ESSAYS ON DYNAMIC MODELS IN FINANCIAL ECONOMICS

by

NEIL DAVID PEARSON

A. B., Economics
Princeton University (1981)

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

April 1990

Signature of Author____

Sloan School of Management
April 3, 1990

Certified by_____

John C. Cox
Nomura Professor of Finance
Thesis Supervisor

Accepted by_____

James B. Orlin
Professor of Management Science and Operations Research
Chair, Doctoral Program Committee

ESSAYS ON DYNAMIC MODELS IN FINANCIAL ECONOMICS

by

Neil D. Pearson

Submitted to the Sloan School of Management
on April 3, 1990 in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Management

ABSTRACT

This thesis consists of three independent essays on dynamic models in financial economics.

In the essay that comprises Chapter 1, I employ a martingale approach to study a consumption and portfolio problem in a discrete time model with dynamically incomplete markets and short-sale constraints. I show that the solution of the original dynamic problem is identical to the solution of a static problem of choosing among consumption bundles satisfying budget constraints formed using all Arrow-Debreu state prices consistent with no arbitrage. The budget feasible set can be defined by finitely many constraints even though infinitely many state price vectors are consistent with no arbitrage, and the approach is convenient for computation.

In the second essay I estimate the parameters of one and two-factor models of the term structure of interest rates due to Cox, Ingersoll, and Ross using the method of maximum likelihood, and present tests of the models. I recover the unobservable state variables and exploit their conditional density in estimation and testing, and use both the time-series and cross-sectional information in a sample of bonds that includes coupon bonds. I compare the basic CIR one-factor model of the term structure (interpreted as a model of nominal bond prices) to one of their two-factor models, and compare these to extended (translated) models. The tests reject the one-factor model in favor of its translated variant, and also reject the CIR one-factor models in favor of the translated two-factor model. I also nest these models within a more general Markov model for yields and reject the restrictions implied by the bond pricing models.

The third essay presents an extended simulated moments estimator (ESME) of the parameters of continuous time asset pricing models in which the underlying state variables follow an exogenously given diffusion process and the endogenously determined asset prices can be expressed as conditional expectations of known functions of the sample path of the state vector. The ESME allows the estimation of models in which the asset pricing function is neither known in closed form nor easily evaluated numerically. The idea underlying the estimation strategy is that, when asset prices can be written as conditional expectations, it is possible to simulate the moments of certain functions of asset prices even when it is not feasible to compute the endogenously determined asset prices as functions of the underlying state variables. I show the consistency and asymptotic normality of the estimator and demonstrate its computational feasibility by using it to estimate a one-factor term structure model due to Cox, Ingersoll, and Ross.

Thesis Supervisor: John C. Cox
Title: Nomura Professor of Finance

CHAPTER 1
CONSUMPTION AND PORTFOLIO POLICIES WITH INCOMPLETE MARKETS AND SHORT-SALE CONSTRAINTS: THE FINITE DIMENSIONAL CASE

Abstract In this essay, I employ a martingale approach to study a consumption-portfolio problem in a discrete time model with dynamically incomplete markets and short-sale constraints. In doing so I show how the approach is useful when markets are incomplete. The solution of the original dynamic problem is identical to the solution of a static problem of choosing among consumption bundles satisfying budget constraints formed using all Arrow-Debreu state prices consistent with no arbitrage. The budget feasible set can be defined by finitely many constraints even though infinitely many state price vectors are consistent with no arbitrage, and the approach is convenient for computation.

1 Introduction

The classical approach to the analysis of optimal intertemporal consumption and portfolio policies is stochastic dynamic programming, the use of which in this context was pioneered by Mossin (1968), Samuelson (1969), and Hakansson (1970) in discrete time and Merton (1969, 1971) in continuous time. Recently Pliska (1982, 1986), Cox and Huang (1987a, 1987b), and Pagès (1987) have used a martingale representation technology instead of dynamic programming to study optimal intertemporal consumption and portfolio policies, while Chamberlain (1988), Duffie and Huang (1985), and Huang (1987) have used it in a general equilibrium setting. However, to date little is known about how the martingale approach may be useful when markets are dynamically incomplete. With the exception of the paper by Pagès, those mentioned above assume that markets are 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 this essay I use the martingale approach to study optimal intertemporal consumption and portfolio policies in a general discrete time, discrete state space "finite dimensional" economy with dynamically incomplete markets and short-sale constraints. A dynamically incomplete (henceforth simply "incomplete") market is one in which not all contingent claims can be created by dynamic trading in the existing securities. I show how the original dynamic problem can be reformulated as a static problem which may then be attacked with the Lagrangian theory. The static formulation is useful for computation, has a natural economic interpretation, and has a geometry which makes it easy to understand the effect of incomplete markets. While the results of course apply only to the finite dimensional case, I impose no additional assumptions on either the asset price process or the nature of the incompleteness of markets.

In the martingale approach one solves the consumption and portfolio problem by separating it into two parts. First one identifies the set of attainable consumption bundles and solves a static optimization problem in order to select the consumption bundle most preferred by the consumer-investor. Then one determines the trading strategy needed to generate the most preferred consumption bundle. Pliska (1982) explicitly carries out these computations for a simple example in a discrete time finite probability space in which the agent consumes only at a single terminal date. Pliska (1986) extends the analysis to a more general continuous time stochastic enviroment (though again with consumption only at a single terminal date) in which asset prices are semi-martingales and consumption can be either positive or negative.

