochastic endowments. Second, we expect that an infinite horizon problem can also be formulated and solved using the martingale approach. Both extensions will be our future research interests.

References


Appendix A.

**Proof of Theorem 1.** We assume for simplicity that $B(t) \equiv 1$, or else, we can always treat $(c^{*}, W^{*})$ as a discounted consumption-final wealth pair. Let $\eta_{\nu}(t) = \xi_{\nu}(t)/\xi(t)$. Since $(c^{*}, W^{*})$ is optimal for $(\nu^{*},)$, we have by Bayes rule that

$$E_{Q} \left[ \int_{0}^{T} \eta_{\nu}(t)c^{*}(t)dt + \eta_{\nu}(T)W^{*} \right] = W_{0}.$$

Define $X(t) = \eta_{\nu \cdot}^{-1}(t)E_{Q} \left[ \int_{t}^{T} \eta_{\nu \cdot}(s)c^{*}(s)ds + \eta_{\nu \cdot}(T)W^{*} \big| \mathcal{F}_{t} \right]$, then according to the martingale representation theorem, there exists an adapted process $\psi(t)$ such that $\int_{0}^{T} |\psi(t)|^{2}dt < \infty$ a.s., and

$$\int_{0}^{t} \eta_{\nu \cdot}(s)c^{*}(s)ds + \eta_{\nu \cdot}(t)X(t) = W_{0} + \int_{0}^{t} \psi(s)^{T}dw^{*}(s),$$

where the right-hand-side is a martingale under $Q$. Applying Ito's lemma, we obtain that

$$\int_{0}^{t} c^{*}(s)ds + X(t) = W_{0} + \int_{0}^{t} \theta(s)^{T}(dw^{*}(s) - \nu^{*}dt),$$

where $\theta(t) = (\int_{0}^{t} \eta_{\nu \cdot}(s)c^{*}(s)ds - Y(t))\nu^{*}(t)/\eta_{\nu \cdot}(t) + \psi(t)/\eta_{\nu \cdot}(t)$ and $Y(t) = W_{0} + \int_{0}^{t} \psi^{T}(s)dw^{*}(s)$.

We intend to show that $\theta$ lies in span($\sigma$).

According to Lemma 4, $\nu^{*}$ is also the solution to

$$\inf_{\nu \in \Pi \nu} J(\nu),$$

where $J(\nu) = E \int_{0}^{T} \tilde{u}(\lambda_{0}^{t}, t) + \tilde{V}(\lambda_{0}^{t}, t) + \lambda_{0}W_{0}$. We want to perturb $\nu$ around $\nu^{*}$ to get an optimality relation for $\nu^{*}$ as in Xu [1988]. Let us fix a $\nu$ such that $\xi_{\nu} \in \Pi$, and define a stopping time $t_{n}$ and $\nu^{*}$ for $\epsilon > 0$ as follows,

$$t_{n} = \inf \{ t : \int_{0}^{t} c(s)ds + \int_{0}^{t} |\nu(s) - \nu^{*}(s)|^{2}ds + \int_{0}^{t} \eta_{\nu \cdot}(s)c^{*}(s)ds + \int_{0}^{t} |\theta(s)^{T}(dw^{*}(s) - \nu^{*}(s))ds| \geq n \},$$

$$\nu^{*} = \nu^{*} + \epsilon(\nu - \nu^{*})1_{\{t \leq t_{n}\}}.$$

Obviously, $\xi_{\nu^{*}} \in \Pi$, and

$$\xi_{\nu^{*}}(t) = \xi_{\nu^{*}}(t) \exp\{-\epsilon\beta(t) - \epsilon^{2}/2 \int_{0}^{t} |\nu(s) - \nu^{*}(s)|^{2}ds\},$$

where $\beta(t) = \int_{0}^{t} (\nu(s) - \nu^{*}(s))^{T}(dw^{*}(s) - \nu^{*}(s)ds).$ Furthermore, $J(\nu^{*})$ reaches minimum at $\epsilon = 0$. This implies that

$$0 \leq \lim_{\epsilon \downarrow 0} \frac{J(\nu^{*}) - J(\nu^{*})}{\epsilon} \leq \lim_{\epsilon \downarrow 0} E \int_{0}^{T} \tilde{u}(\lambda_{0}^{t}, t)\lambda_{0} \xi_{\nu^{*}}(t) - \xi_{\nu^{*}}(t)dt + \tilde{V}(\lambda_{0}^{t}, t)\lambda_{0} \xi_{\nu^{*}}(T) - \xi_{\nu^{*}}(T) \epsilon.$$
where the second inequality follows from the convexity of $\tilde{u}$ and $\tilde{V}$. Since $|\beta(t)| \leq n$ and $\int_0^{t\wedge n} |\nu(s) - \nu(s)|^2 ds \leq n$, we have
\[
|\tilde{u}'(\lambda_0 \xi_{\nu^*}(t), t)\lambda_0 \frac{\xi_{\nu^*}(t) - \xi_{\nu^*}(t)}{\epsilon} | \leq \frac{|\tilde{u}'((\lambda_0 - \delta_{\lambda_0/2})\xi_{\nu^*}(t), t)|}{|\lambda_0 \xi_{\nu^*}(t)|} e^{-m - \epsilon^2 t/2} \
\leq \frac{|\tilde{u}'((\lambda_0 - \delta_{\lambda_0/2})\xi_{\nu^*}(t), t)|}{|\lambda_0 \xi_{\nu^*}(t)|} e^{m(2n + \epsilon n/2)} \
\leq \frac{|\tilde{u}'((\lambda_0 - \delta_{\lambda_0/2})\xi_{\nu^*}(t), t)|}{|\lambda_0 \xi_{\nu^*}(t)|} e^{m(2n + n/2)},
\]
for sufficiently small $\epsilon$ such that $\lambda_0 e^{-m - \epsilon^2 n/2} > \lambda_0 - \delta_{\lambda_0/2}$ and $0 < \epsilon < 1$. $V$ satisfies a similar inequality. According to Assumption $T$, the right-hand-side (35) is integrable, we can past the limit through the integration, which gives us
\[
E \left[ \int_0^T \beta(t) \xi_{\nu^*}(t) c^*(t) + \beta(t_n) \xi_{\nu^*}(T) W^*(T) \right] \geq 0.
\]
To simplify (36), we define a new measure $Q_n$,
\[
Q_n(A) = \int_A \xi_{\nu^*}(\omega, t_n) P(d\omega).
\]
Following the Girsanov's theorem, $Q_n$ is a probability measure, and $w^\nu_{\nu^*}(t) = \nu^*(t \wedge t_n) - \int_0^{t\wedge n} \nu^*(s) ds$ is a Brownian motion under $Q_n$. Next, observe that (33) implies
\[
E_Q \left[ \beta(t_n) \int_0^T \eta_{\nu^*}(s) c^*(s) ds + \beta(t_n) \eta_{\nu^*}(T) X(T) \right] = E_Q \beta(t_n) Y(T).
\]
However, since $Y(t)$ is a martingale under $Q$, we have $E_Q \beta(t_n) Y(T) = E_Q \beta(t_n) Y(t_n)$. It then follows from (33) and (37) that
\[
E_Q \beta(t_n) \left[ \int_0^T \eta_{\nu^*}(s) c^*(s) ds + \beta(t_n) \eta_{\nu^*}(T) X(T) \right] = E_Q \beta(t_n) \eta_{\nu^*}(t_n) X(t_n).
\]
Now applying Itô's lemma to $\beta(t) \int_0^t c^*(s) ds$ for $t \leq t_n$, we have
\[
\beta(t) \int_0^t c^*(s) ds = \int_0^t \gamma(s)^T dw^\nu_{\nu^*}(s) + \int_0^t \beta(s) c^*(s) dt,
\]
where $\gamma(t) = (\int_0^t c^*(s) ds)(\nu(t) - \nu^*(t))$. Since the first term (stopped at $t_n$) on the right-hand-side of the above equation is a martingale under $Q_n$, we obtain
\[
E_Q \beta(t_n) \int_0^{t_n} c^*(s) ds = E_Q \int_0^{t_n} \beta(t) c^*(s) ds.
\]
Finally, substituting (34), (38) and (40) into (36), we obtain
\[
0 \leq E_Q \left[ \int_0^T \beta(t) \eta_{\nu^*}(t) c^*(t) dt + \beta(t_n) \eta_{\nu^*}(T) X(T) \right] = E_Q \left[ \int_0^{t_n} \beta(t) \eta_{\nu^*}(t) c^*(t) dt + \beta(t_n) \eta_{\nu^*}(t_n) X(t_n) \right] = E_Q \beta(t_n) \left[ \int_0^{t_n} c^*(t) dt + X(t_n) \right] = E_Q \left[ \int_0^{t_n} \theta(t)^T (\nu(t) - \nu^*(t)) dt \right].
\]
for all \( \nu \in \ker(\sigma) \). This implies that there exists an adapted process \( \psi(t) \) such that \( \theta(t) = \sigma(t)^T \psi(s) \) for \( t \in [0, t_n] \). However, \( t_n \to T \) as \( n \to \infty \), we conclude that \( \theta \in \text{span}(\sigma) \). Therefore, equation (34) becomes

\[
\int_0^t c^*(s)ds + X(t) = \mathcal{W}_0 + \int_0^t \theta(s)^T dw^*(s).
\]

Clearly, \( (c^*, \mathcal{W}^*) \) is marketed. This implies that \( (c^*, \mathcal{W}^*) \) is also an optimal solution to \( (P) \), since \( \text{val}(P_{c^*}) \geq \text{val}(P) \). We conclude that \( \xi_{c^*} \) defines a minimax local martingale measure. 

\[\text{PROOF OF THEOREM 3.} \] Note first that for any \( \mathcal{W} \in C_2(\mathcal{W}_0) \), the wealth process \( \mathcal{W}(t) \) for \( \mathcal{W} \) is a bounded and strictly positive martingale under \( Q \). This implies that there exists an adapted process \( \psi \in \text{span}(\sigma) \), \( \mathbb{E}_Q \int_0^T |\psi(t)|^2 dt < \infty \), such that

\[
\mathcal{W} = \mathcal{W}_0 \exp \left( \int_0^T \psi(t)dw^*(t) - \int_0^T |\psi(t)|^2dt \right) \equiv \mathcal{W}_\psi.
\]

Therefore, we can define the objective function as a functional of \( \psi \):

\[
f(\psi) = EV(\mathcal{W}_\psi),
\]

where \( \psi \) lies in \( \text{span}(\sigma) \) and \( \mathbb{E}_Q \int_0^T |\psi(t)|^2 dt < \infty \). We make the following two claims.

\textbf{Claim A:} \( f(\psi) \) is concave in \( \psi \). Let \( \lambda_1 + \lambda_2 = 1 \) and \( \lambda_i \geq 0 \), by convexity, we have

\[
|\lambda_1\psi_1(t) + \lambda_2\psi_2(t)|^2 \leq \lambda_1|\psi_1(t)|^2 + \lambda_2|\psi_2(t)|^2.
\]

Since \( R_V(x) \geq 1 \), we know that \( V(e^x) \) is concave in \( x \). Thus,

\[
f(\lambda_1\psi_1 + \lambda_2\psi_2) &= EV \left( \mathcal{W}_0 \exp \left( \int_0^T (\lambda_1\psi_1(t) + \lambda_2\psi_2(t))dw^*(t) - \frac{1}{2} \int_0^T |\lambda_1\psi_1(t) + \lambda_2\psi_2(t)|^2 dt \right) \right) \\
&\geq EV \left( \mathcal{W}_0 \exp \left( \int_0^T (\lambda_1\psi_1(t) + \lambda_2\psi_2(t))dw^*(t) - \frac{1}{2} \int_0^T \lambda_1|\psi_1(t)|^2 + \lambda_2|\psi_2(t)|^2 dt \right) \right) \\
&= EV(\mathcal{W}^\lambda_1\psi_1^\lambda_2\psi_2) \\
&\geq \lambda_1 f(\psi_1) + \lambda_2 f(\psi_2).
\]

\textbf{Claim B:} \( f(\psi) \) is upper semi-continuous in \( L^2(\Omega \times [0,T], P(d\omega) \times dt) \). In fact, since \( E \int_0^T |\kappa(t)|^2 dt \leq 2E \exp \left( \int_0^T \frac{1}{2}|\kappa(t)|^2 dt \right) < \infty \), if \( \psi_n \to \psi^* \) in \( L^2(\Omega \times [0,T], P(d\omega) \times dt) \), then \( \int_0^T \psi_n(t)^T dw(t) \) and \( \int_0^T \psi_n(t)^T \kappa(t) dt \) also converge to \( \int_0^T \psi^*(t)^T dw(t) \) and \( \int_0^T \psi^*(t)^T \kappa(t) dt \) in \( L^2(P(d\omega) \times dt) \), respectively. Therefore, there exists a subsequence, denoted again by \( n \), such that \( \psi_{n_k} \to \psi^* \) in \( L^2(\Omega \times [0,T], P(d\omega) \times dt) \), \( \int_0^T \psi_{n_k}(t)^T dw(t) \) and \( \int_0^T \psi_{n_k}(t)^T \kappa(t) dt \) converge to \( \int_0^T |\psi^*(t)|^2 dt \), \( \int_0^T \psi^*(t)^T dw(t) \) and \( \int_0^T \psi^*(t)^T \kappa(t) dt \) almost surely, respectively. It should be clear that \( \psi^* \in \text{span}(\sigma) \). Since \( V \) is bounded from above, we can apply Fatou’s lemma to get

\[
\lim_{n \to \infty} \sup f(\psi_n) \leq E \lim_{n \to \infty} \sup V(\mathcal{W}_{\psi_n}) = f(\psi^*).
\]

We now proceed the proof for the existence as follows. According to Lemma 5, we can find a sequence of \( \psi_n \), such that \( \mathcal{W}_{\psi_n}(t) \in C_2(\mathcal{W}_0) \) and \( f(\psi_n) \to \text{val}(P) \). However, for any \( \mathcal{W}_{\psi_n} \in C_2(\mathcal{W}_0) \),

\[41\]
if we set $T_m^n = \inf\{t : \int_0^t |\psi_n(s)|^2ds \geq m\}$ and set

$$W_{\psi_n}^m = W_0 \exp \left( \int_0^{T_m^n} \psi_n(t)^T dw^*(t) - \int_0^{T_m^n} |\psi_n(t)|^2 dt \right),$$

then, $W_{\psi_n}^m \to W_{\psi_n}$ almost surely and $EV(W_{\psi_n}^m) \to EV(W_{\psi_n})$ as $m \to \infty$, since they bounded from both above and below by some strictly positive constants. Therefore, we may as well assume that $E \int_0^T |\psi_n(t)|^2 dt < \infty$. Now, by the concavity of $V(e^x)$, we have

$$f(\psi_n) \leq V \left( W_0 \exp(\int_0^T \psi_n(t)dw^*(t) - \frac{1}{2} \int_0^T |\psi_n(t)|^2 dt) \right)$$

$$= V \left( W_0 \exp(-E \int_0^T \kappa(t)^T \psi_n(t) dt - \frac{1}{2} E \int_0^T |\psi_n(t)|^2 dt) \right).$$

Therefore, if $E \int_0^T |\psi_n(t)|^2 dt \to \infty$, then $\lim \sup f(\psi_n) \leq V(0)$, which is certainly less than $\text{val}(P)$. This implies that $E \int_0^T |\psi_n(t)|^2 dt$ must be bounded. Hence, $\{\psi_n\}$ is weakly compact in $L^2(P(d\omega) \times dt)$. Theorem 2.6.1 of Balakrishnan [1981] allows us to claim that there exists $\psi^*, \int_0^T |\psi^*(t)|^2 dt < \infty$ and $\psi^* \in \text{span}(\sigma)$ such that $f(\psi^*) = \text{val}(P)$. □

**Proof of Theorem 4.** Note that $R_u \leq 1$ and $R_V \leq 1$ imply that $u$ and $V$ are unbounded from above as $z \to \infty$. Given Lemma 6, we need only to show that

$$\inf_{\nu} J(\nu) = \inf_{\xi \in \Pi} E \left[ \int_0^T \tilde{u}(\lambda_\nu(t)/B(t), t) dt + \tilde{V}(\lambda_\nu/B(T)) \right]$$

has an optimal solution for any $\lambda > 0$. Since the proof for this statement is similar to the proof for Theorem 3, we sketch the proof.

First, we argue that we only need to consider those $\nu$ such that $E \int_0^T |\nu(t)|^2 dt < \infty$. Second, it can be verified that $J$ is convex in $\nu$, lower semi-continuous in $L^2(P(d\omega) \times dt)$ due to the fact that $\tilde{u}(e^x)$ and $\tilde{V}(e^x)$ are convex, decreasing and bounded from below. Finally, if $\nu_n$ is the sequence such that $J(\nu_n)$ reaches its infimum, then $\nu_n$ is weakly compact in $L^2(P(d\omega) \times dt)$. Theorem 2.6.1 of Balakrishnan [1981] allows us to claim that there exists a solution to $D$. □

**Proof of Theorem 5.** For simplicity, we assume $B(t) \equiv 1$. First, consider the following maximization problem:

$$\sup_{\xi_\phi} E - \tilde{V}(\lambda \xi_\phi(T)), \quad (A)$$

where $\xi_\phi(t) = \xi(t)\eta_\phi(t)$, with $\eta_\phi(t) = 1 + \int_0^t \phi(s)^T dw^*(s)$, $\eta_\phi(t) \geq 0$ and $\phi \in \ker(\sigma)$. Note that $\xi_\phi$ may not be in $\Pi$, since $\xi_\phi(T)$ may be zero with some strictly positive probability. We want to show that the solution to $(A)$ exists and lies in $\Pi$, which implies that $(A)$ is equivalent to maximizing over $\Pi$.

Since $-\tilde{V}$ is strictly concave and strictly increasing, program $(A)$ is very similar to program $(P)$ in Theorem 3, except that $\phi$ lies in the span($\sigma$), whereas $\phi$ in $(A)$ lies in the $\ker(\sigma)$. We claim that for fixed $\lambda > 0$, the dual program of $(A)$ is of the following form:

$$\inf_{\mu > 0} \left\{ \inf_{\phi} E - V(\mu X_\phi(T)) + \mu \lambda \right\}. \quad (A')$$

42
where \( X\phi(T) = \exp(\int_0^T \psi(t)dw^* - \int_0^T |\psi(t)|^2 dt), \int_0^T |\psi(t)|^2 dt < \infty \) a.s. and \( \psi \) lies in the span(\( \sigma \)).

The proof for this statement can be carried out exactly as we did for the proof of Theorem 1.

However, according to Theorem 3, for any \( \mu > 0 \), the solution to \( \inf_{\phi} E - V(\mu X\phi(T)) \) exists.

Define \( h(\mu) = \inf_{\phi} -E V(\mu X\phi(T)) + \lambda \mu \), it is easy to prove as we did in Lemma 6 that \( h \) is convex, and \( h(+\infty) = +\infty \) (since \( V \) is bounded from above and unbounded from below). Moreover, since

\[
    h(\mu) \geq -\mu^{1-b} \sup_{\phi} X\phi^{1-b}_{1-b} + \lambda \mu - C,
\]

and \( \sup_{\phi} X\phi^{1-b}_{1-b} \) exists and negative according to Theorem 3, we conclude that \( h(0) = +\infty \). Therefore, \( \inf_{\lambda > 0} h(\lambda) \) exists, which implies that (A') has a solution.

By the duality between (A) and (A'), we infer that (A) has a solution. Moreover, there exists \( X\phi \) such that the solution to (A) coincides with the solution to

\[
    \sup_{\phi} E - \tilde{V}(\lambda \xi\phi(T))
\]

s.t. \( EX\phi(T)\xi\phi(T) \leq 1 \).

By the Lagrangian theory, there exists a constant \( \gamma > 0 \) such that

\[
    -\lambda \tilde{V}'(\lambda \xi\phi(T)) = \gamma X\phi(T).
\]

Therefore \( \xi\phi(T) = V'(\gamma X\phi(T)/\lambda)/\lambda > 0 \). This implies that \( \xi\phi \in \Pi \). Thus, the solution to (A) exists and lies in \( \pi \).

Next, we want to prove that the dual problem (D')

\[
    \inf_{\lambda > 0} \inf_{\xi\phi \in \Pi} E \tilde{V}(\lambda \xi\phi(T)) + \lambda W_0
\]

has a solution, since this will then imply the existence of a minimax local martingale measure. To do so, let

\[
    h_1(\lambda) = \inf_{\xi\phi \in \Pi} E \tilde{V}(\lambda \xi\phi(T)) + \lambda W_0.
\]

It is easy to see that \( h_1(\lambda) \) is convex. Since \( C + B^{\frac{1}{1-b'}} \leq V(z) \leq C + B^{\frac{1}{1-b}} \), we have

\[
    C + B^{\frac{1}{1-b'}} \tilde{V}(z) \leq C + B^{\frac{1}{1-b}} \tilde{V}(z) \leq \tilde{V}(z) \leq C + B^{\frac{1}{1-b'}} \leq C + B^{\frac{1}{1-b}} \tilde{V}(z).
\]

Therefore

\[
    C + B^{\frac{1}{1-b'}} \tilde{V}(z) + \lambda W_0 \leq h_1(\lambda) \leq C + B^{\frac{1}{1-b'}} \tilde{V}(z) + \lambda W_0, \tag{41}
\]

where we have used the fact that \( E[\xi\phi(T)] \leq 1 \) for establishing the first part of inequality. Equation (41) implies that \( h_1(0) = C \) and \( h_1(+\infty) = +\infty \). Moreover, \( h(\lambda) \) is decreasing at 0, since for sufficiently small \( \lambda \), the right-hand-side of the above inequality is strictly less than \( C \). Thus \( h_1 \) attains its minimum at some \( \lambda_0 > 0 \). We conclude that (D') has a solution, i.e. the minimax local martingale measure exists. \( \blacksquare \)
ESSAY TWO:
CONVERGENCE FROM DISCRETE TO CONTINUOUS
TIME FINANCIAL MODELS

Abstract This essay generalizes the Cox, Ross and Rubinstein [1979] binomial model, and
develops a theory of convergence from discrete time multivariate and multinomial models to
continuous time multi-dimensional diffusion models in the context of contingent claim pricing
and consumption-portfolio selection. The key to our approach is to approximate the \( N \)
dimensional diffusion price process by a sequence of \( N \)-variate and \( N + 1 \)-nomial processes. It
is shown that contingent claim prices and optimal consumption-portfolio policies derived from
discrete time models converge to their corresponding continuous time limits. In contrast to
solving a partial differential equation (PDE) with possibly more than one state variable, this
approach provides a simpler numerical procedure for computing contingent claim prices and
optimal consumption-portfolio policies.

1. Introduction

For the purpose of modeling securities markets and trading behaviors, continuous and discrete
time models have been the two approaches to use. While the continuous time approach has
been prevalent in a large part of financial economic theory, it is widely felt that continuous
and discrete time models are approximations to each other. Kreps [1982] and Merton [1982]
contain excellent discussions on general issues of convergence from discrete to continuous time
financial models. A convergence theory as such also has important practical implications in that
practitioners would feel comfortable using the results of continuous time theory when trades
occur discretely in time and periods between trades are sufficiently small.

A body of convergence theory has focused on the study of limits of discrete time models
in which asset prices evolve as in continuous time models while trades are limited to discrete
dates, e.g., Merton [1975], Rubinstein [1978], Boyle and Emanuel [1980] and Leland [1982]. In
a completely different approach, Cox, Ross and Rubinstein (CRR) [1979] develops a theory of
convergence from binomial price processes to a geometric Brownian motion, and show that the
Black-Scholes [1973] option pricing formula is the limit of the discrete time binomial option
pricing formula.\(^1\) The CRR's binomial approximation for one-dimensional geometric Brownian
motion can be easily generalized for general one-dimensional diffusion processes. However,
to date little is known about how their technique can be generalized for multi-dimensional
diffusion price processes.\(^2\) While convergence theory has been established for option prices

---

\(^1\)The binomial or two state variables approach was also developed independently in Sharp [1978] and Rendle-
mam and Bartter [1979].

\(^2\)Several recent papers, Evnine [1981], Boyle [1988] and Madan, Milne and Shafir [1988], have attempted
to approximate a multi-dimensional geometric Brownian motion; but none of them have provided satisfactory
answers. See the discussion below.
or more generally contingent claim prices in the case of one-dimensional price process, less is known about convergence for optimal consumption and portfolio policies.

The aim of this paper is to generalize the CRR's binomial approximation to the case of multi-dimensional price system in the context of contingent claim pricing and consumption-portfolio selection. Specifically, we want to develop a theory of convergence from discrete time multivariate and multinomial models to continuous time multi-dimensional diffusion models for both contingent claim prices and optimal consumption-portfolio policies.

There are two reasons for establishing such a theory of convergence. The first one is due to the intuitive appeal of discrete time models. While the economic intuition that continuous trading in stocks and bond can span infinitely many states of nature is difficult to convey without using advanced mathematics, the simple discrete time binomial model provides an easy way of explaining how uncertainties are resolved in the continuous time model. The second reason is that the discrete time models often provide an elegant numerical alternative to the PDE's produced by continuous time models. In contrast to solving a PDE with possibly more than one state variable, the multinomial approximation approach provides a simpler numerical procedure for computing contingent claim prices and optimal consumption-portfolio policies.\footnote{Various trinomial or multinomial models also have been used in option pricing literature due to their numerical simplicity, e.g., Boyle [1988] and Hull and White [1988]. An important distinction between our approximations and those mentioned here is that those approximations usually let the discrete time price processes converge to the corresponding continuous time price process under risk neutral probability measure, i.e., the expected return for the risky asset must be equal to the riskless rate. This allows them to price options by taking expectations.}

Similar to CRR, we use weak convergence as our convergence tool. The diffusion model considered in this essay consists of $N$ risky stocks and one riskless bond, where the stocks and the bond form a dynamically complete securities market. We approximate the $N$-dimensional diffusion process for stock prices by a sequence of $N$-variate and $N + 1$-nomial processes. Thus, the stocks and bond in discrete time models also form a dynamically complete securities market. Dynamic completeness allows us to price contingent claims whose payoffs depend upon the value of assets in general form. The main results of this essay are as follows: 1) In the case of contingent claim pricing, we show that contingent claim prices and replicating portfolio strategies derived from the discrete time models converge weakly to the corresponding contingent claim price and replicating portfolio strategy of the limiting continuous time model. In particular, the discrete time contingent claim prices and replicating portfolio strategies calculated at time 0 converge to the corresponding continuous time limits. 2) In the case of consumption-portfolio selection, we show that optimal consumption policies, optimal portfolio policies and optimal conditional expected utilities derived from the discrete time models converge weakly to their corresponding continuous time limits.
The question on how to approximate a price system with two stocks and one bond, where stock prices follow two correlated geometric Brownian motions, has been raised in the literature for a long time, and no satisfactory answer has yet been found. Intuitively, one would think that if one geometric Brownian motion can be approximated by one binomial process, two geometric Brownian motions should be approximated by two binomial processes. This leads to a quadnomial process. Since there are four uncertain states following each trading date and there are only two stocks and one bond traded, markets cannot be completed, and options cannot be priced by arbitrage. This phenomenon is counter-intuitive to the familiar continuous time model from which we already know that markets can be completed by continuous trading in two stocks and one bond.

To overcome this problem, Evnine [1981] proposes a “multiple” binomial model which approximates the increments of two geometric Brownian motions by three subsequent moves. His idea is to let stock one move stochastically first, while letting stock two grow at the riskless rate. He then lets stock two move stochastically, while letting stock one grow at the riskless rate. Finally, he lets both stocks move together so as to capture the correlation. Although he manages to show that the discrete price process matches the continuous price process in distribution in the limit, it is intuitively not clear why the two stocks have to move separately. More important, the dynamic portfolio trading strategy as implied by Evnine’s model is always indeterminant, since the return on one of the two stocks always correlates perfectly with the riskless bond. Madan, Milne and Shefrin [1988] construct an $N + 1$-nomial process for $N$ stocks such that the discrete time price process for each individual stock converges weakly to a one-dimensional geometric Brownian motion. Since they fail to specify the correlations among different assets and establish a joint convergence for $N$ stock prices, their model does not imply convergence for general contingent claims prices, for example, option on the maximum of two stocks.

In this essay, we resolve this old controversy by showing that an $N$-dimensional diffusion process for stock prices can be approximated by an $N$-variate and $N + 1$-nomial process. We utilize the fact that the increments of $N$ independent Brownian motions can be approximated by $N$ uncorrelated, not necessarily independent, random variables. For example, we can use a trinomial model to approximate two geometric Brownian motions. A crucial distinction between our approximation and that of Madan, Milne and Shefrin is that the implicit Arrow-Debreu state price processes derived from the discrete time models converge weakly to the

---

4 The number $N$ is called the martingale multiplicity as discussed in Duffie and Huang [1985]. When $N + 1$-nomial process is used, the martingale multiplicities in both discrete and continuous time models are exactly the same.
corresponding continuous time limit. Mathematically, one could approximate an \( N \)-dimensional diffusion process by an \( N \)-variate and \( M \)-nomial process, where \( M \) could be any integer greater than or equal to \( N + 1 \). However, to preserve market completeness and to allow contingent claims to be priced by arbitrage, \( N + 1 \) is the right number to use.

The rest of this essay is organized as follows. Section 2 sets out the definition of weak convergence and derives some of the basic properties of weak convergence to be used in later sections. Section 3 presents a multivariate and multinomial approximation to the multi-dimensional diffusion process for stock prices. Section 4 deals with convergence of contingent claim prices and replicating portfolio strategies, while Section 5 establishes convergence of optimal consumption-portfolio polices and optimal conditional expected utilities. We conclude this essay in Section 6 with some comments and suggestions for future research.

2. Preliminaries on Weak Convergence

This section sets out the definition of weak convergence, and derives some of the basic properties to be used in later sections. In particular, we will present a martingale central limit theorem, which is to be used as a tool for establishing convergence from a sequence of Markov chains to a diffusion process.

Let the sample space be \( D^M[0,1] \), which is the space of functions from \([0,1]\) to \( \mathbb{R}^M \) that are right-continuous with left limits (RCCLL). Since diffusion processes have continuous sample paths and the discrete time processes to be considered in later sections are piecewise constant, the choice of \( D^M[0,1] \) as our sample space is appropriate for our purpose. The topology to be used for \( D^M[0,1] \) is the Skorohod topology (as explained by Billingsley [1968]). The Borel field generated by the Skorohod topology is denoted by \( \mathcal{D} \). Let \( X^n \) be a sequence of stochastic processes in \( D^M[0,1] \), and let \( P_n \) be the probability measure on \( D^M[0,1] \) associated with \( X^n \), i.e. for any \( A \in \mathcal{D} \), \( P_n(A) \) is the probability of event \( \{X^n \in A\} \). A sequence of stochastic processes \( X^n \) in \( D^M[0,1] \) is said to converge to \( X \) in \( D^M[0,1] \) weakly or in distribution, denoted by \( X^n \Rightarrow X \), if for any bounded continuous function \( h \) mapping from \( D^M[0,1] \) to \( \mathbb{R} \), we have \( E_n[h(X^n)] \to E[h(X)] \), where \( E_n \) denotes the expectation under \( P_n \), and \( E \) denotes the expectation under \( P \) associated with \( X \).

Since the multinomial processes to be used in later sections are Markov chains, we would like to have a mathematical tool for establishing weak convergence from a sequence of Markov chains to a diffusion process. The martingale central limit theorem developed in Ethier and Kurtz [1986] turns out to exactly what we need. Let \( X \) be determined by the following stochastic
differential equation:
\[ dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t, \]  
(1)

where \( b(x) \) is an \( M \times 1 \) vector, \( \sigma(x) \) is an \( M \times N \) matrix for all \( x \in \mathbb{R}^M \) and \( w \) is an \( N \)-dimensional Brownian motion. For the rest of the paper, we use \( \mathbf{T} \) to denote the transpose of a matrix and \( |A| \equiv \text{trace}(A^\mathbf{T}A)^{\frac{1}{2}} \) to denote the norm of matrix \( A \).

The following proposition states the martingale central limit theorem, whose proof can be found in Ch. 7, Ethier and Kurtz \[1986]\.

**Proposition 1 (Martingale Central Limit Theorem)** Let \( b \) and \( \sigma \) be continuous such that (1) admits a unique weak solution.\(^5\) Let \( X \) be the weak solution to (1) with \( X_0 = x_0 \). Suppose that \( X^n \) is a sequence of Markov processes with sample path in \( \mathbb{D}^M[0,1] \). Let \( L^n \) and \( A^n \) be a \( N \times 1 \) and a symmetric \( N \times N \) matrix-valued processes, respectively, such that each of their elements has sample path in \( \mathbb{D}^1[0,1] \) and \( A^n_t - A^n_s \) is non-negative definite for \( t > s \geq 0 \). Define
\[ \tau^n_a = \inf\{t \leq T : |X^n_t| \geq a \text{ or } |X^n_{T-}| \geq a\}, \]
and suppose further that
\[ a) \, X^n_0 \to x_0 \text{ in distribution}; \]
\[ b) \, M^n \equiv X^n - L^n \text{ and } M^n M^n \mathbf{T} - A^n \text{ are martingales}; \]
\[ c) \, \text{for all } q > 0, \]
\[ \lim_{n \to \infty} E_n \left[ \sup_{t \leq \tau^n_a} |X^n_t - X^n_{\tau^n_a}|^2 \right] = 0, \]
\[ \lim_{n \to \infty} E_n \left[ \sup_{t \leq \tau^n_a} |L^n_t - L^n_{\tau^n_a}|^2 \right] = 0, \]
\[ \lim_{n \to \infty} E_n \left[ \sup_{t \leq \tau^n_a} |A^n_t - A^n_{\tau^n_a}| \right] = 0; \]
\[ d) \]
\[ \sup_{t \leq \tau^n_a} |L^n_t - \int_0^t b(X^n_s)ds| \to 0, \text{ as } n \to \infty \text{ in probability for all } q > 0, \]
\[ \sup_{t \leq \tau^n_a} |A^n_t - \int_0^t a(X^n_s)ds| \to 0, \text{ as } n \to \infty \text{ in probability for all } q > 0. \]

Then \( X^n \Rightarrow X \).

\(^5\)Equation (1) is said to have a weak solution \( X \) with initial distribution \( \mu \) if there exists \((\Omega, \{\mathcal{F}_t\}, P)\) such that \( w \) is an \( \{\mathcal{F}_t\}\)-Brownian motion, \( X_0 \) has law \( \mu \), and \( X \) satisfies
\[ X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s, \]
where all the stochastic integrals are well-defined. The weak solution to (1) is said to be unique if whenever \([X_t, t \geq 0]\) and \([X'_t, t \geq 0]\) are two solutions such that the distributions of \( X_0 \) and \( X'_0 \) are the same, then the distributions of \( X_t \) and \( X'_t \) are the same, see Rogers and Williams \[1987\].
The following two types of conditions are sufficient for (1) to have a unique weak solution
(see Rogers and Williams [1987]):

i) $b$ and $\sigma$ satisfy a uniform Lipschitz condition, i.e. there exists a constant $L > 0$ such that
for all $x, y \in \mathbb{R}^M$,
\[ |b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq L|x - y|; \]

ii) $a(x) = \sigma(x)\sigma(x)^T$ is strictly positive definite for each $x \in \mathbb{R}^M$ and there exists a constant
$K$ such that for all $x \in \mathbb{R}^M$,
\[ |a(x)| \leq K(1 + |x|^2), \quad |b(x)| \leq K(1 + |x|). \]

Before leaving this section, we record two lemmas to be used in later sections, whose proofs
can be found in Billingsley [1968].

Lemma 1 (Continuous Mapping Theorem) Assume that $X^n \Rightarrow X$.

i) Suppose that $h(x)$ is a continuous function mapping from $\mathbb{R}^N$ to $\mathbb{R}^M$, except for at most
countably many points, and denote $Y^n_t = h(X^n_t)$, $Y_t = h(X_t)$, then $Y^n \Rightarrow Y$;

ii) Suppose that $H$ is a continuous mapping from $D^M[0,1]$ to $\mathbb{R}$, Then, $H(X^n)$ converges
to $H(X)$ in distribution.

Lemma 2 Let $X_n$ and $X \in \mathbb{R}^1$ be random variables, and suppose that $X_n \rightarrow X$ in distribution
as $n \rightarrow \infty$. If $\{X^n\}$ is uniformly integrable, then $E_n(X_n) \rightarrow E(X)$ as $n \rightarrow \infty$. Conversely, if
$X, X_n \geq 0$ and $E_n(X_n) \rightarrow E(X)$ as $n \rightarrow \infty$, then $\{X_n\}$ is uniformly integrable.

3. Multinomial Approximation

In this section we will construct a sequence of $N$-variate and $N + 1$-nomial processes for stock
prices that converges weakly to the $N$-dimensional diffusion price process. Moreover, we will
show that the implicit Arrow-Debreu state price process derived from the discrete time model
converges weakly to the corresponding continuous time limit. This result will play an important
role in establishing weak convergence of contingent claim prices and optimal consumption-
portfolio policies in later sections.

We consider a securities market consisting of $N$ risky stocks and one locally riskless bond.
The price processes for stocks and bond are described by the following stochastic differential
equation:
\[ dS_t = b(S_t) \, dt + \sigma(S_t) \, dw_t, \quad (2) \]
where \( w_t \) is an \( N \)-dimensional standard Brownian motion defined on some probability space \((\Omega, \mathcal{F}, P)\), \( S_t \) is an \( N \)-dimensional price vector for stocks, and \( B_t \) is the price for the bond. We assume that functions, \( b: \mathbb{R}^N \rightarrow \mathbb{R}^{N \times 1} \), \( \sigma: \mathbb{R}^N \rightarrow \mathbb{R}^{N \times N} \) and \( \tau: \mathbb{R}^N \rightarrow \mathbb{R} \), are continuous and \( \sigma \) is nonsingular. We assume further that \( b \) and \( \sigma \) satisfy the linear growth condition, i.e., there exists a constant \( K > 0 \) such that

\[
|b(x)| + |\sigma(x)| \leq K(1 + |z|), \quad \forall z \in \mathbb{R}^N,
\]

and the uniform Lipschitz condition, i.e., there exists a constant \( L > 0 \) such that

\[
|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq L|x - y|, \quad \forall x, y \in \mathbb{R}^N.
\]

Furthermore, we assume that \( \tau \) is non-negative and the time span is \([0, 1]\). All the regularity conditions imposed on the drift term and the diffusion term are readily satisfied for the Black-Scholes' price system.

As in the Black-Scholes' economy, we assume that markets are dynamically complete. That is, contingent claims written on the stocks such as options can be spanned by dynamic trading in stocks and bond.\(^6\) To rule out arbitrage opportunity, we assume that there exists a unique equivalent martingale measure or a risk neutral probability measure for the price system defined by (2) and (3). Moreover, the martingale measure, denoted by \( Q \), has the following form. Let \( \kappa(S_t) = -\sigma(S_t)^{-1}(b(S_t) - \tau_t S_t) \), and \( \xi_t = \exp \left( \int_0^t \kappa_s \, dw_s - \frac{1}{2} \int_0^t |\kappa_s|^2 \, ds \right) \). Then,

\[
Q(A) \equiv \int_A \xi_1(\omega) \, dP.
\]

One can easily verify that the stochastic process \( \{\xi_t\} \) satisfies the following stochastic differential equation:

\[
d\xi_t = \kappa(S_t) \xi_t \, dw_t, \quad \xi_0 = 1.
\]

In the literature, \( \xi_t(\omega) \) is usually interpreted as the implicit Arrow-Debreu state price per unit of probability for a security that pays off one unit of consumption good at time \( t \) and state \( \omega \), but nothing otherwise (see Cox and Huang [1987]).

We now proceed with the construction of a sequence of multivariate and \( N + 1 \)-nomial processes. The basic idea is to approximate the increments of \( N \) independent Brownian motions by \( N \) uncorrelated random variables. We first construct such random variables.

\(^6\)Rigorously speaking, dynamic completeness requires the existence of a unique equivalent martingale measure and a proper choice of the space of feasible trading strategies, see Cox and Huang [1987] for details.
Let $A$ be an $(N + 1) \times (N + 1)$, real and orthogonal matrix such that the last column of $A$ is $(1/\sqrt{N + 1}, \cdots, 1/\sqrt{N + 1})^\top$ (such a matrix always exists!). Define a sequence of $N$-dimensional, independent and identically distributed random vector, $\hat{\varepsilon}_k = (\hat{\varepsilon}_k^1, \cdots, \hat{\varepsilon}_k^N)^\top$, $k = 1, 2, \cdots, n$, as follows. Let $\Omega_k = \{\omega_k^1, \cdots, \omega_{N+1}^1\}$ be the sample space on which $\hat{\varepsilon}_k$ is defined. For $j = 1, \cdots, N$, let

$$P \left[ \hat{\varepsilon}_k^{(s)}(\omega_k^j) = a_{s,j} \sqrt{N + 1} \right] = \frac{1}{N + 1}, \quad s = 1, \cdots, N + 1,$$

where $a_{s,j}$ is the $s$-th element in $j$-th column of $A$. That is, we assign equal probability to each state $\omega_k^j$. It is easy to verify that, for fix $k$, $\hat{\varepsilon}_k^1, \cdots, \hat{\varepsilon}_k^N$ are uncorrelated to each other, and have mean zero and variance 1.

Next, we divide $[0, 1]$ into $n$ equally spaced subintervals, each with a length of $\frac{1}{n}$. Define a sequence of discrete time stocks and bond prices $(\bar{T}^n_t, \bar{S}^n_t)$ as follows. Let $B^n_k$ and $S^n_k$ be determined by the following stochastic difference equations:

$$S^n_{k+1} = S^n_k + b(S^n_k) \frac{1}{n} \sigma^1(S^n_k) \frac{\hat{\varepsilon}_k^1}{\sqrt{n}} + \cdots + \sigma^N(S^n_k) \frac{\hat{\varepsilon}_k^N}{\sqrt{n}}, \quad (5)$$

$$B^n_{k+1} = B^n_k \left( 1 + \frac{\tau(S^n_k)}{n} \right), \quad (6)$$

$$S^n_0 = S_0, \quad B^n_0 = 1,$$

where $\sigma^j$ is the $j$-th column of $\sigma$. Let $\bar{B}^n_t = B^n_{[nt]}$ and $\bar{S}^n_t = S^n_{[nt]}$, then $\bar{B}^n_t$ and $\bar{S}^n_t$ are well-defined for all $t \in [0, 1]$ and right-continuous with left limits. Moreover, they are piecewise constant and jump only at $t = \frac{k}{n}$. Equation (5) and (6) can be viewed as a finite difference approximation to the stochastic differential equation (2) and (3).

Since there are $N + 1$ assets traded and $N + 1$ possible uncertain states following each trading date, markets are dynamically complete. The unique Arrow-Debreu state price $\pi_{S^n_k}(\omega_k^j)$ at time $\frac{k}{n}$ and stock price $S^n_k$ for a security that pays off one unit of consumption good at time $\frac{k+1}{n}$ and state $\omega_k^j$ but nothing otherwise must satisfy the following relation:

$$\sum_{s=1}^{N+1} \pi_{S^n_k}(\omega_k^j) S^n_{k+1}(\omega_k^s) = S^n_k. \quad (7)$$

Solving this equation by substituting (5) into (7), we obtain that

$$\pi_{S^n_k}(\omega_k^j) = \frac{1}{N + 1} \left( 1 + \frac{\kappa(S^n_k)^\top}{\sqrt{n}} \hat{\varepsilon}_k(\omega_k^j) \right) \left( 1 + \frac{\tau(S^n_k)}{n} \right)^{-1}.$$ 

Note that $\pi_{S^n_k}$ is the one period Arrow-Debreu state price (from time $\frac{k}{n}$ to time $\frac{k+1}{n}$). To obtain the Arrow-Debreu state price at time 0 for a security that pays off one unit of consumption
good at time $t = \frac{k}{n}$, we need to multiply all the one period Arrow-Debreu state prices from period 1 to period $k$, where period $i$ is from time $\frac{i-1}{n}$ to time $\frac{i}{n}$. Let $\pi_k$ denote this state price, then $\pi_k^n = \pi_{S_{k-1}} \pi_{S_{k-2}} \cdots \pi_{S_0}$ for $k \geq 1$, and $\pi_0^n = 1$. Moreover, $\pi_k^n$ satisfies the following stochastic difference equation:

$$\pi_{k+1}^n = \frac{\pi_k^n}{N+1} (1 + \kappa(S_k^n)^T \frac{r(S_k^n)}{\sqrt{n}})^{-1}.\xi_k).$$

We assume that all $\pi$'s are non-negative for sufficiently large $n$. A sufficient condition for this to be true is that $\kappa$ is bounded.$^7$

To relate $\{\pi_k^n\}$ to the implicit Arrow-Debreu state price process $\{\xi_t\}$ defined in (4), we introduce a new variable $\xi_k^n$,

$$\xi_k^n = \pi_k^n B_k^n (N + 1)^k.$$

One can verify that $E_n[\xi_n] = 1$ by using the fact that $\sum_{t=1}^{N+1} \pi_{S_k}^n (\omega_k^n) = (1 + \frac{r(S_k^n)}{n})^{-1}$. Now, define

$$Q^n(A) = \int_A \xi_n^A P_n(d\omega),$$

then (7) implies that $Q^n$ is an equivalent martingale measure, i.e. the discounted stock price processes become martingales under this measure. We therefore call $\xi_k^n$ the implicit Arrow-Debreu state price (per unit of the probability) as we did for $\xi_t$ in the continuous time case. The implicit state price also can be represented by a stochastic difference equation as follows,

$$\xi_{k+1}^n = \xi_k^n \left(1 + \kappa(S_k^n)^T \frac{r(S_k^n)}{\sqrt{n}}\right).$$

(8)

For $t \in [0,1]$, set $\tilde{\xi}_t^n = \xi_{[nt]}^n$. Our first claim for this $N+1$-nomial approximation is the following convergence theorem on asset prices and the implicit Arrow-Debreu state prices.

**Theorem 1** Let $\tilde{X}^n = (\tilde{S}^n, \tilde{B}^n, \tilde{\xi}^n)$ and $X = (S, B, \xi)$. Then $\tilde{X}^n \to X$.

**Proof.** First, we argue that the uniform Lipschitz condition and the linear growth condition for $b$ and $\sigma$ guarantee that the weak solution for (2) and (3) is unique. As a result, the weak solution for (4) is also unique. Next, letting

$$L_t^n = \begin{pmatrix} \int_0^{[nt]} b(\tilde{S}_s^n) ds \\ \int_0^{[nt]} \sigma(\tilde{S}_s^n) \tilde{B}_s^n ds \\ 0 \end{pmatrix},$$

$^7$When this condition is not satisfied, we need to add a higher order term to the right hand side of (5), so that the resulting price system admits no arbitrage opportunity, while still permitting weak convergence for prices and Arrow-Debreu state prices.
$$A^n_t \equiv \begin{pmatrix}
\int_0^t \sigma(\tilde{S}^n_s)\sigma^\top(\tilde{S}^n_s)ds & 0 & -\int_0^t \sigma(\tilde{S}^n_s)\kappa(\tilde{S}^n_s)\xi^n_s ds \\
0 & 0 & 0 \\
-\int_0^t \kappa(\tilde{S}^n_s)^\top\sigma(\tilde{S}^n_s)\xi^n_s ds & 0 & \int_0^t \kappa(\tilde{S}^n_s)\xi^n_s^2 ds
\end{pmatrix},$$

then $A^n_t - A^n_s$ is non-negative definite for $t \geq s$, and $M^n \equiv X^n - L^n$, $M^n M^{n\top} - A^n$ are martingales. If we can verify that condition c) and d) of Proposition 1 are satisfied, we can conclude that $\tilde{X}^n$ converges to $X$ weakly.

To verify c), we observe that with the control of stopping time and the continuity of $r$, $b$, $\sigma$ and $\kappa$, $|X_t^n - X_t^m|$ is of the order $n^{-1/2}$, $|L_t^n - L_t^m|$ is of the order $n^{-1}$ and $|A^n_t - A^m_t|$ is of the order $n^{-1}$. Thus, c) is satisfied. Similar argument applies to d). This completes our proof. □

**Remark 1**

a) This proposition implies that the local movements of $N$ one-dimensional and independent Brownian motions can be approximated by $N$ uncorrelated but possibly dependent random variables $\{\varepsilon_k^j, j = 1, \ldots, N\}$.

b) $N+1$ is the minimum number of branches one can allow in order to keep $\{\varepsilon_k^j, j = 1, \ldots, N\}$ being uncorrelated among themselves.

To give a few examples, let us consider the Black-Scholes' price system with one stock and one bond ($N=1$),

$$B_t = e^{rt},$$

$$S_t = S_0e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma w(t)}.$$

If we choose $A$ to be

$$A = \begin{pmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}
\end{pmatrix},$$

we obtain the following binomial approximation:

$$S_{k+1}^n = \begin{cases}
S_k^n + \frac{\mu S_k^n}{\sqrt{n}} + \frac{\sigma S_k^n}{\sqrt{n}} \\
S_k^n + \frac{\mu S_k^n}{\sqrt{n}} - \frac{\sigma S_k^n}{\sqrt{n}}
\end{cases}$$

More generally, consider the following price system with two correlated lognormal processes,

$$dS_{t,1} = \mu_1 S_{t,1} dt + \sigma_1 S_{t,1} dw_1,$$

$$dS_{t,2} = \mu_2 S_{t,2} dt + \rho \sigma_2 S_{t,2} dw_1 + \sigma_2 \sqrt{1 - \rho^2} S_{t,2} dw_2,$$
where $\sigma_i^2$ is the volatility of the return on the $i$-th asset, and $\rho$ is the correlation coefficient of the returns on these two assets. If we choose $A$ to be

$$A = \begin{pmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
0 & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}}
\end{pmatrix},$$

we obtain the following trinomial approximation:

$$S_{k+1,1}^n = \frac{\mu_1 S_{k,1}^n}{n} + \sigma_1 S_{k,1}^n \frac{\sqrt{3}}{\sqrt{2n}}$$

$$S_{k,1}^n = \frac{\mu_1 S_{k,1}^n}{n} - \sigma_1 S_{k,1}^n \frac{\sqrt{3}}{\sqrt{2n}}$$

$$S_{k,2}^n = \frac{\mu_2 S_{k,2}^n}{n} + \sigma_1 S_{k,2}^n \frac{\sqrt{3}}{\sqrt{2n}} + \sigma_2 \sqrt{1 - \rho^2} S_{k,2}^n \frac{1}{\sqrt{2n}}$$

$$S_{k,2}^n = \frac{\mu_2 S_{k,2}^n}{n} - \sigma_1 S_{k,2}^n \frac{\sqrt{3}}{\sqrt{2n}} + \sigma_2 \sqrt{1 - \rho^2} S_{k,2}^n \frac{1}{\sqrt{2n}}$$

3. Contingent Claim Pricing

An immediate application of Theorem 1 is the convergence of contingent claim prices. We shall demonstrate in this section that the contingent claim price process obtained from the discrete time model based on a no arbitrage argument converges weakly to its continuous time counterpart. Moreover, the dynamic portfolio strategy that replicates the payoff of the underlying contingent claim converges weakly. It then follows that the contingent claim price and the replicating portfolio strategy computed at time 0 converge to the corresponding continuous time limits.

We begin with a definition for contingent claims. Let $g$ be a measurable function mapping from $\mathbb{R}^N$ to $\mathbb{R}$. A contingent claim on stock prices $S$ is defined to be a security which pays $g(S_t)$ units of consumption good at final date (i.e. $t = 1$).8 This formulation subsumes all the usual applications. For example, it includes option on one stock, $\max(S_1^{(1)} - K, 0)$, and option on the maximum of the two stocks, $\max(\max(S_1^{(1)}, S_1^{(2)}) - K, 0)$, as special cases, where $S_1^{(i)}$ denotes the price of the $i$-th security at time 1.

---

8We ignore the possibility of dividend payout before the final date, although it poses no difficulty at all.
Following Harrison and Kreps [1979], the price of a contingent claim at time $t$ can be evaluated by taking the conditional expectation of the discounted final payoff under the equivalent martingale measure $Q$, i.e.

$$
V(S_t, t) = E_Q \left[ e^{-\int_t^T r(S_t, \tau) d\tau} g(S_T) \mid S_t \right]
$$

$$
= E \left[ \frac{\xi_1}{\xi_t} e^{-\int_t^T r(S_t, \tau) d\tau} g(S_T) \mid S_t \right].
$$

(9)

Alternatively, if $V$ is first order continuously differentiable with respect to $t$ and second order continuously differentiable with respect to $z$, then $V$ can be determined as a solution to the following partial differential equation (see Cox, Ingersoll and Ross [1985]):

$$
\frac{1}{2} \text{trace} [\sigma^T V_{SS} \sigma] + rS^T V_S + V_t - rV = 0,
$$

(10)

$$
V(S, 1) = g(S),
$$

where

$$
V_S(S, t) = \left[ \frac{\partial V(S, t)}{\partial S_1}, \ldots, \frac{\partial V(S, t)}{\partial S_N} \right]^T,
$$

$$
V_{SS}(S, t) = \left[ \frac{\partial^2 V(S, t)}{\partial S_i \partial S_j} \right]_{N \times N}.
$$

Equation (10) is usually called fundamental partial differential equation for valuation in the option pricing literature. The dynamic portfolio strategy which replicates the final payoff of this claim is given by

$$
\theta(S_t, t) = V_S(S_t, t),
$$

$$
\alpha(S_t, t) = (V(S_t, t) - \theta(S_t, t)^T S_t) / B_t,
$$

where $\theta_i$ and $\alpha_i$ denote the number of shares held in the $i$-th stock and bond respectively.

This valuation technique can be easily applied to the discrete time model. The price of the contingent claim at any time $k/n$, denoted by $V^n$, can be evaluated by taking the conditional expectation under the equivalent martingale measure $Q^n$, i.e.

$$
V^n(S^n_k, \frac{k}{n}) = E^n \left[ g(S^n_T) B^n_k / B^n_n \mid S^n_k \right]
$$

$$
= E^n \left[ \frac{\xi^n}{\xi_k} g(S^n_T) B^n_k / B^n_n \mid S^n_k \right].
$$

(11)

Alternatively, $V^n$ can be determined as a solution to the following recurrent equation:

$$
V^n(S^n_k, \frac{k}{n}) = \sum_{s=1}^{N+1} \mathbb{I}_{S^n_k(\omega^n_s)} V^n(S^n_{k+1}(\omega^n_s), \frac{k+1}{n}),
$$

(12)

$$
V^n(S^n_n, 1) = g(S^n_n).
$$
The dynamic portfolio strategy that replicates the final payoff of this claim is determined by the following system of linear equations:

\[ \alpha^n_k B^n_{k+1,1}(\omega^k_s) + \theta^n_{k,1} S^n_{k+1,1}(\omega^k_s) + \cdots + \theta^n_{k,N} S^n_{k+1,N}(\omega^k_s) = V(S^n_{k+1}(\omega^k_s), \frac{k+1}{n}), \]

for \( s = 1, \cdots, N + 1 \). Since there are \( N + 1 \) equations and \( N + 1 \) unknowns, the solution for \( (\alpha^n_k, \theta^n_k) \) is uniquely determined. Moreover, \( \alpha^n_k \) and \( \theta^n_k \) are functions of stock prices \( S^n_k \) and time \( \frac{k}{n} \). The following theorem establishes a convergence of contingent claim prices from discrete time models to the continuous time diffusion model. We need a definition to begin with.

A function \( f : \mathbb{R}^N \rightarrow \mathbb{R}^M \) is said to satisfy a polynomial growth condition if there exist \( K_1, \beta > 0 \) such that

\[ |f(z)| \leq K_1(1 + |z|^\beta), \quad \forall z \in \mathbb{R}^N. \]

**Theorem 2** (Contingent Claim Pricing) Suppose the following:

(a) \( r(z) \) and \( \sigma(z)z \) satisfy a uniform Lipschitz condition for \( z \in \mathbb{R}^N \). \( r(z) \) and \( \sigma(z) \) are continuously differentiable w.r.t. to \( z \) up to the six order with all these derivatives satisfying a polynomial growth condition;

(b) \( g(z) \) is piecewise and continuously differentiable up to the sixth order with all these derivatives satisfying a polynomial growth condition.

Then,

1. Let \( V_t = V(S_t, t), V^n_k = V^n(S^n_k, \frac{k}{n}) \) and \( \tilde{V}_t^n = V^n_{[nd]} \). We have \( \tilde{V}_t^n \Rightarrow V_t \);
2. Let \( \alpha_t = \alpha(S_t, t), \theta_t = \theta(S_t, t), \tilde{\alpha}_t^n = \alpha^n_{[nd]} \) and \( \tilde{\theta}_t^n = \theta^n_{[nd]} \). We have \( (\tilde{\alpha}_t^n, \tilde{\theta}_t^n) \Rightarrow (\alpha_t, \theta_t) \).

In particular, \( \tilde{V}_0^n \rightarrow V_0 \), and \( (\tilde{\alpha}_0^n, \tilde{\theta}_0^n) \rightarrow (\alpha_0, \theta_0) \) as \( n \rightarrow \infty \).

**Proof.** The proof for this theorem is rather lengthy. We leave it to Appendix B.1.

**Remark 2** a) All the conditions imposed on the drift term and the diffusion term are readily satisfied for a price system with stock prices following a multi-dimensional geometric Brownian motion and bond price paying a constant interest rate.

b) Piecewise differentiability assumption on \( g(z) \) allows contingent claims to have truncated final payoffs, such as call options.

In contrast to solving a PDE such as (10) with possibly more than one state variable, this theorem provides us with a simple numerical procedure for computing the price of a contingent claim and the replicating portfolio strategy. Specifically, at \( t = 0 \), \( \tilde{V}_0^n \), \( \tilde{\alpha}_0^n \) and \( \tilde{\theta}_0^n \) can be used as the numerical approximations to the corresponding continuous time limits.
5. Optimal Consumption-Portfolio Policies

In this section we study convergence of optimal consumption-portfolio policies. We focus on the following three economic variables: the optimal consumption policy, the optimal portfolio policy, and the optimal conditional expected utility. Our objective is to show that all these variables converge weakly to their continuous time counterparts. While it is natural to require convergence of optimal expected utilities in the context of consumption-portfolio selection, convergence of optimal conditional expected utilities is also important to ask for. This is because at each point in time, the objective of an investor is always to maximize the conditional expected utility of consumption conditioning upon the past resolution of asset prices. Thus, convergence of optimal conditional expected utilities is as important as convergence of optimal expected utilities.

We first consider convergence of optimal consumption policies. According to Cox and Huang [1987], a typical dynamic consumption and portfolio problem for a single agent can be formulated in terms of the following static expected utility maximization problem:

$$\sup_{c,W \geq 0} E \left[ \int_0^T u(c_t, t) dt + V(W) \right]$$

subject to

$$E \left[ \int_0^T \xi c_t / B_t dt + \xi_1 W / B_1 \right] \leq W_0,$$

where $u$ and $V$ are utility functions, assumed to be continuous, increasing and concave, $\{c_t, t \in [0, 1]\}$ is the consumption rate process, and $W$ is the final wealth, representing a possible bequest to the next generation. The decision variable $c$ is assumed to be adapted to the filtration $\mathcal{F} = \{\mathcal{F}_t : t \in [0, 1]\}$, and $W$ is measurable with respect to $\mathcal{F}_1$. Equation (13) is called the budget constraint, which says that the value at time 0 of the consumption process $\{c_t\}$ and the final wealth $W$ is equal to the initial wealth $W_0$.

Under this static setup, one first solves for the optimal consumption policy, and then implements it by some dynamic portfolio strategy. A nice feature of this formulation is that the optimal consumption policy can be expressed explicitly as a function of the implicit Arrow-Debreu state prices. Specifically, if we define

$$\hat{f}(y, t) = \inf \{z \geq 0 : u'_x(z, t) \leq y\},$$

$$\hat{g}(y) = \inf \{z \geq 0 : V'_x(z) \leq y\},$$

Traditionally, dynamic consumption and portfolio problem is solved by using stochastic dynamic programming (see Merton [1971]). However, Cox and Huang [1987]'s static approach turns out to be much easier to work with for our purpose.
then applying the Lagrangian theory, we can find a $\lambda_0 > 0$ such that
\[ c_t = \hat{f}(\lambda_0 \xi_t / B_t, t), \quad W = \hat{g}(\lambda_0 \xi_1 / B_1), \]
where $\lambda_0$ is determined through the following budget constraint:
\[ E \left[ \int_0^1 \hat{f}(\lambda_0 \xi_t / B_t, t) \xi_t / B_t dt + \hat{g}(\lambda_0 \xi_1 / B_1) \xi_1 / B_1 \right] = W_0. \tag{14} \]

We assume that $\lambda_0$ is uniquely determined.\(^{10}\) Moreover, if we introduce a new state variable $Z_t = B_t / \lambda_0 \xi_t$, then $(S, Z)$ forms a Markov process, and the optimal wealth process is a function of $Z_t, S_t$, and $t$. In fact, let $F$ denote this function, we have that
\[ F(S_t, Z_t, t) = B_t E_Q \left[ \int_t^1 \hat{f}(Z_s^{-1}, s) / B_s ds + \hat{g}(Z_1^{-1}) / B_1 | S_t, Z_t \right] \]
\[ = Z_t E \left[ \int_t^1 Z_s^{-1} \hat{f}(Z_s^{-1}, s) / B_s ds + Z_1^{-1} \hat{g}(Z_1^{-1}) / B_1 | S_t, Z_t \right], \]
which satisfies the following linear partial differential equation:
\[ \frac{1}{2} \text{trace}[\sigma^T F_{SS} \sigma] + \frac{1}{2} |Z\kappa|^2 F_{ZZ} + Z\kappa^T \sigma^T F_S + \tau S^T F_S + \tau Z F_Z + F_t - \tau F + \hat{f}(Z^{-1}, t) = 0, \tag{15} \]
\[ F(S, Z, 1) = \hat{g}(Z^{-1}), \tag{16} \]
where $\sigma = \sigma(S)$, $\kappa = \kappa(S)$ and $\tau = \tau(S)$. Furthermore, the dynamic portfolio strategy that produces the optimal consumption policy is determined by
\[ \theta_t = F_S + (\sigma^T \sigma_t)^{-1}(b_t - \tau S)Z F_Z, \tag{17} \]
\[ \alpha_t = (F - \theta_t^T S_t) / B_t. \]

We refer the readers to Cox and Huang [1987] as well as Essay One for details.

This treatment can be easily applied to the discrete time model. Analogously, we formulate a discrete time static consumption problem as follows:
\[ \max_{c_t^0 \geq 0} E_n \left[ \sum_{k=0}^{n-1} u(c_k^n, \xi_k^n) \Delta t_k^n + V(W^n) \right] \]
\[ \text{s.t. } E_n \left[ \sum_{k=0}^{n-1} \xi_k^n c_k^n / B_k^n \Delta t_k^n + \xi_n^n W^n / B_n^n \right] = W_0, \tag{18} \]
where $t_k^n = \frac{k}{n}$ and $\Delta t_k^n = \frac{1}{n}$. Applying the Lagrangian theory, we can find a $\lambda_n > 0$ such that
\[ c_k^n = \hat{f}(\lambda_n \xi_k^n / B_k^n, c_k^n), \quad W^n = \hat{g}(\lambda_n \xi_n^n / B_n^n), \]
\(^{10}\) A sufficient condition for this to be true is that $u$ and $V$ are strictly concave.

58
for \( k = 0, 1, \cdots, n - 1 \), where \( \lambda_n \) is determined by the budget constraint (18). Therefore, the optimal consumption policy for the discrete time model also can be expressed explicitly as a function of the implicit Arrow-Debreu state prices. Now, define \( \tilde{c}_t^n = c_{[nt]}^n \) for \( t \in [0, 1] \). Since \( \bar{f} \) is continuous except for at most countably many points, following Lemma 1 and Theorem 1, \((\tilde{c}^n, W^n)\) converges to \((c, W)\) weakly if and only if \( \lambda_n \to \lambda_0 \). The fact that \( \lambda_n \to \lambda_0 \) is proved in the following theorem. Without loss of generality, we assume that \( V = 0 \) and \( \tau(t) = 0 \).

**Theorem 3 (Consumption Policies)** Suppose that \( \int_0^1 \tilde{c}_t^n \bar{f}(\gamma \tilde{c}_t^n, [nt]) \, dt \) is uniformly integrable for any constant \( \gamma > 0 \), then \( \tilde{c}^n \to c \), as \( n \to \infty \).

**Proof.** We demonstrate \( \lambda_n \to \lambda_0 \) by contradiction. Suppose \( \lambda_n \to \lambda_0 \) is not true, then there exists a subsequence of \( \lambda_{nj} \), which converges to some number \( \bar{\lambda} \neq \lambda_0 \) (\( \bar{\lambda} \) could be zero and infinite, but is always non-negative). Let us first assume that \( \bar{\lambda} \) is finite and positive. According to Lemma 1, \( \tilde{c}^n \Rightarrow \tilde{c} \), where \( \tilde{c}_t = \bar{f}(\bar{\lambda} \xi_t, t) \). Hence, \( \int_0^1 \xi_t^n \tilde{c}_t^n \, dt \) converges to \( \int_0^1 \xi_t \bar{f}(\bar{\lambda} \xi_t, t) \, dt \) in distribution.

If \( \bar{\lambda} > \lambda_0 \), then there exists sufficiently small number \( \epsilon \) such that \( \lambda_{nj} > \bar{\lambda} - \epsilon > \lambda_0 \) for \( j > j_0 \). Hence, \( \bar{f}(\lambda_{nj} \xi_t^n, [nt]/n) \leq \bar{f}((\bar{\lambda} - \epsilon) \xi_t^n, [nt]/n) \) for \( j > j_0 \), where we have used the fact that \( \bar{f}(x, t) \) is decreasing in \( x \). Since \( \int_0^1 \xi_t^n \tilde{c}_t^n \, dt \) is uniformly integrable, we have

\[
E_n \left[ \int_0^1 \tilde{c}_t^n \bar{f}(\lambda_{nj} \xi_t^n, [nt]/n) \, dt \right] \leq E_n \left[ \int_0^1 \tilde{c}_t^n \bar{f}((\bar{\lambda} - \epsilon) \xi_t^n, [nt]/n) \, dt \right] \to E \left[ \int_0^1 \xi_t \bar{f}((\bar{\lambda} - \epsilon) \xi_t, t) \, dt \right],
\]

(19)

where the convergence on the right-hand-side follows from Lemma 4. Since the left-hand side of (19) is always \( W_0 \) according to (18), we obtain that \( E \left[ \int_0^1 \xi_t \bar{f}((\bar{\lambda} - \epsilon) \xi_t, t) \, dt \right] \geq W_0 \). But, this contradicts (14).

If \( \bar{\lambda} < \lambda_0 \), similar technique can lead to a contradiction. We omit the proof for the case when \( \bar{\lambda} = 0 \), or \( \infty \).  

We would like to remind the readers that the uniform integrability condition required by Theorem 3 is not difficult to check. For example, if \( \kappa \) is bounded and \( \bar{f}(z^{-1}, t) \) satisfies a polynomial growth condition, i.e. \( \exists K, \beta > 0, \) such that

\[
\bar{f}(z, t) \leq K(1 + z^{-\beta}), \quad \forall z > 0,
\]

then the uniform integrability condition is satisfied.\(^{11}\) To demonstrate this, we need only to show that for some \( \eta > 0 \),

\[
\sup_{n} E_n \left[ \left( \frac{1}{n} \sum_{k=0}^{n-1} (\xi_k^n + (\xi_k^n)^{1-\beta}) \right)^{1+\eta} \right] < \infty,
\]

\(^{11}\) A specific example is that \( u(z) = \frac{z^{1-b}}{1-b} \), where \( b > 0 \).
see Stroock [1987]. In fact, by convexity, we have

\[
E_n \left[ \frac{1}{n} \sum_{k=0}^{n-1} \left( \xi_k^n + (\xi_k^n)^{-\beta} \right)^{1+\eta} \right] \leq E_n \left[ \frac{1}{n} \sum_{k=0}^{n-1} \left( \xi_k^n + (\xi_k^n)^{1-\beta} \right)^{1+\eta} \right] \\
\leq 2^{1+\eta} E_n \left[ \frac{1}{n} \sum_{k=0}^{n-1} \left( (\xi_k^n)^{1+\eta} + (\xi_k^n)^{(1-\beta)(1+\eta)} \right) \right].
\]

Applying the technique used in the proof of Lemma 3 in Appendix B.1, one can show that for all \( \alpha \in \mathbb{R} \),

\[
E_n[\xi_k^n]^\alpha \leq C,
\]

where \( C \) depends only on \( \alpha \) and the bound of \( |\kappa| \), but not on \( k, n \). We conclude that

\[
\sup_n E_n \left[ \frac{1}{n} \sum_{k=0}^{n-1} \xi_k^n + (\xi_k^n)^{-\beta} \right]^{1+\eta} < \infty,
\]

for any \( \eta > 0 \). Therefore, when \( \kappa \) is bounded and \( \tilde{f}(x^{-1}, t) \) satisfies a polynomial growth condition, we have \( \tilde{\xi}^n \Rightarrow c \).

Theorem 3 basically says that if the implicit Arrow-Debreu state price process converges to the continuous time limit, then, subject to some regularity conditions on the utility function and the parameters of price processes, the optimal consumption process always converges to the corresponding continuous time limit. However, this theorem does not tell us anything as to whether the optimal portfolio policy converges or not.

It turns out that convergence of optimal portfolio policies can be established in the same way we did for the replicating portfolio strategies in contingent claim pricing. In the previous case, we established convergence of replicating portfolio strategies by comparing the solution to the PDE with the solution to the recurrent equation that determines the price of the claim.

This motivates us to compare the discrete time portfolio strategy with the continuous time portfolio strategy by comparing the optimal wealth processes and their partial derivatives. For notational convenience, we assume \( N = 1 \) (\( N > 1 \) is analogous).

Let \( F \) be the optimal wealth process that satisfies (15) and (16), and \( \theta \) be the optimal portfolio strategy defined by (17). Correspondingly, we define \( F^n \) to be the discrete time optimal wealth process, and \( \theta^n \) be the discrete time optimal portfolio strategy that replicates the optimal consumption process. As in the continuous time case, we introduce an additional state variable \( \tilde{Z}_n \) by setting \( Z_k^n = \lambda_k B_k^n / \xi_k^n \) and \( \tilde{Z}_n = Z^n_{[nd]} \). Then, it should be obvious that the value process \( F^n \) is a function of \( (S_k^n, Z_k^n, t_k^n) \) (see (20) below). Under these notations, \( F^n \) must satisfy the following recurrent equation,

\[
F^n(S_k^n, Z_k^n, t_k^n) = \frac{1}{n} f(Z_k^{n-1}, t_k) + \pi S_k^n(+)F^n(S_{k+1}^{n+}, Z_{k+1}^{n+}, t_{k+1}^n) + \pi S_k^n(-)F^n(S_{k+1}^{n-}, Z_{k+1}^{n-}, t_{k+1}^n),
\]
\[ F^n(S^n, Z^n, t^n) = \hat{g}(Z^{n-1}), \quad (20) \]

where \( + \) and \( - \) denote the states \( \hat{e}_k = 1 \) and \( \hat{e}_k = -1 \) respectively, and

\[
S^n_{k+1}^+ = S^n_k + \frac{b(S^n_k)}{n} + \frac{\sigma(S^n_k)}{\sqrt{n}}, \\
S^n_{k+1}^- = S^n_k + \frac{b(S^n_k)}{n} - \frac{\sigma(S^n_k)}{\sqrt{n}}, \\
Z^n_{k+1}^+ = Z^n_k(1 + \frac{\tau(S^n_k)}{n})(1 + \kappa(S^n_k)\sqrt{\frac{1}{n}})^{-1}, \\
Z^n_{k+1}^- = Z^n_k(1 + \frac{\tau(S^n_k)}{n})(1 - \kappa(S^n_k)\sqrt{\frac{1}{n}})^{-1}.
\]

Equation (20) says that the sum of the values of the current consumption and future wealth is equal to the current wealth. Given the wealth process, the optimal portfolio policy strategy that replicates the optimal consumption policy is determined by

\[
\theta^n_k = \frac{F^n(S^n_{k+1}^+, Z^n_{k+1}^+, t^n_{k+1}) - F^n(S^n_{k+1}^-, Z^n_{k+1}^-, t^n_{k+1})}{S^n_{k+1}^+ - S^n_{k+1}^-}, \\
\alpha^n_k = \left(F^n(S^n_k, Z^n_k, t^n_k) - \theta^n_k S^n_k\right)/B^n_k.
\]

We want to show that \((\tilde{\alpha}^n, \tilde{\theta}^n)\) converges to \((\alpha, \theta)\). To do that, we apply a similar approximation technique used in the proof of Theorem 2. Let \(e^n_k\) be the truncation errors of the optimal wealth process defined as,

\[ e^n_k = F(S^n_k, Z^n_k, t^n_k) - F^n(S^n_k, Z^n_k, t^n_k). \]

Substituting this into (21), we get

\[
\theta^n_k = \frac{\sqrt{n}}{2\sigma(S^n_k)}(F(S^n_{k+1}^+, Z^n_{k+1}^+, t^n_{k+1}) - F(S^n_{k+1}^-, Z^n_{k+1}^-, t^n_{k+1})) + \frac{\sqrt{n}}{2\sigma(S^n_k)}(e^n_{k+1}^+ - e^n_{k+1}^-).
\]

The following theorem shows that

(a) \( \frac{\sqrt{n}}{2\sigma(S^n_k)}(F(S^n_{k+1}^+, Z^n_{k+1}^+, t^n_{k+1}) - F(S^n_{k+1}^-, Z^n_{k+1}^-, t^n_{k+1})) \Rightarrow \theta, \)

(b) \( \frac{\sqrt{n}}{2\sigma(S^n_k)}(e^n_{k+1}^+ - e^n_{k+1}^-) \Rightarrow 0. \)

**Theorem 4 (Portfolio Policies)** Suppose the following

- a) \( \tilde{Z}^n \) converges to \( Z \) weakly;
- b) \( \sigma, \tau \) and \( \kappa \) are continuously differentiable up to the sixth order with all these derivatives satisfying a polynomial growth condition;
- c) \( r(x)x, r(x)y \) and \( \kappa(x)y \) satisfy a uniform Lipschitz condition and a linear growth condition;
d) \( \hat{f}(z^{-1}, t) \) and \( \hat{g}(z^{-1}) \) satisfy a polynomial growth condition.

Then \((\tilde{\alpha}^n, \tilde{\theta}^n) \Rightarrow (\alpha, \theta)\). In particular, \(\tilde{\alpha}_0^n \to \alpha_0\) and \(\tilde{\theta}_0^n \to \theta_0\) as \(n \to \infty\).

**Proof.** The proof for this theorem is similar to that of Theorem 2, we delegate the proof to Appendix B.1. ■

**Remark 3** The uniform Lipschitz condition imposed on \(r(x) y\) and \(\kappa(x) y\) implies that both \(r\) and \(\kappa\) are bounded. This condition can be weakened and replaced by a uniform Lipschitz condition on \(r, \kappa\) and \(|\kappa|^2\), provided that \(\hat{f}(e^{-x}, t)\) and \(\hat{g}(e^{-x})\) satisfy a polynomial growth condition in \(x\).

Similar to Theorem 2, we obtain a simpler numerical procedure for computing the optimal consumption and the optimal portfolio policy as compared to the standard method of solving a PDE such as (15) with at least two state variables, namely \(S\) and \(Z\).

Our last investigation is on convergence of optimal conditional expected utilities. Let \(\tilde{X}^n = (\tilde{S}^n, \tilde{Z}^n)\) and \(X = (S, Z)\) denote the state variables which characterize the information sets of the discrete and continuous time economies, respectively. Let \(\mathbb{F}^n\) and \(\mathbb{F}\) be the filtrations generated by \(\tilde{S}^n\) and \(\tilde{S}\), respectively. We have the following convergence theorem on optimal conditional expected utilities.

**Theorem 5** (Conditional Expected Utilities) Suppose that the sufficient conditions for Theorem 3 are satisfied. Suppose further that \(\{\int_0^1 u(\tilde{c}^n_t, [nt/n])dt\}\) is uniformly integrable, then

\[
E_n \left[ \int_0^1 u(\tilde{c}^n_t, [nt/n]) dt \vbar \mathbb{F}^n_t \right] \Rightarrow E \left[ \int_0^1 u(c_t, t) dt \vbar \mathbb{F}_t \right].
\]

Moreover, for any \(A \in \mathbb{F}_t\) such that the boundary of \(A\) has zero probability under \(P\), then

\[
E_n \left[ \int_0^1 u(\tilde{c}^n_t, [nt/n]) dt \vbar X_n \in A \right] \to E \left[ \int_0^1 u(c_t, t) dt \vbar X \in A \right], \quad \text{as } n \to \infty.
\]

**Proof.** First, we claim that according to Theorem 3, \(\tilde{X}^n \Rightarrow X\) and \(\tilde{c}^n \Rightarrow c\). To prove (22), we need to utilize a notion of extended weak convergence, which is developed by Aldous [1981]. This convergence notion allows us to show that (22) follows directly from the fact that \(\tilde{c}^n\) converges to \(c\) extended weakly. Since the exact definition of extended weak convergence is rather technical, we refer the readers to Appendix B.2 for a formal definition as well as some of the basic properties for this type of convergence. It follows from Remark 4, Lemma 7, and Proposition 2 in Appendix B.2 that

\[
E_n \left[ \int_0^1 u(\tilde{c}^n_t, [nt/n]) dt \vbar \mathbb{F}^n_t \right] \Rightarrow E \left[ \int_0^1 u(c_t, t) dt \vbar \mathbb{F}_t \right].
\]
To prove (23), let \( Y_n \equiv \int_0^1 u(\tilde{c}_t^n, \frac{[nt]}{n}) dt \), then \( |Y_n| \) is uniformly integrable. Let \( \tilde{Y}_n \equiv |Y_n| \).

Following Lemma 2, \( E_n[\tilde{Y}_n|\mathcal{F}_t^n] \) is uniformly integrable. Hence, \( E_n[Y_n|\mathcal{F}_t^n] \) is also uniformly integrable. Next, consider

\[
E_n \left[ \int_0^1 u(\tilde{c}_t^n, \frac{[nt]}{n}) dt | X^n \in A \right] = \int_{X^n \in A} E_n[Y_n|\mathcal{F}_t^n] dP_n
\]

\[
= \int 1_{X^n \in A} E_n[Y_n|\mathcal{F}_t^n] dP_n.
\]

Since the boundary of \( A \) has zero probability under \( P \) and \( X^n \) converges to \( X \) weakly, we conclude that the integrand \( 1_{X^n \in A} E_n[Y_n|\mathcal{F}_t^n] \) converges to \( 1_{X \in A} E[\int_0^1 u(c_t, t) dt | \mathcal{F}_t] \) weakly. Furthermore, it is easy to check that \( 1_{X^n \in A} E_n[Y_n|\mathcal{F}_t^n] \) is uniformly integrable, given that \( E_n[Y_n|\mathcal{F}_t^n] \) is uniformly integrable. Hence, the limit goes through the integration, which gives (23).

This concludes our demonstration of the convergence of optimal consumption-portfolio policies and optimal conditional expected utilities. To summarize our results, the convergence theorems established in Sections 3, 4 and 5 confirm our economic intuition that discrete time models and continuous time models are approximations to each other. In cases where closed form solutions exist, the continuous time approach provides a convenient analytic tool for studying contingent claim prices and optimal consumption-portfolio policies. However, in cases where closed form solutions do not exist, the discrete time approach provides an elegant numerical solution to the continuous time model.

6. Concluding Remarks

The leading assumption for the convergence results established in this essay is that markets in continuous time models are dynamically complete. Dynamic completeness is necessary for contingent claims to be priced by arbitrage, but it is not necessary for solving dynamic consumption and portfolio problem. We conjecture that similar convergence results would also be obtained for optimal consumption-portfolio policies in which markets are dynamically incomplete. In this case, it is reasonable to expect that markets in discrete time models would also be dynamically incomplete. The exact implementation is subject to future research.

References


Appendix B.1. Proofs

Proof for Theorem 2. We need two lemmas to proceed our proof.

Lemma 3. For any integer $m, l, k \geq 0$ and $k \geq l$, there exists a constant $A > 0$, independent of $m$ and $K$, such that

$$\hat{E}_n[S_k^n]^{2m} \leq A(1 + \hat{E}_n[S_k^n]^{2m})$$

where $\hat{E}_n$ denotes the expectation under $Q_n$.

Proof. We demonstrate this inequality for $N = 1$. The proof follows closely from Friedman [1975], Theorem 2.3 on p.107, where he derives this result for the diffusion process.

Apparently, we can find some $K' > 0$, depending only upon $K$, such that for any $z \in \mathbb{R}$,

$$|b(z)|^2 \leq K'(1 + |z|^2), \quad |\sigma(z)|^2 \leq K'(1 + |z|^2),$$

$$|z b(z)| \leq K'(1 + |z|^2), \quad |b(z)\sigma(z)| \leq K'(1 + |z|^2),$$

$$|z^2 \tau(z)| \leq K'(1 + |z|^2).$$

Applying Taylor expansion to function $x^{2m}$, we get

$$|S_{k+1}^n|^{2m} = |S_k^n|^{2m} + 2m|S_k^n|^{2m-1}(S_k^n - S_{k+1}^n) + m(2m - 1)|S_k^n|^{2m-2}(S_k^n - S_{k+1}^n)^2$$

$$= |S_k^n|^{2m} + 2m|S_k^n|^{2m-1} \left( \frac{b}{n} + \frac{\sigma}{\sqrt{n}} \frac{\epsilon_k}{n} \right) + m(2m - 1)|S_k^n|^{2m-2} \left( \frac{b^2}{n^2} + \frac{2b\sigma}{n^{3/2}} \epsilon_k + \frac{\sigma^2}{n} \epsilon_k^2 \right)$$

where $\tilde{S}_k^n = S_k^n + \beta \left( \frac{b}{n} + \frac{\sigma}{\sqrt{n}} \frac{\epsilon_k}{n} \right)$ for some $\beta \in [0,1]$. Invoking expectation $\hat{E}_n$ on both sides and noticing that $|\tilde{S}_k^n| \leq |S_k^n| + |b| + |\sigma|$, we get

$$\hat{E}_n[S_{k+1}^n]^{2m} \leq \hat{E}_n[S_k^n]^{2m} + \frac{2mK'}{n} \hat{E}_n(|S_k^n|^{2m-2} + |S_k^n|^{2m})$$

$$\leq \hat{E}_n[S_k^n]^{2m} + \frac{2mK'}{n} \hat{E}_n(|S_k^n| + |b| + |\sigma|)^{2m-2}(2b^2 + 2|b\sigma| + \sigma^2),$$

where we have used the fact that $|z^{2m-2}r| \leq K'(1 + |z|^2)$ and $\hat{E}_n[\epsilon_k] = \frac{\kappa(S^n_k)}{\sqrt{n}}$. Notice further that $x^{2m-2} \leq 1 + z^{2m}$ and $(z + z')^m \leq 2^m(x^m + y^m)$ for $x > 0$ and $y > 0$, we obtain

$$\hat{E}_n[S_{k+1}^n]^{2m} \leq \hat{E}_n[S_k^n]^{2m} + \frac{2mK'}{n} \hat{E}_n[1 + 2|S_k^n|^{2m}]$$

$$+ \frac{4K'm(2m - 1)}{n} \hat{E}_n[(2K' + (1 + 2K')|S_k^n|^{2m-2}(1 + |S_k^n|^{2m})]$$

$$\leq \hat{E}_n[S_k^n]^{2m} + \frac{2mK'}{n} \hat{E}_n[1 + 2|S_k^n|^{2m}]$$

$$+ \frac{4K'm(2m - 1)}{n} 2^{2m-2} \left( (2K')^{2m-2}(1 + \hat{E}_n[S_k^n]^{2m}) + (1 + 2K')^{2m-2} \hat{E}_n[|S_k^n|^{2m-2} + |S_k^n|^{2m}] \right)$$

However, $\hat{E}_n[S_k^n]^{2m} \leq 1 + \hat{E}_n[S_k^n]^{2m}$ for $m \geq 1$, and $\hat{E}_n[S_k^n]^{2m-2} \leq 1 + \hat{E}_n[S_k^n]^{2m}$, there must exist a constant $C > 0$, depending on $m, K'$ such that

$$\hat{E}_n[S_{k+1}^n]^{2m} \leq C_n + (1 + \frac{C}{n})\hat{E}_n[S_k^n]^{2m}.$$
This implies that
\[ \hat{E}_n|S_k^n| \leq (1 + \frac{C}{n})^{k-1}(1 + \hat{E}_n|S_k^n|^{2m}). \]

Then, \( A = \sup_n(1 + \frac{C}{n})^n \) will be our choice. 

\[ \square \]

**Lemma 4** Under assumption (a) and (b), all the partial derivatives of \( V \) up to the third order satisfy a polynomial growth condition.

**Proof.** First, we rewrite \( V \) as
\[ V(x, t) = E_Q[e^{-Y_t}g(S_1)], \]
where \((\hat{Y}_r, \hat{S}_r)\) satisfies the following equation:
\[
\begin{align*}
    d\hat{Y}_r &= r(\hat{S}_r)dt \\
    d\hat{S}_r &= r(\hat{S}_r)\hat{S}_rdr + \sigma(\hat{S}_r)d\omega^r \\
    \hat{Y}_t &= 0, \quad \hat{S}_t = x,
\end{align*}
\]
where \( r \geq t, \hat{Y} \geq 0 \) and \( \omega^r = \omega_r - \int_0^t \kappa(s)ds \) is a Brownian motion under \( Q \) by Girsanov theorem, see Friedman [1975].

Following the argument used in Theorem 5.5 of Friedman [1975], pp.122 and Theorem 1 of Gihman and Skorohod [1972], pp.61, we conclude that \( V \) is continuously differentiable with respect to \( x \) up to the sixth order, and all of these derivatives satisfy a polynomial growth condition. Moreover, since \( V \) satisfies the PDE, we deduce that \( V \) is also continuously differentiable with respect to \( t \) up to the third order, since \( V_t, V_{tt} \) and \( V_{ttt} \) can be expressed as a function of partial derivatives of \( V \) w.r.t. \( x \) up to at most the sixth order. This plus the fact that \( b, \sigma \) and \( r \) and all their derivatives satisfy a polynomial growth condition lead us to further deduce that all the derivatives with respect to \( x, t \) up to the third order satisfy a polynomial growth condition. 

\[ \square \]

**Proof of Theorem 2.** We prove this theorem for \( N = 1 \) (\( N > 1 \) is analogous to \( N = 1 \)).

The basic idea of this proof is to substitute the true function \( V \) into the recurrent equation that defines \( V^n \) so as to get better estimates for the truncation errors, defined as
\[ e_k^n = V(S^n_k, \frac{k}{n}) - V^n(S^n_k, \frac{k}{n}). \]

We use + and − to denote the state \( \bar{\epsilon}_k = 1 \) and \( \bar{\epsilon}_k = -1 \) respectively, and define
\[
\begin{align*}
    S_{k+1}^{n+} &= S^n_k + \frac{b(S^n_k)}{n} + \frac{\sigma(S^n_k)}{\sqrt{n}}, \\
    S_{k+1}^{n-} &= S^n_k + \frac{b(S^n_k)}{n} - \frac{\sigma(S^n_k)}{\sqrt{n}},
\end{align*}
\]
where \( S_{k+1}^{n+} \) and \( S_{k+1}^{n-} \) denote the price at time \( t_{k+1}^n \) when \( \bar{\epsilon}_k^n = 1 \) and \( \bar{\epsilon}_k^n = -1 \) respectively. Correspondingly, we define two functions, \( f_{+,k}^n \) and \( f_{-,k}^n \) as follows, setting \( t_{k+1}^n = \frac{k}{n} \):
\[
    f_{+,k}^n(\tau) = V(S^n_k + \tau(S^n_{k+1} - S^n_k), t_k^n + \tau(t_{k+1}^n - t_k^n)),
\]

68
\[
f^{k,n}(\tau) = V(S_k^n + \tau(S_{k+1}^n - S_k^n), t_k^n + \tau(t_{k+1}^n - t_k^n)).
\]

Since \( V \) is third order differentiable, \( f^{k,n}_+ \) and \( f^{k,n}_- \) admit the following Taylor expansions (we omit the superscripts for \( f \)):
\[
\begin{align*}
f_+(1) &= f_+(0) + f'_+(0) + \frac{1}{2} f''_+(0) + R^n_k, \\
f_-(1) &= f_-(0) + f'_-(0) + \frac{1}{2} f''_-(0) + Q^n_k,
\end{align*}
\]
where
\[
R^n_k = \frac{1}{6} \int_0^1 (1 - s)^3 f^{(3)}_+(s) ds,
\]
\[
Q^n_k = \frac{1}{6} \int_0^1 (1 - s)^3 f^{(3)}_-(s) ds.
\]

More explicitly, we have,
\[
\begin{align*}
f'_+(0) &= V_S(S_{k+1}^{n+} - S_k^n) + \frac{1}{n} V_t, \\
f'_-(0) &= V_S(S_k^{n-} - S_k^n) + \frac{1}{n} V_t, \\
f''_+(0) &= V_{SS}(S_{k+1}^{n+} - S_k^n)^2 + V_t \frac{1}{n^2} + 2V_{St} \frac{1}{n} (S_{k+1}^{n+} - S_k^n), \\
f''_-(0) &= V_{SS}(S_k^{n-} - S_k^n)^2 + V_t \frac{1}{n^2} + 2V_{St} \frac{1}{n} (S_k^{n-} - S_k^n),
\end{align*}
\]
where \( V \) and the partial derivatives of \( V \) are evaluated at \( (S_k^n, t_k^n) \). Now, to get an estimate for \( e^n_k \), we substitute the expressions of \( f^{k,n}_+(1) \) and \( f^{k,n}_-(1) \) into the recurrent equation and use the fact that \( V \) satisfies (10), which yields, after rearranging terms,
\[
\pi S_k^n (+) f_+(1) + \pi S_k^n (-) f_-(1) = V(S_k^n) + \frac{1}{2} V_{SS} \left( \frac{b^2}{n^2} + \frac{2b\sigma \kappa}{n^3} + \frac{\tau S_k^n}{n^2} \right) (1 + \frac{\tau}{n})^{-1} \frac{1}{2} V_t \frac{1}{n^2} (1 + \frac{\tau}{n})^{-1} - \gamma^n_k (1 + \frac{\tau}{n})^{-1},
\]
where \( \gamma^n_k = -(\pi S_k^n (+) R_k^n + \pi S_k^n (-) Q_k^n) \). Let \( h = -\frac{1}{2} V_{SS} (b^2 + 2b\sigma \kappa + \tau S) - \frac{1}{2} V_t \). Then we have the following recurrent equation for \( e^n_k \):
\[
e^n_k = \pi S_k^n (+) e^{n+}_{k+1} + \pi S_k^n (-) e^{n-}_{k+1} + \frac{1}{n^2} h(S_k^n, \frac{k}{n}) (1 + \frac{\tau}{n})^{-1} + \gamma^n_k (1 + \frac{\tau}{n})^{-1}.
\]

This equation allows us to argue that
\[
|e^n_k| \leq \frac{C}{n} (1 + |S_k^n|)^{2q}, \tag{24}
\]
for some constants \( C > 0 \) and integer \( q > 0 \), which are independent of \( k, n \). To see that, we express \( e^n_k \) as follows:
\[
\begin{align*}
e^n_0 &= 0, \\
e^n_k &= \frac{1}{n^2} \hat{E}_n \left[ \sum_{m=k}^{n-1} h(S_m^n, \frac{m}{n}) B_m^n/B_m^n + n^2 \gamma^n_m B_m^n/B_m^n \mid S_k^n \right]. \tag{25}
\end{align*}
\]
By Lemma 2, \( h \) satisfies a polynomial growth condition, thus there exists a constant \( C' > 0 \) and integer \( q > 0 \) such that
\[
|h(x, t)| \leq C'(1 + |x|^{2q}).
\]

Applying Lemma 3, we get
\[
\hat{E}_n \left[ h(S^n, \frac{m}{m}) \mid S^n \right] \leq C'(1 + A(1 + |S^n|^{2q}))
\]
for \( m \geq k \). This takes care of the first term inside the summation on the right hand side of equation (25). For the last term, we argue that the summation is also of the order \( \frac{1}{n} \). To see that, we can express \( R^n_k \) and \( Q^n_k \) explicitly by writing down \( f'_+^{(3)}(s) \) and \( f'_-^{(3)}(s) \), which allows us to argue that they are the sum of higher order terms, of the order \( n^{-2} \) or higher. For example, one typical term of \( R^n_k \) has the following form:
\[
\int_0^1 (1-s)^3 V_{SSS}((1-s)S^n_k + sS^n_{k+1}, \ldots)(S^n_{k+1} - S^n_k)^3 ds
\]
\[
= \int_0^1 (1-s)^3 V_{SSS}((1-s)S^n_k + sS^n_{k+1}, \ldots) \frac{b^3(S^n_k)}{n^3} ds + \ldots
\]
Since \( V_{SSS} \) satisfies a polynomial growth condition, so is the integrand. We can choose \( q \) to be sufficiently large so that all the polynomial growth conditions have the same power \( q \). Now applying the same procedure, we can get an inequality for \( r^{(n)}_m \), which is of the order \( n^{-2} \), hence the summation is of the order \( n^{-1} \). Altogether, these imply that (24) is true. Thus \( \bar{V}^n \Rightarrow V \) weakly.

We now substitute \( f_- \) and \( f_+ \) into the equation that defines \( \theta^n_k \). This yields
\[
\theta^n_k = \frac{f_+^{(n)}(1) - f_-^{(n)}(1)}{S^n_{k+1} - S^n_{k+1}}
\]
\[
= V_S + \frac{b}{n} V_{SS} + \frac{V_{St}}{n} + \frac{V_{St}}{n} + \frac{2\sqrt{n}}{n} (R^n_k - Q^n_k).
\]
(26)

Since \( R^n_k \) and \( Q^n_k \) are of the order \( n^{-1} \) or higher, we conclude that
\[
\frac{b}{n} V_{SS} + \frac{V_{St}}{n} + \frac{V_{St}}{n} + \frac{2\sqrt{n}}{n} (R^n_k - Q^n_k) \Rightarrow 0.
\]

Hence, \( \theta^n_k \Rightarrow V_S(S, t) \). Finally, since \( \bar{V}^n \Rightarrow V \) and \( \bar{V}_t^n = \tilde{\alpha}^n_t \tilde{B}_t^n + \tilde{\theta}_t^n \tilde{S}_t^n \), we conclude that \( \tilde{\alpha}^n \Rightarrow \alpha \). This proves part 2). The claim that \( \bar{V}_0^n \Rightarrow V_0 \), \( \tilde{\alpha}_0^n \Rightarrow \alpha_0 \) and \( \tilde{\theta}_0^n \Rightarrow \theta_0 \) follow directly from the fact that they are non-stochastic.

\[
\text{PROOF FOR THEOREM 4. Notice, we can rewrite } F \text{ as follows,}
\]
\[
F(x, y, t) = E \left[ \int_0^1 e^{-\bar{\gamma}_t} f(Z^{-1}, \tau) d\tau + e^{-\bar{\gamma}_t} \bar{g}(Z^{-1}) \right],
\]

70
where \((\tilde{B}_r, \tilde{Z}_r, \tilde{S}_r)\) satisfies the following differential equation:

\[
\begin{align*}
\frac{d\tilde{Y}_r}{d\tau} &= r(\tilde{S}_r)d\tau \\
\frac{d\tilde{S}_r}{d\tau} &= r(\tilde{S}_r)\tilde{S}_r d\tau + \sigma(\tilde{S}_r)dw^*_r \\
\frac{d\tilde{Z}_r}{d\tau} &= r(\tilde{S}_r)Z_r d\tau - Z_r \kappa(\tilde{S}_r)dw^*_r \\
\tilde{Y}_t &= 0, \quad \tilde{S}_t = x, \quad Z_t = y, \quad \tau \in [t, 1],
\end{align*}
\]

and \(w^*_r = w_r - \int_0^r \kappa_s ds\). A similar argument as used in the proof of theorem 2 leads us to conclude that conditions a), b) and c) imply that \(F\) is continuously differentiable up to the third order, and all these derivatives satisfy a polynomial growth condition.

The basic idea of this proof is exactly the same as in the proof for Theorem 2. We substitute the true function \(F\) into the finite difference equation that defines \(F^n\) so as to get better estimates for the truncation errors. We define two functions, \(f^{k,n}_+\) and \(f^{k,n}_-\) as follows:

\[
\begin{align*}
f^{k,n}_+(\tau) &= F(S^+_k + \tau(S^n_{k+1} - S^n_k), Z^+_k + \tau(Z^n_{k+1} - Z^n_k), t^n_k + \tau(t^n_{k+1} - t^n_k)), \\
f^{k,n}_-(\tau) &= F(S^-_k + \tau(S^n_{k+1} - S^n_k), Z^-_k + \tau(Z^n_{k+1} - Z^n_k), t^n_k + \tau(t^n_{k+1} - t^n_k)).
\end{align*}
\]

Then \(f^{k,n}_+\) and \(f^{k,n}_-\) admit the following Taylor expansions (we omit the superscripts for \(f\)):

\[
\begin{align*}
f_+(1) &= f_+(0) + f'_+(0) + \frac{1}{2} f''_+(0) + R^n_k, \\
f_-(1) &= f_-(0) + f'_-(0) + \frac{1}{2} f''_-(0) + Q^n_k,
\end{align*}
\]

where

\[
\begin{align*}
R^n_k &= \frac{1}{6} \int_0^1 (1 - s)^3 f^{(3)}_+(s)ds, \\
Q^n_k &= \frac{1}{6} \int_0^1 (1 - s)^3 f^{(3)}_-(s)ds.
\end{align*}
\]

We need to show that \(\frac{\sqrt{n}}{2\pi(S^n_k)}(f^{k,n}_+(1) - f^{k,n}_-(1)) \Rightarrow \theta\). If we express \(f^{k,n}_+\) and \(f^{k,n}_-\) explicitly, and substitute them into the right hand side of (a), we would obtain, after rearranging terms,

\[
(a) = F_S(S^+_k, Z^+_k, t^n_k) - \frac{1}{\sigma(S^n_k)} F_Z(S^n_k, Z^n_k, t^n_k) \kappa(S^n_k) Z^n_k + \varepsilon^n_k,
\]

where \(\varepsilon^n_k\) is the sum of several terms, all of them are of the order \(n^{-1}\) or higher, which goes to zero.

To show (b), we need to use the fact that \(F\) satisfies equation (15). Expressing \(\pi_S^+(+)f_+(1) + \pi_S^-(+)f_-(1)\) in terms of \(f_+(0)\) and \(f_-(0)\) etc., we get,

\[
\pi_S^n(+)F(S^+_{k+1}, Z^+_{k+1}, t^n_{k+1}) + \pi_S^n(-)F(S^-_{k+1}, Z^-_{k+1}, t^n_{k+1}) + \frac{1}{n(1 + \varepsilon^n_k)} \hat{f}(Z^n_k) = F(S^n_k, Z^n_k, t^n_k) + O(\frac{1}{n^2}).
\]

This plus the budget equation (20) should give us the following relation for the truncation errors:

\[
\varepsilon^n_k = \pi_S^n(+)\varepsilon^n_{k+1} + \pi_S^n(-)\varepsilon^n_{k+1} + O(\frac{1}{n^2}),
\]

71
with \( e^n = 0 \), where \( O(\frac{1}{n^p}) \) denotes the sum of those terms which are of the order \( \frac{1}{n^p} \) or higher. As in Theorem 2, this allows us to conclude that

\[
|e^n_k| = O(\frac{1}{n}).
\]

Thus, \( \tilde{F}^n \Rightarrow F \) and (b) holds. This further implies that \( \tilde{\alpha}^n \Rightarrow \alpha \).

### B.2. Extended Weak Convergence

This appendix introduces the notion of extended weak convergence developed by Aldous [1981] and derives some of the basic properties for this type of convergence.

We start by introducing the notions of regular conditional distribution and prediction process, which are needed for defining extended weak convergence. For ease of notation, we will use \( \Omega \) rather than \( D^M[0,1] \) to denote the sample space. An element of \( \Omega \) will be denoted generically as \( \omega \).

**Definition 1** Let \( X \) be a stochastic process in \( D^M[0,1] \), \( B_0 \) be a sub-sigma field in \( D \). A function \( \xi(\omega, A) : \Omega \times D \rightarrow \mathbb{R} \) is called a regular conditional distribution of \( X \) given \( B_0 \) if

(a) for each \( \omega \), the map \( A \rightarrow \xi(\omega, A) \) is a probability measure on \( \Omega \);

(b) for each \( A \), the random variable \( \omega \rightarrow \xi(\omega, A) \) is a version of \( E(1_{\{X \in A\}} | B_0) \), where \( E \) denotes the expectation under \( P \) induced by \( X \).

The existence of a regular conditional distribution is guaranteed by the Doob-Kuratowski theorem when \( B_0 \) is countably generated (see Williams [1979] p.100). Based on (b), we can simply write \( \xi(\omega, A) \) as \( E(1_{\{X \in A\}} | B_0) \).

**Definition 2** \( \{Z_t\}_{0 \leq t \leq 1} \) is said to be a prediction process of \( X \) if for every \( t \), \( Z_t \) is a regular conditional distribution of \( X \) given \( \mathcal{F}_t \), where \( \mathcal{F}_t = \sigma(X_s, s \leq t) \), the sigma-field generated by \( \{X_s, s \leq t\} \).

The prediction process always exists, since \( D \), therefore, \( \mathcal{F}_t \) is countably generated. An important mathematical fact is that the set of regular conditional distributions \( \{Z_t : t \in [0,1]\} \) can be put together to form a RCLL stochastic process from \( \Omega \) to \( \Pi \), where \( \Pi \) denotes the space of all possible probability measures on \( D^M[0,1] \) endowed with weak topology (see Theorem 13.1, Aldous [1981]). The space of RCLL stochastic process mapping from \( \Omega \) to \( \Pi \) will be denoted as \( D(\Pi) \), which is endowed with the Skorohod topology induced by the weak topology on \( \Pi \).

If we take the underlying process \( X \) to be asset prices, then the prediction process of \( X \) completely describes both the distributions and the conditional distributions of the asset returns given the past resolutions of the prices.

We now give the definition of extended weak convergence.

**Definition 3** Let \((X^n, F^n)\) and \((X, F)\) be in \( D^M[0,1] \) with prediction processes \( Z^n \) and \( Z \), respectively, where \( F^n = \{\mathcal{F}_t^n ; t \in [0,1]\} \), \( F^n = \sigma(X^n_t ; s \leq t) \) and \( F = \{\mathcal{F}_t ; t \in [0,1]\} \), \( F_t = \sigma(X_t ; s \leq t) \). Then \( X^n \) converges to \( X \) extended weakly, written as \((X^n, F^n) \Rightarrow (X, F)\), if \( (X^n, Z^n) \) converges weakly to \((X, Z)\) in \( D^M[0,1] \times D(\Pi) \), the natural product space of \( D^M[0,1] \) and \( D(\Pi) \).

When the price processes converge extended weakly, the distributions and the conditional distributions of the price processes converge weakly to the corresponding limits. Hence, extended weak convergence is stronger than weak convergence insofar as the conditional probabilities structures are concerned. Unfortunately, Definition 3 is difficult to interpret and work with. The following lemmas will help us understand about extended weak convergence more intuitively.
Lemma 5. Let \((X^n, F^n) \Rightarrow (X, F)\). Let \(t\) be a continuity point of the prediction process \(Z\) of \(X\) (continuous with respect to the weak topology), and let \(\phi : D^M[0, 1] \rightarrow R\) be a continuous function such that \(\{\phi(X^n)\}\) is uniformly integrable. Then \(E_n[\phi(X^n)|F^n_t] \Rightarrow E[\phi(X)|F_t]\).

PROOF. See Lemma 16.2, Section 16 of Aldous [1981].

Lemma 6. Let \(X\) be a diffusion process, and let \(Z\) be the prediction process of \(X\). Then \(Z_t\) is continuous in \(t\) with respect to the weak topology for every \(t > 0\).

PROOF. We want to show that for any bounded continuous function \(\phi\) mapping from \(D^M[0, 1]\) to \(R\), \(\int \phi(X)Z_t(\omega)\) is continuous in \(t\). However,

\[
\int_{\Omega} \phi(X)Z_t(\omega) = E[\phi(X)|F_t].
\]

The fact that the right-hand side of this equation is continuous in \(t\) is recorded in Huang [1985].

Lemmas 5 and 6 have the following important implication. For any bounded continuous function \(\phi\), mapping from \(D^M[0, 1]\) to \(R\), we have \(E_n[\phi(X^n)|F^n_t] \Rightarrow E[\phi(X)|F_t]\) provided that \(Z\) is continuous at \(t\). Following Theorem 2.1 of Billingsley [1968], we conclude that \(P_n(X^n \in B|F^n_t) \Rightarrow P(X \in B|F_t)\) for all measurable sets \(B\) so that \(P(X \in \partial B|F_t) = 0\). In words, this is to say that when agents update their posterior probabilities or beliefs for any such event \(B\), the posterior probabilities (which are random variables too) converge in distribution to the limiting posterior probabilities. In contrast, weak convergence only entails the convergence of ex ante (unconditional) probabilities, i.e. \(P_n(X^n \in B) \rightarrow P(X \in B)\). The following proposition gives a necessary and sufficient condition for extended weak convergence (see Section 16 of Aldous [1981]).

Proposition 2. \((X^n, F^n) \Rightarrow (X, F)\) if and only if for all bounded continuous functions \(h_1, \ldots, h_k\) on \(D^M[0, 1],\)

\[
(X^n_t, E_n[h_1(X^n)|F^n_t], \ldots, E_n[h_k(X^n)|F^n_t])_{t \geq 0} \Rightarrow (X_t, E[h_1(X)|F_t], \ldots, E[h_k(X)|F_t])_{t \geq 0}.
\]

In general, it is still very difficult to establish extended weak convergence via Proposition 1. However, when \(X\) is a diffusion process, the martingale central limit theorem introduced in Section 2 for weak convergence also provides easily verifiable sufficient conditions for establishing extended weak convergence.

Proposition 3. Let \(X^n\) and \(X\) be defined as in Proposition 1, then \(X^n\) converges to \(X\) extended weakly.

PROOF. We know from Proposition 1 that the Markov process \(X^n\) converges to \(X\) weakly in \(D^M[0, 1]\). Then, following Theorem 21.2 of Aldous [1981], we conclude that \(X^n\) converges to \(X\) extended weakly.

Remark 4. 1) The markovian assumption on \(X^n\) is crucial for proving convergence in prediction processes. This is because, with the markovian structure, the regular conditional distribution depends only on the current value of \(X^n\).

2) Given this proposition, it is clear that \(\bar{X}^n\) of Theorem 1 converges to \(X\) extended weakly. Hence, \(\bar{X}^n\) of Theorem 3 converges to \(X\) extended weakly.

Finally, the continuous mapping theorem introduced in Section 2 also can be generalized as follows.
Lemma 7 (Continuous Mapping Theorem) Let $(X^n, F^n) \Rightarrow (X, F)$.

i) Suppose that $h(x)$ is a continuous function mapping from $\mathbb{R}^N$ to $\mathbb{R}^M$, except for at most countably many points, and denote $Y^n_t = h(X^n_t)$, then $(Y^n, F^n) \Rightarrow (Y, F)$.

ii) Suppose that $H$ is a continuous mapping from $D^M[0, 1]$ to $D^N[0, 1]$, such that for $f, g \in D^M[0, 1]$ and $t \in [0, 1]$, $f(u) = g(u)$ for $u \leq t$ implies $H(f)(u) = H(g)(u)$ for $u \leq t$. Then, $(H(X^n), F^n) \Rightarrow (H(X), F)$.

ESSAY THREE:
METHOD OF MOMENTS ESTIMATION FOR
DIFFUSION PROCESSES

Abstract This essay develops a generalized method of moments (GMM) estimation procedure for estimating parameters of general diffusion processes using discretely sampled data. When the functional forms of the conditional moments are known, GMM estimation is straightforward. When the functional form of the conditional moments are unknown, a numerical approximation procedure is introduced, and GMM estimation is then applied by replacing the actual conditional moments by their numerical approximations. Numerical approximations for conditional moments are obtained through approximating the continuous time diffusion process by a sequence of discrete time multinomial processes. The conditional moments computed from the discrete time processes are shown to converge to the actual conditional moments of the limiting diffusion process. Large sample properties are investigated for the GMM estimators obtained by this method.

1. Introduction

With the theoretical success of dynamic capital asset pricing models, parametric estimation for diffusion asset price process is now becoming an important empirical subject in financial economics. Standard estimation procedures for diffusion process often require that the transition density function of the underlying process be analytically tractable, so that a maximum likelihood (ML) estimation can be applied to a series of discretely sampled observations, cf. Marsh and Rosenfeld [1983], Lo [1986] and Gibbons and Jacklin [1988]. However, except for a limited class of diffusion processes, it is in general difficult to obtain a closed form representation of the transition density for an arbitrarily given diffusion process. Although the transition density function is determined by the Fokker-Plank partial differential equation (PDE), solving the PDE numerically is clearly a formidable task, especially when the underlying state variables are of large dimension.

Alternatively, Gibbons and Jacklin [1988] use generalized method of moments (GMM) estimation to estimate the parameters of a CEV diffusion process using discretely sampled data for stock prices.¹ They obtain moment conditions by deriving analytical expressions of conditional moments. An important feature of the generalized method of moments estimation is that it allows one to perform a specification test through testing over-identifying restrictions. GMM estimation suffers from the difficulty of finding the exact analytical expressions for the moment functions when the diffusion process is arbitrarily given. However, since moment functions are expectations of functionals of the underlying diffusion process, they can be approximated by various numerical and statistical procedures, so that GMM could be applied without knowing

¹The constant elasticity of variance (CEV) diffusions were originally introduced in Cox [1975].
the functional forms of moment functions. An example of such is Duffie and Singleton [1989], where they have used Monte-Carlo simulation technique to approximate the moment functions.

In this essay, we develop a relatively simple numerical approximation procedure that allows us to approximate the conditional moments for a general diffusion process. The basic idea of our approach is similar to the binomial option pricing approach used by Cox, Ross and Rubinstein [1979], where they approximate the Black-Scholes' option pricing formula by the binomial option prices obtained from the discrete time processes. We approximate the continuous time diffusion process by a sequence of discrete time binomial or multinomial processes, such that along the discrete time sequence the conditional moments can be computed easily and they converge to the actual conditional moments of the limiting diffusion process. Our approximation procedure is simple even when the diffusion process is multi-dimensional. This approach, therefore, allows us to apply the generalized method of moments estimation to a series of discretely sampled observations without knowing the functional form of the conditional moments.

The idea that a continuous time process can be approximated by discrete time processes has been explored by many others in various estimation contexts, see for example, Melino and Turnbull [1988] and Scott [1987]. In those studies, the stochastic differential equation that governs the asset price process is discretized and a conditional normal distribution is obtained for the sample observations. In order to ensure the resulting estimator to be consistent, it is always assumed that the underlying asset price process follows exactly the distribution of the discretized process, i.e. the normal distribution. Strictly speaking, the parameter estimators derived by this method are inconsistent if the underlying asset price process does follow the diffusion process (see Lo [1986]). Our approach differs from the previous ones in that we always assume that the underlying process follows a diffusion process, and our approximation for the continuous time process is used only for the purpose of finding numerical estimates for the conditional moments.

A closely related work on estimating the parameters of general diffusion processes using GMM is the recent paper by Duffie and Singleton [1989]. Duffie and Singleton develop a simulated moment estimation technique, which approximates the steady state (unconditional) moments through simulating the sample paths of the underlying process. Their approach is different from ours in that they focus on approximations for the steady state moment functions, rather than the conditional moments as in this essay. Although theoretically, the simulated moments estimation could be applied to approximate the conditional moments as well, it is practically intractable, since the amount of simulations may be overwhelming. On the other hand, our discrete time approximation approach is not appropriate for finding the steady state
moment function, since we do not know the initial distribution of the diffusion process. In addition, the simulated moments approach has the advantage of estimating a diffusion process with unobservable state variables. In our studies, we have to require that all the state variables appearing in the diffusion equations be observable.

For the purpose of developing large sample properties for the GMM estimators, we require that the diffusion process be stationary and ergodic. Stationarity and ergodicity ensure that a strong law of large numbers and a central limit theorem obtain even though the sample observations are serially dependent. In statistical literature, there exist many other sufficient conditions to ensure these large sample properties to hold. For instance, the stationarity and ergodicity condition can be replaced by a mixing condition. For stationary sequence, the mixing condition implies ergodicity. The mixing condition is always satisfied when the sample observations are generated from a finite Gaussian ARMA process. More generally, Galant and White [1988] use a notion of near epoch dependence, which allows the serial observations to depend upon the entire history of another time series while still permitting the law of large number and the central limit theorem to hold. In practice, ergodicity and mixing condition are difficult to verify. Alternatively, Duffie and Singleton [1989] introduce a notion of geometric ergodicity, which ensures that when the diffusion process is simulated with some arbitrary starting point, the law of large numbers and the central limit theorem are applicable. Intuitively speaking, geometric ergodicity guarantees the sample observations to be asymptotically stationary and ergodic. There exist easily verifiable conditions for a process to be geometrically ergodic, although these conditions may be too strong to apply in practice. In principle, any of the above mentioned conditions is sufficient for the law of large numbers and the central limit theorem to obtain. Without loss of generality, we adopt the stationarity and ergodicity assumption, which would then allow us to quote many existing theorems from Hansen [1982] for establishing consistency and asymptotic normality for the GMM estimators.

As an illustration, GMM estimation is applied to a mean reverting square root process originally used by Cox, Ingersoll and Ross [1985] for modeling the dynamics of spot interest rate. We conduct Monte-Carlo simulations, which examine the finite sample properties of the GMM estimators, in particular, the properties of asymptotic normality.

Finally, it is worth pointing out that many other statistical studies have focused on estimating diffusion process using continuous records, cf. Brown and Hewitt [1975], Le Breton [1976], Kutoyants [1976], and Tsirovish [1976]. These studies allow the unknown parameters to appear only in the drift term, which rules out the situations we are interested in.

The balance of this essay is organized as follows. Section 2 sets out the estimation pro-
cEDURE under GMM. Section 3 presents numerical approximations of the conditional moments for general diffusion process. In Section 4, we develop large sample properties for the GMM estimators introduced in Section 2. Section 5 reports the empirical results of GMM estimation applying to a mean reverting square root process. Simulations are conducted to examine the finite sample properties of the GMM estimators.

2. GMM Estimation

For simplicity, only univariate diffusion processes are considered here although the results can be extended readily to the general multi-dimensional case.

Let \( X(t) \in \mathbb{R} : t \in [0, \infty) \) be a stochastic process defined on a complete probability space \((\Omega, P, \mathcal{F})\), and suppose that \( X(t) \) satisfies the following stochastic differential equation:

\[
\begin{align*}
    dX(t) &= b(X(t), \beta)dt + \sigma(X(t), \beta)dw(t), \\
    X(0) &= Y,
\end{align*}
\]

where \( w(t) \) is a one dimensional standard Brownian motion defined on \((\Omega, P, \mathcal{F})\), \( \beta \) is a \( r \)-dimensional vector of unknown parameters to be estimated, and \( \beta \) lies in some compact parameter space \( \Theta \subset \mathbb{R}^r \). The true parameter, denoted by \( \beta_0 \), is assumed to lie in the interior of \( \Theta \). Functions, \( b(z, \beta) : \mathbb{R} \times \mathbb{R}^r \to \mathbb{R} \) and \( \sigma(z, \beta) : \mathbb{R} \times \mathbb{R}^r \to \mathbb{R} \), are assumed to be continuous in both \( z \) and \( \beta \), and satisfy the usual linear growth condition and uniform Lipschitz condition. That is, there exist constants \( K, L > 0 \) such that for all \( z, \gamma \in \mathbb{R} \) and \( \beta \in \Theta \),

\[
|b(z, \beta)| + |\sigma(z, \beta)| \leq K(1 + |z|),
\]

\[
|b(z, \beta) - b(\gamma, \beta)| + |\sigma(z, \beta) - \sigma(\gamma, \beta)| \leq L|z - \gamma|.
\]

Note that \( K \) and \( L \) are independent of \( \beta \), which is reasonable to assume, given that \( \Theta \) is compact. Moreover, the drift term \( b \) and the diffusion term \( \sigma \) are independent of \( t \), therefore \( X \) is time-homogenous. The time-homogenous assumption is necessary for diffusion processes to be stationary, which we will assume later. Finally, the starting point \( Y \) is assume to be a random variable defined on \((\Omega, P, \mathcal{F})\).

A generalized method of moments estimation for diffusion process can be conducted as follows. First, a series of observations for \( X \), denoted by \( X(\tau_t) \) or \( X_t \), is sampled at dates \( \{\tau_t\} \) for \( t = 1, 2, \ldots, T \). Second, for each \( t \), an \( I \)-dimensional vector valued measurable function of \( X_t, X_{t-1}, \ldots, X_{t-k} \) and the unknown parameter \( \beta \), denoted by \( f(X_t, \ldots, X_{t-k}; \beta) \) ( \( f_t(X, \beta) \) or
For simplicity, is selected such that \( f_t^{\beta} \) satisfies the following moment condition under the null hypothesis that \( \beta = \beta_0 \),
\[
E[f_t(X, \beta_0)] = 0.
\] (5)

This type of moment conditions is often called the orthogonality condition in various equilibrium asset pricing models, where the orthogonality condition is derived from the first order condition for optimality. Usually, \( I \) must be greater than or equal to \( r \) in order for all the unknown parameters to be identifiable. Lastly, for any parameter vector \( \beta \in \Theta \), let
\[
G_T(\beta) \equiv \frac{1}{T} \sum_{t=1}^{T} f_t(X, \beta)
\]
denote a vector of sample moments. If \( f_t^{\beta} \) satisfies a strong or weak law of large numbers, then \( \text{plim}_T G_T(\beta_0) = 0 \). With some identifiability condition, we have \( \text{plim}_T G_T(\beta) = 0 \) if and only if \( \beta = \beta_0 \). Thus, the GMM estimator is defined to be the solution to the following non-linear minimization program:
\[
\hat{\beta}_T \equiv \arg \min_{\beta \in \Theta} G_T(\beta)^T W_T \quad G_T(\beta),
\] (6)

where \( W_T \) is a symmetric, semi-positive, and \( I \times I \) weighting matrix, measuring the length of the sample moments \( G_T(\beta) \). Under suitable regularity conditions, \( \hat{\beta}_T \) is consistent and asymptotically normal.

In general, the moment function \( f_t^{\beta} \) can be formed in two different ways. The first one, as adopted by Duffie and Singleton [1989], uses an unconditional expectation relation, which sets
\[
f_t(X, \beta) = g_t(X, \beta) - E^{\beta}[g_t(X, \beta)],
\] (7)

where \( g_t = g(X_t, X_{t-1}, \cdots, X_{t-k}, \beta) \) is a known measurable function. The expectation operator \( E^{\beta} \) denotes the expectation taken as if the true parameter is \( \beta \). The second one, as adopted by Gibbons and Jacklin [1988], uses a conditional expectation relation, which sets
\[
f_t(X, \beta) = g_t(X, \beta) - E^{\beta}[g_t(X, \beta)|X_{t-l}],
\] (8)

for some integer \( l > 0 \). In both cases, the moment conditions (5) for \( f \) are satisfied.

One sees immediately that a complete knowledge of the functional form of \( E^{\beta}[g_t(X, \beta)] \) or \( E^{\beta}[g_t(X, \beta)|X_{t-l}] \) is essential for GMM estimation to apply. Gibbons and Jacklin [1988] derive the exact analytical expressions of the conditional mean and variance for the CEV diffusion, and apply GMM to a series of weekly and monthly observed stock prices, while assuming stock prices to follow a CEV diffusion process. However, in general, it is very difficult to calculate
conditional moments for arbitrary diffusion process. It is even more difficult to calculate the unconditional moments, since we may not know the initial distribution of \(Y\).\(^2\)

In this essay we focus on GMM estimation using the conditional moment condition such as (8). When the functional form of the conditional moments are known, GMM estimation is straightforward. When the functional form of the conditional moments are unknown, the numerical approximation procedure to be developed in the next section can be applied to approximate the conditional moments for general diffusion process. The basic idea of our approach is to approximate \(X\) by a sequence of binomial process (or multinomial processes if \(X\) is multi-dimensional), and replace the actual conditional moments by the conditional moments calculated from the binomial process. We will demonstrate that conditional moments obtained from the binomial processes converge to the true conditional moments of the limiting diffusion process uniformly.

Once the conditional moments are approximated, the GMM estimator can then be obtained as follows:

\[
G_T^n(\beta) \equiv \frac{1}{T} \sum_{t=1}^{T} f_t^n(X, \beta),
\]

\[
\hat{\beta}_T^n \equiv \arg \max_{\beta \in \Theta} G_T^n(\beta) W_T^n G_T^n(\beta), \tag{9}
\]

where \(f_t^n(X, \beta)\) is an approximation for \(f_t(X, \beta)\) and \(W_T^n\) is an approximation for the weighting matrix \(W_T\). If \(G_T^n(\beta)\) and \(W_T^n\) converge in probability to \(\operatorname{plim}_T G_T(\beta)\) and \(\operatorname{plim}_T W_T\) uniformly on \(\Theta\) as \((T, n) \to \infty\), respectively, then it is reasonable to expect that \(\hat{\beta}_T^n\) has the same asymptotic distribution as \(\hat{\beta}_T\).

3. Numerical Approximations for Conditional Moments

We now describe how the conditional moments in (8) can be approximated by conditional moments calculated from a sequence of discrete time processes, where the sequence of discrete time processes approximates the diffusion process in distribution. Our objective can be stated more generally as to develop a numerical procedure to approximate expectations of functionals of the solution of a stochastic differential equation. Applications of such procedure have appeared in contingent claims pricing literature, cf. Cox, Ross and Rubinstein [1979].

\(^2\)Duffie and Singleton [1989] use a Monte-carlo simulation technique to approximate the unconditional moments. More specifically, they set

\[
E^\beta[g_t(X, \beta)] \approx \frac{1}{T} \sum_{t=1}^{T} g_t(\tilde{X}, \beta),
\]

where \(\tilde{X}\) is the simulation of a sample path of the solution to the stochastic differential equation (1).
For ease of notation, we assume that the time horizon is \([0, 1]\), and the conditional moment of interest has the following form:

\[
F(\beta) = E^\beta[g(X(1), \beta)|X(0)],
\]

where \(X(0)\) is fixed.\(^3\) Function \(g\), mapping from \(\mathbb{R} \times \Theta\) to \(\mathbb{R}\), is measurable and satisfies a polynomial growth condition. That is, there exist constants \(K', \gamma > 0\), such that for all \(x \in \mathbb{R}\) and \(\beta \in \Theta\),

\[
|g(x, \beta)| \leq K'(1 + |x|^\gamma).
\]

(10)

We approximate the one dimensional diffusion process \(X\) by a sequence of binomial processes as follows. Divide \([0, 1]\) into \(n\) equally spaced subintervals, each with a length of \(\frac{1}{n}\), and define a binomial process \(X^n_k\) for \(k = 0, 1, \ldots, n\) as the solution to the following stochastic difference equation:

\[
\begin{align*}
X^n_{k+1} &= X^n_k + b(X^n_k, \beta) \frac{1}{n} + \sigma(X^n_k, \beta) \frac{\xi_k}{\sqrt{n}}, \\
X^n_0 &= X(0),
\end{align*}
\]

(11)

where \(P[\xi_k = 1] = P[\xi_k = -1] = \frac{1}{2}\).

We call equation (11) a discretization to the stochastic differential equation (1).\(^4\) Now, define function \(F^n\) as follows:

\[
F^n(\beta) = E^n[g(X^n_1, \beta)],
\]

(12)

where \(E^n\) denotes the expectation under the probability measure for \(\{X^n_k\}\). Obviously, \(F^n\) is the conditional moment for \(X^n\). Since the transition probabilities for \(X^n_{k+1}\) conditioning on \(X^n_k\) are constant and equal to \(\frac{1}{2}\), the expectation of \(g(X^n_k, \beta)\) can be easily computed as

\[
F^n(\beta) = \frac{1}{2^n} \sum_{X^n_k} g(X^n_k, \beta).
\]

(13)

Moreover, the derivatives of \(F^n\) with respect to \(\beta\) can be computed explicitly. Applying the chain rule, one can verify that

\[
\frac{\partial F^n(\beta)}{\partial \beta} = E^n \left[ \frac{\partial g(X^n_k, \beta)}{\partial x} D^n_k + \frac{\partial g(X^n_k, \beta)}{\partial \beta} \right],
\]

where \(D^n_k = \frac{\partial X^n_k}{\partial \beta}\) is a \(\tau\)-dimensional vector of derivative process of \(X^n_k\) with respect to \(\beta\), and \(D^n_k\) satisfies the following stochastic difference equation:

\[
\begin{align*}
D^n_{k+1} &= D^n_k + (b_\beta(X^n_k, \beta) + b_x(X^n_k, \beta) D^n_k) \frac{1}{n} + (\sigma_\beta(X^n_k, \beta) + \sigma_x(X^n_k, \beta) D^n_k) \frac{\xi_k}{\sqrt{n}}, \\
D^n_0 &= 0,
\end{align*}
\]

\(^3\) Generalization to \(E^n[g(X(\tau), \beta)|X(\tau-1)]\) is obvious.

\(^4\) Let \(\hat{X}^n = X^n_{[n\delta]}\). According to Essay Two, \(\hat{X}^n\) converge to \(X\) weakly in \(D[0, 1]\).
where \( b_\beta, b_\tau, \sigma_\beta \) and \( \sigma_\tau \) denote the partial derivatives of \( b(z, \beta) \) and \( \sigma(\tau, \beta) \) with respect to \( z \) and \( \beta \), respectively.\(^5\) The following proposition establishes a uniform convergence from \( F^n \) and \( \frac{\partial F^n}{\partial \beta} \) to \( F \) and \( \frac{\partial F}{\partial \beta} \), respectively.

**Proposition 1** Suppose the following:

a) \( b \) and \( \sigma \) are continuously differentiable with respect to \( z \) up to the seventh order, \( b_\beta \) and \( \sigma_\beta \) exist and are continuously differentiable with respect to \( z \) up to the sixth order, and all these derivatives satisfy a polynomial growth condition (with possibly different \( K' \) and \( \gamma' \));

b) \( g \) and \( g_\beta \) are continuously differentiable up to the seventh and the sixth order, respectively, and all these derivatives satisfy a polynomial growth condition;

c) \( b_\beta(z, \beta) y, \sigma_\beta(z, \beta) y, b_\beta(z, \beta) \) and \( \sigma_\beta(z, \beta) \) satisfy a uniform Lipschitz condition in \( z \) and \( y \) for fixed \( \beta \in \Theta \), respectively.

Then there exist constants \( C \) and \( \gamma' > 0 \), independent of \( n \) and \( \beta \), such that for all \( \beta \in \Theta \),

\[
\begin{align*}
1) \ & |F(\beta) - F^n(\beta)| \leq \frac{C}{n} (1 + |X_0|^{\gamma'}), \\
2) \ & \left| \frac{\partial F(\beta)}{\partial \beta} - \frac{\partial F^n(\beta)}{\partial \beta} \right| \leq \frac{C}{n} (1 + |X_0|^{\gamma'}).
\end{align*}
\]

**Proof.** See Appendix C.

**Remark 1** Condition c) can be relaxed to requiring only that \( b(z, \beta) \) and \( \sigma(z, \beta) \) satisfy a uniform Lipschitz condition in \( z \). However, the stronger assumptions given by c) greatly simplifies the proof for 2).

Proposition 1 shows that the discretization scheme defined by (11) produces approximations for conditional moments of the order \( \frac{1}{n} \). Milstein [1974] develops a discretization scheme that produces approximations for conditional moments of the order \( \frac{1}{n^2} \). However, when \( X \) is multi-dimensional, the Milstein's scheme requires additional assumptions on the diffusion term \( \sigma \), which may not be satisfied in general.

The discrete time approximation for \( X \) when it is multi-dimensional can be dealt with similarly. This involves approximating the \( N \)-dimensional Brownian motion by a sequence of multinomial processes. For example, a multinomial process formed by putting \( N \) independent binomial processes together makes up an approximation for the \( N \)-dimensional Brownian motion. We refer the readers to Essay Two for an \( N + 1 \)-nomial approximation scheme for the \( N \)-dimensional Brownian motion.

\(^5\)More generally, following Gihman and Skorohod [1975], we have \( \frac{\partial F(\beta)}{\partial \beta} = E \left[ \frac{\partial f(X(1), \beta)}{\partial \beta} D(1) + \frac{\partial f(X(1), \beta)}{\partial \beta} \right] \), where \( D \) is the derivative process of \( X \) w.r.t. \( \beta \), satisfying

\[
dD(t) = (b_\beta(X(t), \beta) + b_\tau(X(t), \beta))D(t)dt + (\sigma_\beta(X(t), \beta) + \sigma_\tau(X(t), \beta))D(t)dw(t).
\]
Finally, in order to obtain an approximation for the objective function, the above discretization scheme is used repeatedly to get approximations for the conditional moments \( F(X_t, \beta) = E^{\beta}[g_t(X, \beta)|X_{t-1}] \), where \( t = 1, 2, \cdots, T \). Once \( F \) is approximated by \( F^n \), we can then replace \( f_t \) by
\[
 f^n_t = g_t(X, \beta) - F^n(X_t, \beta),
\]
and apply GMM estimation to get the GMM estimator \( \hat{\beta}^n_T \) as illustrated in (9).

4. Large Sample Properties

In this section, we will establish large sample properties for the GMM estimators \( \hat{\beta}_T \) and \( \hat{\beta}^n_T \) defined in Section 2. The basic treatment follows closely from Hansen [1982].

4.1. GMM Estimation when \( f_t^\beta \) is Known.

For the purpose of large sample properties, we need to make some primitive assumptions on the maximal degree of dependence allowable for the discretely sampled observations \( \{X_t\}_{t=1}^\infty \) while still permitting some form of the law of large numbers and the central limit theorem to obtain.\(^6\) The standard assumption requires that \( \{X_t\} \) be stationary and ergodic, as in Billingsley [1969] and Hansen [1982]. Mcleish [1975] shows that one could replace the stationarity and ergodicity assumptions by a mixing condition (for an elementary discussion of the mixing condition, see White [1988]). For stationary sequence, the mixing condition is stronger than ergodicity, since it implies ergodicity. More recently, Galant and White [1988] use a notion of near epoch dependency, which generalizes the notion of mixing to a time series whose observations may depend upon the entire history of another time series. In practice, both mixing and ergodicity condition are difficult to verify. Alternatively, Duffie and Singleton [1989] introduce a notion of geometric ergodicity, for which easily verifiable sufficient conditions exist, although they may be too strong to apply in practice.\(^7\) Intuitively, geometric ergodicity implies that the stochastic process is asymptotically stationary and ergodic. For the purpose of this essay, any of the above mentioned assumptions is a valid candidate. Without loss of generality, we adopt

\(^6\)In cases where the diffusion process has a long-run growth trend, \( X \) may be a transformation of the original process such that \( X \) is trend-free.

\(^7\)Formally, let \( P_t^\rho \) denote the transition probability of \( X(t) \) given the initial point \( X(0) = x \) for a time-homogeneous Markov process \( \{X(t)\} \). The process \( \{X(t)\} \) is called geometric ergodic, if there exists some \( \rho \in (0, 1) \), and a probability measure \( \pi \) on the state space of the process such that, for every initial point \( x \),
\[
\rho^{-t}||P_t^\rho - \pi||_* \to 0, \text{as } t \to \infty,
\]
where \( || \|_* \) denotes the total variation norm. According to Duffie and Singleton [1989], geometric ergodicity implies \( \alpha \)-mixing.
the stationarity and ergodicity assumptions, which would then allow us to quote from Hansen [1982] the existing theorems for establishing consistency and asymptotic normality for the GMM estimators.

We summarize several important regularity conditions for consistency.

**Assumption 1 (Ergodicity)** \( \{X_t, t = 1, 2, \ldots, \} \) is stationary and ergodic.

When the diffusion process \( \{X(t), t \in [0, \infty)\} \) is stationary and ergodic, the sequence of discretely sampled observations \( \{X_t, t = 1, 2, \ldots, \} \) is also stationary and ergodic. Hence, \( \{f_t^\beta, t = 1, 2, \ldots, \} \) is stationary and ergodic for all \( \beta \in \Theta \).

**Assumption 2 (Weighting Matrix)**

\[
p_{\lim\limits_{T \to \infty}} W_T = W_0,
\]

where \( W_0 \) is an \( I \times I \), non-stochastic, and semi-positive matrix with \( \text{rank}(W_0) \geq r \).

The weighting matrix \( W_T \) defines a measure of the length of the sample moments \( G_T(\beta) \). As suggested in Hansen [1982], one may choose \( W_0 = S_0^{-1} \), which yields an efficient GMM estimator, where

\[
S_0 \equiv \sum_{j=-\infty}^{\infty} E(f_t^{\beta_0} f_{t-j}^{\beta_0}^T).
\]

In our setup, since the moment conditions are formed by using conditional expectations, \( \{f_t^{\beta_0}\} \) is a martingale difference sequence, which implies that \( E(f_t^{\beta_0} f_{t-j}^{\beta_0}^T) = 0 \) for \( j \neq 0 \). Hence, \( S_0 \) simply equals to \( E[f_t^{\beta_0} f_t^{\beta_0}^T] \).

**Assumption 3 (Regularity Condition I)**

i) \( f_t(X, \beta) \) is measurable in \( X \) for each \( \beta \in \Theta \) and continuous in \( \beta \) for fixed \( X \);

ii) \( E f_t(X, \beta) \) exists and is finite for all \( \beta \in \Theta \);

iii) \( f_t(X, \beta) \) is first moment continuous at \( \beta \) for all \( \beta \in \Theta \);

iv) Let \( H(\beta) = E f_t^T(X, \beta) W_0 E f_t(X, \beta) \). Then \( H(\beta) \) has a unique minimum at \( \beta_0 \).

Formally, a random function \( f(\omega, \beta) \) is first moment continuous at \( \beta \) if

\[
\lim_{\delta \to 0} E[\epsilon(\omega, \beta, \delta)] = 0,
\]

where \( \epsilon(\omega, \beta, \delta) = \sup \{|f(\omega, \beta) - f(\omega, \alpha)| : \alpha \in \Theta, \ |\beta - \alpha| < \delta\} \). A sufficient condition for \( f \) to be first moment continuous is that there exists a \( \delta > 0 \) such that \( E[\epsilon(\omega, \beta, \delta)] < \infty \) (see Lemma 2.1, Hansen [1982]). As a consequence, \( f \) is first moment continuous if

\[
E \sup_{\beta \in \Theta} f_t^\beta < \infty. \quad (14)
\]
The first moment continuity condition ensures that when we apply the weak law of large numbers to \( \{f_t^\beta\} \), the sample moments converge in probability to their population moments uniformly in \( \beta \).

We present the following theorem on the consistency of the GMM estimator \( \hat{\beta}_T \) for the case where the functional form of conditional moments is known. The proof of this theorem follows directly from Theorem 2.1, Hansen [1982].

**Theorem 1 (Consistency for \( \hat{\beta}_T \))** Suppose Assumptions 1 - 3 are satisfied, then \( \hat{\beta}_T \) exists and converges in probability to \( \beta_0 \).

We need two more assumptions for deriving the asymptotic distribution of \( \hat{\beta}_T \).

**Assumption 4 (Regularity Condition II)**

(i) The estimators \( \hat{\beta}_T \) lie in the interior of \( \Theta \);  
(ii) \( f_t^{\beta} \) is continuously differentiable with respect to \( \beta \in \Theta \) for all \( t \);  
(iii) \( \partial f_t^{\beta} / \partial \beta \) is first moment continuous at \( \beta_0 \);  
(iv) \( d_0 = E[\partial f_t^{\beta_0} / \partial \beta] \) exists, and has full rank.

**Assumption 5** \( S_0 = E[f_t^{\beta_0} f_t^{\beta_0}] \) exists and is invertible. Moreover, \( \text{plim}_T W_T = S_0^{-1} \).

The following proposition offers the asymptotic distribution for \( \hat{\beta}_T \), the proof of which also follows directly from Theorem 3.1, Hansen [1982].

**Theorem 2 (Asymptotic Normality for \( \hat{\beta}_T \))** Under Assumptions 1 - 5, \( \sqrt{T}(\hat{\beta}_T - \beta_0) \rightarrow N(0, D) \) as \( T \rightarrow \infty \), where \( D = (d_0^T S_0^{-1} d_0)^{-1} \).

4.2. GMM Estimation when \( f_t \) is Unknown

We now investigate large sample properties for the GMM estimator \( \hat{\beta}_T^n \) in cases where the functional forms of the conditional moments are unknown and the actual conditional moments are replaced by their discrete time approximations. We will show that under some regularity conditions, \( \hat{\beta}_T^n \) has the same large sample distribution as \( \hat{\beta}_T \) when \( T \) and \( n \) go to infinity.

We need to make a similar set of regularity assumptions.

\(^4\) Stronger regularity conditions such as Lipschitz continuity (or smoothness) are required for the uniform law of large numbers (ULLN's) to hold, if \( X \) is mixing or geometrically ergodic, but not stationary (see Andrew [1987], Duffie and Singleton [1980] and Potiescher and Prucha [1989]).
Assumption 2' (Weighting Matrix)

\[ \lim_{T \to \infty} W^n_T \rightarrow W^n, \]

where \( W^n \) is an \( I \times I \), non-stochastic, and semi-positive matrix with \( \text{rank}(W^n) \geq \tau \). Moreover,

\[ \| W^n - W_0 \| \rightarrow 0, \quad \text{as } n \rightarrow \infty, \]

where \( W_0 \) is defined as in Assumption 2, and \( \| A \| = \text{trace}(A^\top A)^{\frac{1}{2}} \) denotes the norm of matrix \( A \).

Assumption 3' (Regularity Condition I')

i) \( f^n_t(X, \beta) \) is measurable in \( X \) for each \( \beta \in \Theta \) and continuous in \( \beta \) for fixed \( X \);

ii) \( E f^n_t(X, \beta) \) exists and is finite for all \( \beta \in \Theta \);

iii) \( f^n_t(X, \beta) \) is first moment continuous at \( \beta \) for all \( \beta \in \Theta \);

iv) Let \( H^n(\beta) = E f^n_t(X, \beta)^\top W^n E f^n_t(X, \beta) \), \( H^n(\beta) \) has a unique minimum at \( \beta^n \), where \( \beta^n \) lies in the interior of \( \Theta \).

To get a feeling of Assumption 3', we introduce the following lemma.

Proposition 2 Let

\[ R^n(\beta) = E[f^n_t(X, \beta)], \quad R(\beta) = E[f_t(X, \beta)], \]

\[ J^n(\beta) = E[\frac{\partial f^n_t(X, \beta)}{\partial \beta}], \quad J(\beta) = E[\frac{\partial f_t(X, \beta)}{\partial \beta}], \]

\[ H^n(\beta) = E[f^n_t(X, \beta)^\top W^n E[f^n_t(X, \beta)], \quad H(\beta) = E[f_t(X, \beta)^\top W_0 E[f_t(X, \beta)], \]

Suppose that the assumptions in Proposition 1 are satisfied, and suppose further that \( E|Y|^\lambda < \infty \) for any \( \lambda > 0 \). Then \( R^n, J^n, \) and \( H^n \) converge uniformly on \( \Theta \) to \( R, J \) and \( H \), respectively.

PROOF. See Appendix C. \[ \]

Given this proposition, condition ii) and iii) of Assumption 3' can be easily derived from condition ii) and iii) of Assumption 3. Moreover, since we have assumed that \( H(\beta) \) has a unique zero at \( \beta_0 \), condition iv) of Assumption 3' is reasonable to impose, given that \( H(\beta) \) converges to \( H^n(\beta) \) uniformly. We now present the following consistency result for \( \hat{\beta}_n^T \).

Theorem 3 (Consistency for \( \hat{\beta}_n^T \))
Under Assumptions 1, 2', 4' and the assumptions for Proposition 2, $\hat{\beta}_T^n$ defined in (9) is consistent as $(n,T)$ goes to infinity in the following sense. For any $\epsilon, \delta > 0$, there exist $n_0 > 0$ and $T(n) > 0$ for $n > n_0$, such that

$$P[|\hat{\beta}_T^n - \beta_0| > \delta] < \epsilon,$$

for all $n > n_0$ and $T > T(n)$.

**Proof.** Under the given assumptions, for fixed $n$, $G_T^n(\beta)^T W_T^n G_T^n(\beta) \rightarrow H^n(\beta)$ in probability uniformly on $\Theta$ as $T$ goes to infinity. This implies that $\hat{\beta}_T^n \rightarrow \beta^n$ in probability as $T \rightarrow \infty$, see Amemiya [1985].

According to Proposition 2, $H^n$ converges to $H$ uniformly, therefore, $\beta^n \rightarrow \beta_0$ as $n \rightarrow \infty$. Thus, for any $\delta > 0$, there exists $n_0 > 0$ such that $|\beta^n - \beta_0| < \delta/2$ for all $n > n_0$. Now, for any $\epsilon > 0$, since $\text{plim}_T \hat{\beta}_T^n = \beta^n$, we can find a $T(n) > 0$ such that as $T > T(n)$, $P[|\hat{\beta}_T^n - \beta^n| > \delta/2] < \epsilon$. This implies that

$$P[|\hat{\beta}_T^n - \beta_0| > \delta] \leq P[|\hat{\beta}_T^n - \beta^n| > \delta/2] < \epsilon,$$

for all $n > n_0$ and $T > T(n)$. \qed

It is clear from the proof that for fixed $n$, $\hat{\beta}_T^n$ may be biased even though $T$ goes to infinity. We must let both $n$ and $T$ go to infinity in order for $\hat{\beta}_T^n$ to be consistent. We make some additional regularity assumptions, which will yield an asymptotic distribution for $\hat{\beta}_T^n$.

Assumption 4' (Regularity Condition II')

(i) The estimators $\hat{\beta}_T^n$ lie in the interior of $\Theta$;

(ii) $f_t^n(X, \beta)$ is continuously differentiable with respect to $\beta \in \Theta$ for all $t$;

(iii) $\partial f_t^n(X, \beta)/\partial \beta$ is first moment continuous at $\beta^n$;

(iv) $d^n = E[\partial f_t^n(X, \beta^n)/\partial \beta]$ exists, and has full rank.

Assumption 5' $S^n = E[f_t^n(X, \beta^n)f_t^n(X, \beta^n)^T]$ exists and is invertible. Moreover, $\text{plim}_T W_T^n = S^{n-1}$.

The following theorem derives the asymptotic distribution for the GMM estimator $\hat{\beta}_T^n$.

**Theorem 4 (Asymptotic Normality for $\hat{\beta}_T^n$)** Under Assumptions 1, 2', 4' and the assumptions for Proposition 2, $\sqrt{T}(\hat{\beta}_T^n - \beta_0) \rightarrow N(0, D)$ as $(n, \infty) \rightarrow \infty$, where $D = (d_0^T S_0^{-1} d_0)^{-1}$.

**Proof.** Following Theorem 3.1, Hansen [1982], we know that for fixed $n$,

$$\sqrt{T}(\hat{\beta}_T^n - \beta^n) \rightarrow N(0, (d^n^T S^n d^n)^{-1}), \quad \text{as } T \rightarrow \infty,$$

87
where \( d^n = E^{\frac{\partial f_t(X, \beta^n)}{\partial \theta}} \).

However, according to Proposition 2, \( E^{\frac{\partial f_t(X, \beta^n)}{\partial \theta}} \) converges to \( E^{\frac{\partial f_t(X, \beta)}{\partial \theta}} \) uniformly on \( \Theta \). Following the proof for Proposition 2, it is also clear that \( E f_t(X, \beta) f_t(X, \beta)^\top \) converges to \( E f_t(X, \beta) f_t(X, \beta)^\top \) uniformly on \( \Theta \). We conclude that \( d^n \to d_0 \) and \( S^n \to S_n \) as \( n \to \infty \). Hence,

\[
\sqrt{T}(\hat{\beta}_T - \beta_0) \to N(0, D),
\]

in distribution as \( (n, T) \to \infty \) in a similar manner as in Theorem 3, where \( D = (d_0^\top S_0^{-1} d_0)^{-1} \).

This concludes our demonstrations for the large sample properties of the GMM estimators.

5. Empirical Example: Square Root Process

In this section, we apply the proposed GMM estimation to a mean reverting square root process originally used by Cox, Ingersoll and Ross [1985] for modeling the dynamics of spot interest rate. We will present the results of several sampling experiments which explore the finite sample properties of the GMM estimators \( \hat{\beta}_T \) and \( \hat{\beta}_T^n \) developed in the previous sections. In particular, these experiments will examine whether the finite distributions of \( \hat{\beta}_T \) and \( \hat{\beta}_T^n \) are close to their theoretical asymptotic normal distributions as \( T \) and \( n \) become large.

Formally, the mean reverting square root process \( \{X(t), t \geq 0\} \) is determined by the following stochastic differential equation:

\[
dX(t) = \kappa(\theta - X(t)) + \sigma \sqrt{X(t)} \, dw(t),
\]

where \( \theta \) is interpreted to be the long-run mean of \( X(t) \) and \( \kappa \) to be the mean reverting intensity. We assume that \( \kappa, \theta > 0 \) and \( 2\kappa \theta \geq \sigma^2 \), which imply that zero is inaccessible. The initial distribution of \( X(0) \) is assumed to be drawn from a gamma distribution, whose density function has the following form:

\[
f(x) = \frac{\omega^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\omega x}, \quad x \geq 0,
\]

where \( \omega = \frac{2\kappa}{\sigma^2} \), and \( \nu = \frac{2\kappa \theta}{\sigma^2} \). It is known that the mean reverting square root process \( \{X(t), t \geq 0\} \) is stationary and ergodic if zero is inaccessible.\(^9\)

\(^9\)Stationarity can be verified directly. To prove ergodicity, one can show that there exists a constant \( \lambda > 0 \) such that

\[
\|P_t^\nu - \pi\| \leq \lambda e^{-\alpha t},
\]

where \( \pi \) is the steady state probability measure on the state space that has a gamma distribution function. This implies that \( X \) is geometrically ergodic, and therefore is \( \alpha \)-mixing, according to Duffie and Singleton [1989]. However, mixing implies ergodicity.
The transition density function of \( X(s) \) given \( X(t) \), when \( s > t \), is given by

\[
f(X(s), s ; X(t), t) = ce^{-u-v} \left( \frac{v}{u} \right)^{\frac{q}{2}} I_{q}(2(uv)^{\frac{1}{2}}),
\]

where

\[
c \equiv \frac{2\kappa}{\sigma^{2}(1 - e^{-\kappa(s-t)})},
\]
\[
u \equiv cX(t)e^{-\kappa(s-t)},
\]
\[
v \equiv cX(s),
\]
\[
q \equiv \frac{2\kappa\theta}{\sigma^{2}} - 1,
\]

and \( I_{q}(\ ) \) is the modified Bessel function of the first kind of order \( q \), defined as

\[
I_{q}(z) = \sum_{j=0}^{\infty} \frac{1}{\Gamma(j+1)\Gamma(j+1+q)} \left( \frac{z}{2} \right)^{2j+q},
\]

with

\[
\Gamma(a) = \int_{0}^{\infty} t^{a-1}e^{-t}dt,
\]

the gamma function. Given the transition density, it is easy to verify that the conditional moments of \( X(s) \) and \( X(s)^{2} \) given \( X(t) \) for \( s > t \) are as follows,

\[
E[X(s)|X(t)] = X(t)e^{-\kappa(s-t)} + \theta(1 - e^{-\kappa(s-t)}),
\]
\[
E[X(s)^{2}|X(t)] = X(t)\frac{\sigma^{2}}{\kappa}(e^{-\kappa(s-t)} - e^{-2\kappa(s-t)}) + \theta\frac{\sigma^{2}}{2\kappa}(1 - e^{-\kappa(s-t)})^{2}
\]
\[
  (X(t)e^{-\kappa(s-t)} + \theta(1 - e^{-\kappa(s-t)})])^{2}.
\]

Although it is possible to perform ML estimation, given that we have obtained the analytical expression for the transition density, we will focus only on GMM estimation.

We conduct sampling experiments in the following manner. First, for a fixed set of parameters of \( \kappa, \sigma \) and \( \theta \), a random sample of weekly observations of \( X \) is generated. Second, a GMM estimation is performed, which yields the point estimates and the estimated variance-covariance matrix for \( \kappa, \sigma \) and \( \theta \). This sampling-estimation procedure is then repeated independently for 500 times, which produces 500 replications of sampling experiment. Since the mean reverting intensity parameter \( \kappa \) has a strong effect on the pattern of the sample path of \( X \), three sets of parameters \((\kappa = 0.8, 4, 40)\) are used in these experiments, while \( \theta \) and \( \sigma \) are held constant at 0.1. The effect of mean reversion intensity on the statistical properties of point estimates will be explored later.
In order to generate a random sample of size $T$ for $X$, the stochastic differential equation (1) is discretised and simulated in the following way,

$$
\tilde{X}_{k+1}^h = \tilde{X}_k^h + \kappa (\theta - \tilde{X}_k^h) h + \sigma \sqrt{X_k^h} \sqrt{h} \, \tilde{e}_k,
$$

(17)

where $h$ is set to be $1/(52 \times 200)$, $\tilde{e}_k$ is a sequence of i.i.d. random variables and is distributed as $N(0,1)$. The initial distribution of $X(0)$ is drawn from the gamma distribution defined as in (16). The sample observations, $(X_1, X_2, \ldots, X_T)$, are then obtained by setting

$$X_t = \tilde{X}_{200(t-1)}, \quad t = 1, 2, \ldots, T.$$

We admit that the discretization procedure in (17) may incur errors for $X$ to become a random sample drawn from the square root process. However, it is reasonable to believe that such effect is insignificant, since it is known that $\tilde{X}_t^h$ converges to the square root process in distribution as $h \to 0$. Experiments are performed for sample sizes of $T = 260, 520, 780$ and $1040$, which correspond in real time to series of weekly observations of lengths 5, 10, 15 and 20 years, respectively.

Since there are three unknown parameters, we need at least three moment conditions. The following three are used in our estimation,

$$
\begin{align*}
\ell_1(\beta) &= X_{t+1} - E^\beta[X_{t+1}\mid X_t], \\
\ell_2(\beta) &= X_t^2 - E^\beta[X_t^2\mid X_t], \\
\ell_3(\beta) &= \frac{(X_{t+1} - E^\beta[X_{t+1}\mid X_t])^2}{\text{Var}^\beta(X_{t+1}\mid X_t)} - 1,
\end{align*}
$$

where $\beta = (\kappa, \sigma, \theta)^T$,

$$
\begin{align*}
E^\beta[X_{t+1}\mid X_t] &= X_t e^{-\kappa \Delta} + \theta(1 - e^{-\kappa \Delta}), \\
E^\beta[X_t^2\mid X_t] &= X_t \kappa^2 (e^{-\kappa \Delta} - e^{-2\kappa \Delta}) + \theta \kappa^2 (1 - e^{-\kappa \Delta})^2 + (X_t e^{-\kappa \Delta} + \theta(1 - e^{-\kappa \Delta}))^2, \\
\text{Var}^\beta(X_{t+1}\mid X_t) &= X_t \kappa^2 (e^{-\kappa \Delta} - e^{-2\kappa \Delta}) + \theta \kappa^2 (1 - e^{-\kappa \Delta})^2,
\end{align*}
$$

and $\Delta = 1/52$, which gives a time interval of one week. It is not difficult to verify that Regularity Conditions I and II are satisfied, as long as $\kappa$, $\sigma$ and $\theta$ are assumed to be positive and lie in a compact set that does not contain 0. Now, define

$$G_T(\beta) = \frac{1}{T} \sum_{t=1}^{T} \ell_t(X_t, \beta).$$
Our estimation procedure is carried out by performing the following minimization over \( \beta \),

\[
\min_{\beta} G_T(\beta)^T G_T(\beta).
\]

Since the parameters are exactly identified, the optimal solution is obtained when \( \hat{\beta}_T \) is such that \( G(\hat{\beta}_T) = 0 \). To calculated the variance-covariance matrix, we define \( \hat{W}_T \equiv \Sigma^{-1}(\hat{\beta}_T) \), where

\[
\Sigma(\hat{\beta}_T) \equiv \frac{1}{T} \sum_{t=1}^{T} f_t(\hat{\beta}_T) f_t(\hat{\beta}_T)^T.
\]

The asymptotic variance-covariance matrix for \( \hat{\beta}_T \) is then calculated as

\[
\text{Var}(\hat{\beta}_T) = \frac{1}{T} \left[ D^T(\hat{\beta}_T) \hat{W}_T D(\hat{\beta}_T) \right]^{-1},
\]

(18)

where

\[
D(\hat{\beta}_T) \equiv \frac{1}{T} \sum_{t=1}^{T} \frac{\partial f_t}{\partial \beta} |_{\beta = \hat{\beta}_T}.
\]

Tables 1, 2, and 3 summarize the basic results of simulations for \( \kappa = 0.8 \) with \( T = 520, 780, 1040 \) and \( \kappa = 4, 40 \) with \( T = 260, 620, 780 \). Each table corresponds to a separate independent experiment, which in turn corresponds to 500 independent replications. The second column displays the values of the true parameters. The third column displays the means and standard deviations of the point estimates of \( \hat{\beta}_T \). The fourth column reports the theoretical values for the asymptotic variance and standard deviation of estimators \( \hat{\beta}_T \), while the fifth column reports the means and standard deviations of the estimated asymptotic variance calculated according to (18). The next three columns display estimated 1, 5 and 10 percent of tail probabilities, respectively.\(^{11}\) Asymptotic \( t \)-statistics for the hypothesis that the true tail probabilities are 1, 5 and 10 percent respectively are reported in parentheses. The last column reports the Pearson \( \chi^2 \)-statistics and their probability values under the null hypothesis that \( \hat{\beta}_T \)

\(^{10}\) Usually, GMM estimation is performed in two stages. In the first stage, a consistent estimator is found by using an arbitrary weighting matrix. This consistent estimate is then used to form an optimal weighting matrix to yield an efficient consistent estimator. However, when the number of moment conditions equals exactly to the number of unknown parameters, any weight matrix will yield the same asymptotic variance-covariance matrix for \( \hat{\beta}_T \).

\(^{11}\) For given size \( \alpha \), let \( A_\alpha \) be such that

\[
P[|\hat{\kappa}_T - \kappa_0| / \sigma(\hat{\kappa}_T) > A_\alpha] = \alpha,
\]

where \( \hat{\kappa}_T \sim N(\kappa_0, \sigma(\hat{\kappa}_T)^2) \), then the tail probability for \( \hat{\kappa}_T \) is defined to be \( \frac{m}{100} \), where \( m \) is the total number of times among 500 replications that \( \hat{\kappa}_T \) lies outside of \((\kappa_0 - \sigma(\hat{\kappa}_T)A_\alpha, \kappa_0 + \sigma(\hat{\kappa}_T)A_\alpha)\). The tail probabilities for \( \hat{\kappa}_T \) and \( \hat{\beta}_T \) are defined similarly.
has an (asymptotic) normal distribution.\textsuperscript{12}

The results reported in Tables 1, 2 and 3 suggest that the point estimates for $\sigma$ and $\theta$ are consistent in all nine experiments. As we would have expected, the standard deviations decrease as the sample size $T$ increases. While the standard deviations for $\hat{\sigma}_T$ in column three are largely unchanged for three different levels of $\kappa$, the standard deviations for $\hat{\theta}_T$ decrease as $\kappa$ increases. Moreover, the standard deviations of point estimates in column three are close to their theoretical values for the asymptotic standard deviation in column four, while the means of the estimated asymptotic variances in column five are also close to their theoretical values for the asymptotic variance in column four. In addition, the $t$-statistics for the tail probabilities as well as the $\chi^2$ probability values are largely insignificant. These results suggest that the finite sample distributions of $\hat{\sigma}_T$ and $\hat{\theta}_T$ match the asymptotic normal distributions established in Section 4 reasonably well.

However, the finite sample properties of the point estimates for $\kappa$ are much worse than those of $\sigma$ and $\theta$. When the true $\kappa$ is 0.8, the point estimates for $T = 520, 780$ and 1040 are strongly biased, and the tail probabilities as well as the $\chi^2$ statistics are significantly different from their corresponding true values, although the point estimate for $T = 1040$ improves quite a bit. It is clear from Table 1 that when the true $\kappa$ is equal to 0.8, a longer time series is needed in order to obtain good estimates for $\kappa$. As the true $\kappa$ increases to 4 and 40, the finite sample properties of the point estimates improve significantly, and point estimates for $\kappa$ are largely consistent for $T \geq 520$. The $t$-statistics for the tail probabilities and $\chi^2$ probability values for $T \geq 520$ are largely insignificant, and the means and standard deviations of the point estimates are close to their corresponding theoretical values, which suggest that the finite distributions of $\hat{\kappa}_T$ resemble their theoretical asymptotic normal distributions quite well.

Overall, the results strongly suggest that it is much easier to estimate $\kappa$ when the true $\kappa$ is large than it is small. This is not surprising. When the true $\kappa$ is small, the sequence looks more or less like a random walk or a unit root process, which makes it difficult to identify $\kappa$. When the true $\kappa$ is large, the mean reverting effect dominates the random effect, which helps identify the mean reverting intensity parameter $\kappa$.

We now consider the GMM estimator $\hat{\beta}^2_T$, where we replace $f^1$, $f^2$ and $f^3$ by the conditional

\textsuperscript{12}We divide the real line into 20 intervals such that each interval has a probability of 0.05 under the null hypothesis. The Pearson $\chi^2$-statistics is constructed as

$$\chi^2 = \sum_{i=1}^{20} \frac{(f_i - 0.05 \times 500)^2}{0.05 \times 500},$$

where $f_i$ is the frequency of the $i$-th interval. The probability value is calculated as $P[\chi^2(19) \geq \chi^2]$. 

92
moments calculated from (13). Since the uniform Lipschitz condition required for Proposition 1 is not satisfied for the square root process, we can not claim that the moments obtained from the binomial process will produce good approximations to the actual moments. We therefore compare the values and derivatives of the actual conditional moments and the actual objective function to their corresponding approximations. The results are reported in Tables 4, 5, 6 and 7, where we have fixed the parameters at $\kappa = 0.8, 4, 40$ and $\sigma = \theta = 0.1$. Clearly, from these tables, the approximations for the first three moments and the objective function are quite good, with $\kappa = 0.8$ being the best. These comparisons are only meant to be suggestive. A thorough investigation that involves comparing moments for parameters at different levels is needed before we can claim that conditional moments obtained from the binomial processes produce good approximations in the case of square root process.

As an illustration, we apply GMM estimation for $\kappa = 4, T = 520$, and $n = 12$ to $10$ independent sample observations. The results of comparisons between $\hat{\beta}_T$ and $\hat{\beta}_T^n$ are reported in Table 8, where each row corresponds to one pair of estimates obtained by using the same sequence of time series observations of size 520. The point estimates of $\hat{\beta}_T^n$ and $\hat{\beta}_T$ and the estimated standard deviations are virtually identical. We conclude that $\hat{\beta}_T^n$ has the same asymptotic distribution as $\hat{\beta}_T$.

6. Concluding Comments

We have developed an approach for estimating the parameters of general diffusion process using GMM. Our procedure involves approximating the conditional moments when the functional form of conditional moments are not available. As we have demonstrated, this procedure is easy to apply in practice. We observe that when a diffusion process is approximated by a multinomial process, the probability distribution of the diffusion process is also approximated by that of the multinomial process. This suggests that maximum likelihood estimation may be applicable. ML estimation is interesting since it will produce an efficient estimator if the parametric specification of the diffusion process is correct. The main difficulty for applying ML estimation is how to recover the density function from the estimated distribution function, which is not a continuous function in this case. A good reference on recovering density function from sample observations is Silverman [1981].
References


Appendix C.

Proof of Proposition 1. We sketch the proof, since it is similar to the proof for Theorem 2 in Essay Two. First, we define

$$V(z, t, \beta) = E[g(X(1), \beta) | X(t) = z],$$

then $V$ satisfies the following partial differential equation:

$$\frac{1}{2} \sigma(z, \beta)^2 V_{xx}(z, t, \beta) + b(z, \beta) V_x(z, t, \beta) + V_t(z, t, \beta) = 0,$$

(19)

$$V(z, 1, \beta) = g(z, \beta).$$

(20)

Next, we define

$$V^n(X_{k+1}^n, \frac{k}{n}, \beta) = E_n[g(X_n^k, \beta) | X_n^k],$$

then $V^n$ satisfies the following recurrent (difference) equation:

$$V^n(X_{k+1}^n, \frac{k}{n}, \beta) = \frac{1}{2} V^n(X_{k+1}^n, \frac{k+1}{n}, \beta) + \frac{1}{2} V^n(X_{k+1}^n, \frac{k+1}{n}, \beta),$$

(21)

where

$$X_{k+1}^n = X_k^n + \frac{b(X_k^n, \beta)}{n} + \frac{\sigma(X_k^n, \beta)}{\sqrt{n}},$$

$$X_{k+1}^{n-} = X_k^n + \frac{b(X_k^n, \beta)}{n} - \frac{\sigma(X_k^n, \beta)}{\sqrt{n}}.$$

Now, let $e^n_k = V(X_{k+1}^n, \frac{k}{n}, \beta) - V^n(X_{k+1}^n, \frac{k}{n}, \beta)$, we need to show that $|e^n_k| \leq \frac{C}{n} (1 + |X_0|)^\gamma'$ for some constants $C$ and $\gamma'$. This can be done as follows.

Applying the Taylor expansion to $V$, we get

$$\frac{1}{2} V(X_{k+1}^n, \frac{k+1}{n}, \beta) + \frac{1}{2} V(X_{k+1}^{n-}, \frac{k+1}{n}, \beta) = V(X_k^n, \frac{k}{n}, \beta) + \frac{1}{n^2} h(X_k^n) + \gamma^n_k,$$

where $h = \frac{1}{2} V_{xx}\sigma^2 + V_{xt} b + \frac{1}{2} V_{tt}$ and $\gamma^n_k$ is in the order of $n^{-3}$. This allows us to get a recurrent equation for $e^n_k$,

$$e^n_k = \frac{1}{2} (e^n_{k+1} + e^n_{k+1}) - \frac{1}{n^2} h(X_k^n, \frac{k}{n}, \beta) - \gamma^n_k.$$

(22)

In the proof of Theorem 2 in Essay Two, it was shown that (22) implies that there exist constants $C$ and $\gamma' > 0$, independent of $n$ and $\beta$, such that

$$|e^n_k| \leq \frac{C}{n} (1 + |X_0|)^{\gamma'}.$$

The proof for the second part is exactly the same, if we treat $(X, D)$ as a two-dimensional diffusion process, where $D = \frac{\partial^2}{\partial^2}$. More specifically, denote $\frac{\partial^2}{\partial^2}$ by $H$, we have

$$J\delta(\beta) = E[g(X(1), D(1), \beta) | X(0) = z, D(0) = 0],$$

where $g(x, y, \beta) = g_x(x, \beta) y + g_{\beta}(x, \beta).$  

96
PROOF OF PROPOSITION 2. Without loss of generality, we assume that

\[ f_t(X, \beta) = g(X_t, X_{t-1}, \beta) - E^\theta [g(X_t, X_{t-1}, \beta)|X_{t-1}], \]

where \( g \) satisfies a polynomial growth condition, i.e. there exist constants \( K, \gamma > 0 \), such that

\[ |g(X_t, X_{t-1}, \beta)| \leq K(1 + |X_{t-1}|^\gamma + |X_t|^\gamma). \]

Following Proposition 1, and Lemma 1 of Essay Two, we know that there exist constants \( C', \gamma' > 0 \), independent of \( n \) and \( \beta \), such that

\[ |f_t(X, \beta) - f_t^n(X, \beta)| \leq \frac{C'}{n}(1 + |X_{t-1}|^{\gamma'}), \quad (23) \]

\[ |E^\theta [g(X_t, X_{t-1}, \beta)|X_{t-1}]| \leq C'(1 + |X_{t-1}|^{\gamma'}), \quad (24) \]

\[ |E_n^\theta [g(X^n_t, X_{t-1}, \beta)|X^n_{t-1}]| \leq C'(1 + |X_{t-1}|^{\gamma'}). \quad (25) \]

Although Proposition 1 and Lemma 1 may give us different constants \( C' \) and \( \gamma' \) for (23), (24) and (25), we can always find \( C', \gamma' \), large enough, such that all three equations are satisfied. Equation (23) implies that

\[ |R(\beta) - R^n(\beta)| \leq \frac{C'}{n}(1 + E|X_{t-1}|^{\gamma'}). \]

Equation (24) and (25) imply that there exist constants \( K', \tilde{\gamma} > 0 \), large enough, such that

\[ |f_t(X, \beta)| \leq K'(1 + |X_t|^{\tilde{\gamma}} + |X_{t-1}|^{\tilde{\gamma}}), \]

\[ |f_t^n(X, \beta)| \leq K'(1 + |X_t|^{\tilde{\gamma}} + |X_{t-1}|^{\tilde{\gamma}}). \]

Now, consider the following,

\[ [H(\beta) - H^n(\beta)] \leq |(E f_t(X, \beta))^TW_0(E f_t(X, \beta)) - (E f_t^n(X, \beta))^TW^n(E f_t^n(X, \beta))| \]

\[ \leq E|f_t(X, \beta) - f_t^n(X, \beta)| || W_0 || E|f_t(X, \beta)| \]

\[ + E|f_t^n(X, \beta)| || W_0 - W^n || E|f_t(X, \beta)| \]

\[ + E|f_t^n(X, \beta)| || W^n || E|f_t^n(X, \beta) - f_t(X, \beta)| \]

\[ \leq \frac{C'K'}{n}E(1 + |X_{t-1}|^{\gamma'})E(1 + |X_{t-1}|^{\tilde{\gamma}} + |X_t|^{\tilde{\gamma}}) || W_0 || \]

\[ + K'^2E(1 + |X_{t-1}|^{\tilde{\gamma}} + |X_t|^{\tilde{\gamma}})^2 || W_0 - W^n || \]

\[ + \frac{C'K'}{n}E(1 + |X_{t-1}|^{\gamma'})E(1 + |X_{t-1}|^{\tilde{\gamma}} + |X_t|^{\tilde{\gamma}}) || W^n || . \]

Since \( E|X_{t-1}|^{\tilde{\gamma}} = E|X_t|^{\tilde{\gamma}} = E|Y|^{\tilde{\gamma}} < \infty \), and \( E|X_{t-1}|^{\gamma'} = E|Y|^{\gamma'} < \infty \), \( R^n \) and \( H^n \) converge to \( R \) and \( H \) uniformly on \( \Theta \).

Similar arguments can be used to prove that \( J^n \) converges to \( J \) uniformly on \( \Theta \). \( \blacksquare \)
Table 1: (a) Finite sample properties of $\hat{\theta}_T$ for $T = 520.$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\tilde{\text{Var}}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail&lt;sup&gt;a&lt;/sup&gt; (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>0.80</td>
<td>1.301 (0.623)</td>
<td>0.215 (0.464)</td>
<td>0.2806 (0.132)</td>
<td>0.174 (25.62)</td>
<td>0.212 (16.62)</td>
<td>0.268 (12.52)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.10</td>
<td>0.1001 (0.311e-2)</td>
<td>0.954e-5 (0.309e-2)</td>
<td>0.989e-5 (0.119e-5)</td>
<td>0.008 (-0.449)</td>
<td>0.054 (0.410)</td>
<td>0.106 (0.447)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.0994 (0.126e-1)</td>
<td>0.162e-3 (0.127e-1)</td>
<td>0.201e-3 (0.101e-2)</td>
<td>0.014 (0.899)</td>
<td>0.034 (-1.641)</td>
<td>0.068 (-2.381)</td>
</tr>
</tbody>
</table>

Table 1: (b) Finite sample properties of $\hat{\beta}_T$ for $T = 780.$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\tilde{\text{Var}}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail&lt;sup&gt;a&lt;/sup&gt; (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>0.80</td>
<td>1.102 (0.428)</td>
<td>0.143 (0.379)</td>
<td>0.187 (0.884e-1)</td>
<td>0.080 (15.73)</td>
<td>0.150 (10.26)</td>
<td>0.200 (7.453)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.10</td>
<td>0.1000 (0.241e-2)</td>
<td>0.636e-5 (0.252e-2)</td>
<td>0.659e-5 (0.797e-6)</td>
<td>0.000 (-2.247)</td>
<td>0.038 (-1.231)</td>
<td>0.106 (0.447)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.1003 (0.107e-1)</td>
<td>0.108e-3 (0.104e-1)</td>
<td>0.133e-3 (0.671e-3)</td>
<td>0.016 (1.348)</td>
<td>0.048 (-0.205)</td>
<td>0.106 (0.447)</td>
</tr>
</tbody>
</table>

Table 1: (c) Finite sample properties of $\hat{\beta}_T$ for $T = 1040.$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\tilde{\text{Var}}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail&lt;sup&gt;a&lt;/sup&gt; (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>0.80</td>
<td>0.993 (0.343)</td>
<td>0.107 (0.328)</td>
<td>0.108 (0.377e-1)</td>
<td>0.034 (5.393)</td>
<td>0.102 (5.334)</td>
<td>0.150 (3.727)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.10</td>
<td>0.0999 (0.215e-2)</td>
<td>0.477e-5 (0.218e-2)</td>
<td>0.492e-5 (0.423e-6)</td>
<td>0.010 (0.000)</td>
<td>0.044 (0.000)</td>
<td>0.092 (0.000)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.1003 (0.951e-2)</td>
<td>0.809e-4 (0.899e-2)</td>
<td>0.753e-4 (0.787e-4)</td>
<td>0.014 (0.897)</td>
<td>0.068 (1.846)</td>
<td>0.116 (1.192)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Asymptotic t-statistics for the hypothesis that the true proportion is $p$ is calculated as $(\hat{p} - p)/[p(1-p)/N]^{1/2}$, where $N = 500$ is the number of replications and $\hat{p}$ is the estimated proportion.
Table 2: (a) Finite sample properties of $\hat{\beta}_T$ for $T = 260$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>$\text{Var}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>$\overline{\text{Var}}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail$^a$ (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>4.00</td>
<td>4.919 (1.578)</td>
<td>1.994 (1.412)</td>
<td>2.185 (0.813)</td>
<td>0.06 (11.23)</td>
<td>0.122 (7.386)</td>
<td>0.178 (5.813)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.10</td>
<td>0.1003 (0.455e-2)</td>
<td>0.197e-4 (0.443e-2)</td>
<td>0.204e-4 (0.417e-5)</td>
<td>-2.247 (0.00)</td>
<td>0.06 (1.026)</td>
<td>0.116 (1.192)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.1000 (0.352e-2)</td>
<td>0.131e-4 (0.363e-2)</td>
<td>0.113e-4 (0.803e-5)</td>
<td>-0.899 (-0.615)</td>
<td>0.044 (0.092)</td>
<td>0.092 (1.520)</td>
</tr>
</tbody>
</table>

Table 2: (b) Finite sample properties of $\hat{\beta}_T$ for $T = 520$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>$\text{Var}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>$\overline{\text{Var}}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail$^a$ (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>4.00</td>
<td>4.274 (1.063)</td>
<td>0.997 (0.998)</td>
<td>0.993 (0.266)</td>
<td>0.024 (3.146)</td>
<td>0.068 (1.846)</td>
<td>0.132 (2.385)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.10</td>
<td>0.0999 (0.326e-2)</td>
<td>0.984e-5 (0.313e-2)</td>
<td>0.104e-4 (0.140e-5)</td>
<td>0.014 (0.899)</td>
<td>0.054 (0.410)</td>
<td>0.098 (-0.149)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.1000 (0.233e-2)</td>
<td>0.658e-5 (0.256e-2)</td>
<td>0.589e-5 (0.295e-5)</td>
<td>-1.348 (-1.846)</td>
<td>0.032 (0.078)</td>
<td>0.078 (-1.639)</td>
</tr>
</tbody>
</table>

Table 2: (c) Finite sample properties of $\hat{\beta}_T$ for $T = 780$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.error)</th>
<th>$\text{Var}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>$\overline{\text{Var}}(\hat{\beta}_T)$ (Std.error)</th>
<th>0.01 tail$^a$ (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>4.00</td>
<td>4.124 (0.842)</td>
<td>0.665 (0.815)</td>
<td>0.636 (0.143)</td>
<td>0.02 (2.247)</td>
<td>0.046 (1.040)</td>
<td>0.114 (0.28)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.10</td>
<td>0.0999 (0.258e-2)</td>
<td>0.656e-5 (0.256e-2)</td>
<td>0.693e-5 (0.765e-6)</td>
<td>0.008 (1.348)</td>
<td>0.042 (-0.821)</td>
<td>0.100 (0.00)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.1000 (0.204e-2)</td>
<td>0.439e-5 (0.209e-2)</td>
<td>0.397e-5 (0.156e-7)</td>
<td>0.004 (0.000)</td>
<td>0.050 (0.152)</td>
<td>0.102 (0.87)</td>
</tr>
</tbody>
</table>

$^a$ Asymptotic t-statistics for the hypothesis that the true proportion is $p$ is calculated as $(\hat{p} - p)/(p(1 - p)/N)^{1/2}$, where $N = 500$ is the number of replications and $\hat{p}$ is the estimated proportion.
Table 3: (a) Finite sample properties of $\hat{\beta}_T$ for $T = 260$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\text{Var}^*(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail$^a$ (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2$(19) (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>40.0</td>
<td>41.24 (6.104)</td>
<td>40.86 (6.392)</td>
<td>41.29 (13.892)</td>
<td>0.02</td>
<td>0.058</td>
<td>0.118</td>
<td>22.72</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.10</td>
<td>0.1004 (0.636e-2)</td>
<td>0.374e-4 (0.611e-2)</td>
<td>0.408e-4 (0.109e-4)</td>
<td>0.028</td>
<td>0.062</td>
<td>0.104</td>
<td>20.32</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.1000 (0.354e-3)</td>
<td>0.128e-6 (0.358e-3)</td>
<td>0.131e-6 (0.284e-7)</td>
<td>0.004</td>
<td>0.048</td>
<td>0.098</td>
<td>25.52</td>
</tr>
</tbody>
</table>

Table 3: (b) Finite sample properties of $\hat{\beta}_T$ for $T = 520$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\text{Var}^*(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail$^a$ (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2$(19) (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>40.0</td>
<td>40.47 (4.427)</td>
<td>20.42 (4.519)</td>
<td>19.75 (4.651)</td>
<td>0.012</td>
<td>0.046</td>
<td>0.088</td>
<td>16.16</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.10</td>
<td>0.1004 (0.447e-2)</td>
<td>0.187e-4 (0.432e-2)</td>
<td>0.200e-4 (0.371e-5)</td>
<td>0.022</td>
<td>0.058</td>
<td>0.098</td>
<td>10.40</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.1000 (0.253e-3)</td>
<td>0.641e-7 (0.253e-3)</td>
<td>0.661e-7 (0.104e-7)</td>
<td>0.014</td>
<td>0.058</td>
<td>0.106</td>
<td>20.64</td>
</tr>
</tbody>
</table>

Table 3: (c) Finite sample properties of $\hat{\beta}_T$ for $T = 780$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\text{Var}^*(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail$^a$ (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2$(19) (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>40.0</td>
<td>40.17 (3.699)</td>
<td>13.62 (3.691)</td>
<td>13.34 (2.676)</td>
<td>0.008</td>
<td>0.046</td>
<td>0.106</td>
<td>11.84</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.10</td>
<td>0.1007 (0.380e-2)</td>
<td>0.125e-4 (0.353e-2)</td>
<td>0.136e-4 (0.216e-5)</td>
<td>0.012</td>
<td>0.060</td>
<td>0.112</td>
<td>14.88</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.0999 (0.202e-3)</td>
<td>0.427e-7 (0.207e-3)</td>
<td>0.434e-7 (0.056e-7)</td>
<td>0.004</td>
<td>0.048</td>
<td>0.098</td>
<td>13.52</td>
</tr>
</tbody>
</table>

$^a$ Asymptotic $t$-statistics for the hypothesis that the true proportion is $p$ is calculated as $(\hat{p} - p)/[p(1 - p)/N]^{1/2}$, where $N = 500$ is the number of replications and $\hat{p}$ is the estimated proportion.
Table 4: (a) First moment approximations $F(\beta) = E^{\theta}[X_{t+1}|X_t]$.  

\[ \kappa = 0.8, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>n</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.8489e-1</td>
<td>0.6075e-2</td>
<td>0.000</td>
<td>0.1528e-1</td>
</tr>
<tr>
<td>10</td>
<td>0.8489e-1</td>
<td>0.6069e-2</td>
<td>0.000</td>
<td>0.1528e-1</td>
</tr>
<tr>
<td>12</td>
<td>0.8489e-1</td>
<td>0.6065e-2</td>
<td>0.000</td>
<td>0.1527e-1</td>
</tr>
<tr>
<td>14</td>
<td>0.8488e-1</td>
<td>0.6062e-2</td>
<td>0.000</td>
<td>0.1527e-1</td>
</tr>
<tr>
<td>16</td>
<td>0.8488e-1</td>
<td>0.6061e-2</td>
<td>0.000</td>
<td>0.1526e-1</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.8488e-1</td>
<td>0.6059e-2</td>
<td>0.000</td>
<td>0.1526e-1</td>
</tr>
</tbody>
</table>

Table 4: (b) Second moment approximations $F(\beta) = E^{\theta}[X_{t+1}^2|X_t]$.  

\[ \kappa = 0.8, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>n</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.7596e-2</td>
<td>0.1037e-2</td>
<td>0.1559e-2</td>
<td>0.2626e-2</td>
</tr>
<tr>
<td>10</td>
<td>0.7596e-2</td>
<td>0.1037e-2</td>
<td>0.1559e-2</td>
<td>0.2626e-2</td>
</tr>
<tr>
<td>12</td>
<td>0.7596e-2</td>
<td>0.1037e-2</td>
<td>0.1560e-2</td>
<td>0.2627e-2</td>
</tr>
<tr>
<td>14</td>
<td>0.7596e-2</td>
<td>0.1036e-2</td>
<td>0.1560e-2</td>
<td>0.2627e-2</td>
</tr>
<tr>
<td>16</td>
<td>0.7596e-2</td>
<td>0.1036e-2</td>
<td>0.1560e-2</td>
<td>0.2627e-2</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.7596e-2</td>
<td>0.1036e-2</td>
<td>0.1561e-2</td>
<td>0.2628e-2</td>
</tr>
</tbody>
</table>

Table 4: (c) Third moment approximations $F(\beta) = E^{\theta}[X_{t+1}^3|X_t]$.  

\[ \kappa = 0.8, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>n</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.7133e-3</td>
<td>0.1340e-3</td>
<td>0.4155e-3</td>
<td>0.3564e-3</td>
</tr>
<tr>
<td>10</td>
<td>0.7134e-3</td>
<td>0.1340e-3</td>
<td>0.4163e-3</td>
<td>0.3566e-3</td>
</tr>
<tr>
<td>12</td>
<td>0.7135e-3</td>
<td>0.1399e-3</td>
<td>0.4169e-3</td>
<td>0.3567e-3</td>
</tr>
<tr>
<td>14</td>
<td>0.7136e-3</td>
<td>0.1399e-3</td>
<td>0.4174e-3</td>
<td>0.3569e-3</td>
</tr>
<tr>
<td>16</td>
<td>0.7136e-3</td>
<td>0.1399e-3</td>
<td>0.4180e-3</td>
<td>0.3571e-3</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.7138e-3</td>
<td>0.1399e-3</td>
<td>0.4187e-3</td>
<td>0.3574e-3</td>
</tr>
</tbody>
</table>

Note: $n = \infty$ corresponds to the values and derivatives of the actual conditional moments.
Table 5: (a) First moment approximations $F(\beta) = E^{\beta}[X_{t+1}|X_t]$.

\[ \kappa = 4, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>$n$</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>8.1487e-2</td>
<td>0.3594e-3</td>
<td>0.0000</td>
<td>0.7436e-1</td>
</tr>
<tr>
<td>10</td>
<td>8.1486e-2</td>
<td>0.3588e-3</td>
<td>0.0000</td>
<td>0.7427e-1</td>
</tr>
<tr>
<td>12</td>
<td>8.1485e-2</td>
<td>0.3582e-3</td>
<td>0.0000</td>
<td>0.7421e-1</td>
</tr>
<tr>
<td>14</td>
<td>8.1485e-2</td>
<td>0.3577e-3</td>
<td>0.0000</td>
<td>0.7416e-1</td>
</tr>
<tr>
<td>16</td>
<td>8.1484e-2</td>
<td>0.3569e-3</td>
<td>0.0000</td>
<td>0.7411e-1</td>
</tr>
<tr>
<td>$\infty$</td>
<td>8.1481e-2</td>
<td>0.3561e-3</td>
<td>0.0000</td>
<td>0.7403e-1</td>
</tr>
</tbody>
</table>

Table 5: (b) Second moment approximations $F(\beta) = E^{\beta}[X_{t+1}^2|X_t]$.

\[ \kappa = 4, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>$n$</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.6654e-2</td>
<td>0.5834e-4</td>
<td>0.2902e-3</td>
<td>0.1213e-1</td>
</tr>
<tr>
<td>10</td>
<td>0.6654e-2</td>
<td>0.5823e-4</td>
<td>0.2898e-3</td>
<td>0.1213e-1</td>
</tr>
<tr>
<td>12</td>
<td>0.6654e-2</td>
<td>0.5810e-4</td>
<td>0.2895e-3</td>
<td>0.1211e-1</td>
</tr>
<tr>
<td>14</td>
<td>0.6653e-2</td>
<td>0.5802e-4</td>
<td>0.2892e-3</td>
<td>0.1210e-1</td>
</tr>
<tr>
<td>16</td>
<td>0.6653e-2</td>
<td>0.5796e-4</td>
<td>0.2891e-3</td>
<td>0.1209e-1</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.6653e-2</td>
<td>0.5780e-4</td>
<td>0.2879e-3</td>
<td>0.1207e-1</td>
</tr>
</tbody>
</table>

Table 5: (c) Third moment approximations $F(\beta) = E^{\beta}[X_{t+1}^3|X_t]$.

\[ \kappa = 4, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>$n$</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.5446e-3</td>
<td>0.7125e-5</td>
<td>0.7110e-4</td>
<td>0.1487e-1</td>
</tr>
<tr>
<td>10</td>
<td>0.5446e-3</td>
<td>0.7110e-5</td>
<td>0.7097e-4</td>
<td>0.1485e-1</td>
</tr>
<tr>
<td>12</td>
<td>0.5446e-3</td>
<td>0.7100e-5</td>
<td>0.7087e-4</td>
<td>0.1484e-1</td>
</tr>
<tr>
<td>14</td>
<td>0.5445e-3</td>
<td>0.7093e-5</td>
<td>0.7079e-4</td>
<td>0.1484e-1</td>
</tr>
<tr>
<td>16</td>
<td>0.5445e-3</td>
<td>0.7088e-5</td>
<td>0.7072e-4</td>
<td>0.1483e-1</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.5445e-3</td>
<td>0.7068e-5</td>
<td>0.7054e-4</td>
<td>0.1481e-1</td>
</tr>
</tbody>
</table>

Note: $n = \infty$ corresponds to the values and derivatives of the actual conditional moments.
Table 6: (a) First moment approximations $F(\beta) = E^{\theta}[X_{t+1}|X_t]$.

\[ \kappa = 40, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>$n$</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.9109e-1</td>
<td>0.1845e-3</td>
<td>0.000</td>
<td>0.5516</td>
</tr>
<tr>
<td>10</td>
<td>0.9096e-1</td>
<td>0.1812e-3</td>
<td>0.000</td>
<td>0.5448</td>
</tr>
<tr>
<td>12</td>
<td>0.9085e-1</td>
<td>0.1795e-3</td>
<td>0.000</td>
<td>0.5401</td>
</tr>
<tr>
<td>14</td>
<td>0.9079e-1</td>
<td>0.1791e-3</td>
<td>0.000</td>
<td>0.5387</td>
</tr>
<tr>
<td>16</td>
<td>0.9075e-1</td>
<td>0.1788e-3</td>
<td>0.000</td>
<td>0.5373</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.9073e-1</td>
<td>0.1782e-3</td>
<td>0.000</td>
<td>0.5366</td>
</tr>
</tbody>
</table>

Table 6: (b) Second moment approximations $F(\beta) = E^{\theta}[X_{t+1}^2|X_t]$.

\[ \kappa = 40, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>$n$</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.8301e-2</td>
<td>0.3415e-4</td>
<td>0.1819e-3</td>
<td>0.1010</td>
</tr>
<tr>
<td>10</td>
<td>0.8283e-2</td>
<td>0.3356e-4</td>
<td>0.1789e-3</td>
<td>0.0998</td>
</tr>
<tr>
<td>12</td>
<td>0.8264e-2</td>
<td>0.3304e-4</td>
<td>0.1771e-3</td>
<td>0.0990</td>
</tr>
<tr>
<td>14</td>
<td>0.8256e-2</td>
<td>0.3279e-4</td>
<td>0.1754e-3</td>
<td>0.0983</td>
</tr>
<tr>
<td>16</td>
<td>0.8251e-2</td>
<td>0.3231e-4</td>
<td>0.1741e-3</td>
<td>0.0979</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.8241e-2</td>
<td>0.3223e-4</td>
<td>0.1714e-3</td>
<td>0.0974</td>
</tr>
</tbody>
</table>

Table 6: (c) Third moment approximations $F(\beta) = E^{\theta}[X_{t+1}^3|X_t]$.

\[ \kappa = 40, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>$n$</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.7583e-3</td>
<td>0.4621e-5</td>
<td>0.4962e-4</td>
<td>0.1383e-1</td>
</tr>
<tr>
<td>10</td>
<td>0.7557e-3</td>
<td>0.4546e-5</td>
<td>0.4825e-4</td>
<td>0.1371e-1</td>
</tr>
<tr>
<td>12</td>
<td>0.7542e-3</td>
<td>0.4489e-5</td>
<td>0.4753e-4</td>
<td>0.1363e-1</td>
</tr>
<tr>
<td>14</td>
<td>0.7523e-3</td>
<td>0.4452e-5</td>
<td>0.4695e-4</td>
<td>0.1351e-1</td>
</tr>
<tr>
<td>16</td>
<td>0.7511e-3</td>
<td>0.4429e-5</td>
<td>0.4684e-4</td>
<td>0.1344e-1</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.7493e-3</td>
<td>0.4406e-5</td>
<td>0.4673e-4</td>
<td>0.1331e-1</td>
</tr>
</tbody>
</table>

Note: $n = \infty$ corresponds to the values and derivatives of the actual conditional moments.
Table 7: (a) Approximations for the objective function.

$T = 520, \kappa = 0.8, \sigma = 0.1, \text{ and } \theta = 0.1.$

<table>
<thead>
<tr>
<th>$n$</th>
<th>OJE</th>
<th>GRD(1)</th>
<th>GRD(2)</th>
<th>GRD(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.260921</td>
<td>0.33350e-1</td>
<td>-30.6410</td>
<td>-1.8424</td>
</tr>
<tr>
<td>10</td>
<td>0.260920</td>
<td>0.33347e-1</td>
<td>-30.6410</td>
<td>-1.8420</td>
</tr>
<tr>
<td>12</td>
<td>0.260920</td>
<td>0.33345e-1</td>
<td>-30.5410</td>
<td>-1.8417</td>
</tr>
<tr>
<td>14</td>
<td>0.260919</td>
<td>0.33343e-1</td>
<td>-30.6410</td>
<td>-1.8413</td>
</tr>
<tr>
<td>16</td>
<td>0.260919</td>
<td>0.33341e-1</td>
<td>-30.6410</td>
<td>-1.8409</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.260919</td>
<td>0.33338e-1</td>
<td>-30.6409</td>
<td>-1.8405</td>
</tr>
</tbody>
</table>

Table 7: (b) Approximations for the objective function.

$T = 520, \kappa = 4, \sigma = 0.1, \text{ and } \theta = 0.1.$

<table>
<thead>
<tr>
<th>$n$</th>
<th>OBJ</th>
<th>GRD(1)</th>
<th>GRD(2)</th>
<th>GRD(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.9002e-1</td>
<td>0.2137e-1</td>
<td>-15.3764</td>
<td>-4.7632</td>
</tr>
<tr>
<td>10</td>
<td>0.8998e-1</td>
<td>0.2133e-1</td>
<td>-15.3665</td>
<td>-4.7532</td>
</tr>
<tr>
<td>12</td>
<td>0.8994e-1</td>
<td>0.2130e-1</td>
<td>-15.3653</td>
<td>-4.7466</td>
</tr>
<tr>
<td>14</td>
<td>0.8992e-1</td>
<td>0.2128e-1</td>
<td>-15.3641</td>
<td>-4.7391</td>
</tr>
<tr>
<td>16</td>
<td>0.8991e-1</td>
<td>0.2126e-1</td>
<td>-15.3631</td>
<td>-4.7311</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.8987e-1</td>
<td>0.2118e-1</td>
<td>-15.3607</td>
<td>-4.7298</td>
</tr>
</tbody>
</table>

Table 7: (c) Approximations for the objective function.

$T = 520, \kappa = 40, \sigma = 0.1, \text{ and } \theta = 0.1.$

<table>
<thead>
<tr>
<th>$n$</th>
<th>OBJ</th>
<th>GRD(1)</th>
<th>GRD(2)</th>
<th>GRD(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.2024</td>
<td>0.2108e-1</td>
<td>-25.886</td>
<td>-100.351</td>
</tr>
<tr>
<td>10</td>
<td>0.2016</td>
<td>0.2088e-1</td>
<td>-25.821</td>
<td>-98.891</td>
</tr>
<tr>
<td>12</td>
<td>0.2011</td>
<td>0.2076e-1</td>
<td>-25.781</td>
<td>-97.746</td>
</tr>
<tr>
<td>14</td>
<td>0.2007</td>
<td>0.2063e-1</td>
<td>-25.756</td>
<td>-96.871</td>
</tr>
<tr>
<td>16</td>
<td>0.2005</td>
<td>0.2054e-1</td>
<td>-25.736</td>
<td>-95.981</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.1991</td>
<td>0.2019e-1</td>
<td>-25.613</td>
<td>-94.541</td>
</tr>
</tbody>
</table>

Note: $n = \infty$ corresponds to the values and derivatives of the actual objective functions.
Table 8: Comparisons between $\hat{\beta}_T$ and $\hat{\beta}^n_T$.

$\kappa = 4$, $\sigma = 0.1$, $\theta = 0.1$, $T = 520$, and $n = 12$

<table>
<thead>
<tr>
<th>$\hat{k}_T$</th>
<th>$\hat{k}^n_T$</th>
<th>$\hat{\sigma}_T$</th>
<th>$\hat{\sigma}^n_T$</th>
<th>$\hat{\beta}_T$</th>
<th>$\hat{\beta}^n_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6514925 (1.0211)</td>
<td>4.6514925 (1.0237)</td>
<td>0.1013149 (0.3475e-2)</td>
<td>0.0101348 (0.3480e-2)</td>
<td>0.1011897 (0.2212e-2)</td>
<td>0.1011866 (0.2198e-2)</td>
</tr>
<tr>
<td>4.1857486 (1.0514)</td>
<td>4.1857485 (1.0523)</td>
<td>0.0990128 (0.3075e-2)</td>
<td>0.0990119 (0.3077e-2)</td>
<td>0.1014964 (0.2388e-2)</td>
<td>0.1014965 (0.2375e-2)</td>
</tr>
<tr>
<td>4.1202013 (0.9060)</td>
<td>4.1202037 (0.9064)</td>
<td>0.1024504 (0.3053e-2)</td>
<td>0.1024501 (0.3056e-2)</td>
<td>0.1014151 (0.2426e-2)</td>
<td>0.1014178 (0.2414e-2)</td>
</tr>
<tr>
<td>4.4843505 (1.0332)</td>
<td>4.4843505 (1.0339)</td>
<td>0.1035081 (0.3268e-2)</td>
<td>0.1035085 (0.3271e-2)</td>
<td>0.0970444 (0.2280e-2)</td>
<td>0.0970441 (0.2267e-2)</td>
</tr>
<tr>
<td>3.9555679 (1.0234)</td>
<td>3.9555697 (1.0243)</td>
<td>0.0968098 (0.3375e-2)</td>
<td>0.0968098 (0.3379e-2)</td>
<td>0.0977091 (0.2414e-2)</td>
<td>0.0977099 (0.2402e-2)</td>
</tr>
<tr>
<td>3.8589386 (0.9804)</td>
<td>3.8589386 (0.9811)</td>
<td>0.1022397 (0.3325e-2)</td>
<td>0.1022398 (0.3327e-2)</td>
<td>0.1013215 (0.2702e-2)</td>
<td>0.1013174 (0.2689e-2)</td>
</tr>
<tr>
<td>3.9622223 (1.0068)</td>
<td>3.9622223 (1.0072)</td>
<td>0.1116794 (0.3679e-2)</td>
<td>0.1116777 (0.3682e-2)</td>
<td>0.0969922 (0.2773e-2)</td>
<td>0.0969948 (0.2758e-2)</td>
</tr>
<tr>
<td>4.6551534 (1.0248)</td>
<td>4.6551534 (1.0266)</td>
<td>0.1015327 (0.3441e-2)</td>
<td>0.1015341 (0.3445e-2)</td>
<td>0.1026368 (0.2208e-2)</td>
<td>0.1026367 (0.2194e-2)</td>
</tr>
<tr>
<td>4.0139144 (0.9541)</td>
<td>4.0139144 (0.9549)</td>
<td>0.0987532 (0.3095e-2)</td>
<td>0.0987530 (0.3098e-2)</td>
<td>0.0984243 (0.2432e-2)</td>
<td>0.0984240 (0.2419e-2)</td>
</tr>
<tr>
<td>4.7025483 (0.9669)</td>
<td>4.7025483 (0.9686)</td>
<td>0.1081610 (0.3581e-2)</td>
<td>0.1081625 (0.3585e-2)</td>
<td>0.1014039 (0.2298e-2)</td>
<td>0.1014028 (0.2287e-2)</td>
</tr>
</tbody>
</table>

Note: The numbers in parentheses are the corresponding estimated standard errors.