Cox and Huang (1987a) allow intermediate consumption and provide an easily verifiable set of sufficient conditions for existence weaker than those required in dynamic programming when asset prices follow a diffusion process. They find the unique set of Arrow-Debreu state prices (or, after normalization by the bond price, the unique equivalent martingale measure) that is consistent with the absence of arbitrage, and show that the solution of a static problem of maximizing utility subject to a single budget constraint formed using this set of state prices and the portfolio policies needed to implement it are identical to the optimal consumption and portfolio policies given by dynamic programming. Cox and Huang (1987b) characterize the optimal policies and compute explicitly the optimal consumption and portfolio policies in certain situations in which it is difficult if not impossible to use dynamic programming. An appealling feature of their approach is that the Lagrangian theory may be used to study the static problem.

While Cox and Huang only consider the continuous time case, the use of the martingale approach in the discrete time finite dimensional case with complete markets follows immediately from their analysis. Just as in the continuous time case, one finds the Arrow-Debreu state prices implied by the asset prices, solves a static problem of maximizing utility subject to a budget constraint formed using the state prices, and then implements the solution of the static problem.

A limitation of the analyses of Pliska and Cox and Huang is that these authors assume that markets are complete. In the papers by Cox and Huang the completeness of markets gives a unique set of Arrow-Debreu state prices or a unique equivalent martingale measure that is used to form the budget constraint in the static problem that comprises the first part of the martingale approach. When markets are incomplete there are infinitely many sets of state prices or equivalent martingale measures that are consistent with the absence of arbitrage, and the static problem of maximizing utility subject to the requirement that consumption be feasible with respect to budget constraints formed using all state prices consistent with no arbitrage involves infinitely many budget constraints. Hence it is not immediately obvious that the Lagrangian theory can be used to attack the problem.

The problem I study is that of a consumer-investor who selects his most preferred consumption bundle from among those that can be generated by a dynamic trading strategy using his endowment and the available securities. I first characterize the set of Arrow-Debreu state prices that are consistent with the given securities prices and the absence of arbitrage, and then show that the set of consumption bundles that can be generated by a dynamic trading strategy is identical to the set of consumption bundles that are budget feasible with respect to all of the state prices consistent with no arbitrage. Therefore the solution of the agent's original problem is identical to the solution of a static problem of maximizing utility subject to the constraint that the consumption bundle lies in the budget feasible set.

This static problem involves infinitely many budget constraints. I show that the set of feasible consumption bundles is generated by a finite number of budget constraints corresponding to the extreme points of the closure of the set of state prices consistent with no arbitrage. That is, the budget feasible set can be defined by a *finite* number of constraints even when markets are incomplete and there are *infinitely* many state price vectors consistent with no arbitrage. Hence the solution of the consumer-investor's problem can be easily characterized using the Lagrangian theory and numerical solutions can be computed using standard nonlinear programming techniques. The set of feasible consumption bundles has a nice geometry and a pleasing interpretation in terms of prices. I also extend the results to include short-sale constraints.

If the agent's preferences have an expected utility representation the optimal consumption plan may be obtained using dynamic programming. In this case the agent's original program and the static problem I obtain have identical solutions. An advantage of the static formulation vis-a-vis dynamic programming is that solving the static problem requires significantly less computation.

In the finite dimensional setting, a static formulation of Breeden (1987) represents another alternative to the martingale approach. The relationship between consumption and the trading strategies (shares of securities held in various states) is given by a set of linear equations. Breeden uses these to eliminate consumption from the problem and obtains a static problem in which maximization is performed with respect to the shares of the securities held. Breeden's approach has the same computational advantages vis-a-vis dynamic programming as the approach in this essay. However, the approach here provides an interpretation of the feasible consumption set in terms of the prices of consumption in the various states that Breeden's does not. The formulation in this essay is also more convenient for computation with the utility functions commonly assumed in finance. In addition, the characterization of the solution of the general problem makes it easy to obtain closed form solutions for the optimal consumption policy in the case of an agent with a time-additive, state-independent utility function of the hyperbolic absolute risk aversion (HARA) class (with no nonnegativity restrictions on consumption) when the investment opportunity set is constant and there are no short-sale constraints (see Hakansson (1970) for these solutions obtained via dynamic programming).

The balance of the essay is organized as follows. Section 2 presents an example with which I introduce the main ideas of the essay and illustrate the geometry of the approach. Section 3 describes the economy and its vector space representation and introduces most of the notation. I exploit the vector space representation to show that the set of consumption bundles that can be generated by a dynamic trading strategy using the agent's endowment and the available securities is identical to the set of consumption bundles that are budget feasible with respect to all of the state prices that are consistent with the absence of arbitrage, and show that the budget feasible set can be defined by a finite number of constraints. This program is carried out in Section 4. In Section 5 I present a simple characterization of the solution of the agent's problem, indicate how to implement the optimal consumption policy using a dynamic trading strategy, and discuss the advantages of the approach for computations. Section 6 consists of an example which illustrates the approach. Section 7 contains a few concluding remarks.

2 An Example

I consider an economy with one consumption good, only two dates $t = 0, 1$ and three possible outcomes $\{\omega_1, \omega_2, \omega_3\}$ at time 1. In order to be able to graph the feasible consumption set I require that the good be consumed only at time 1, and use $c = (c_1, c_2, c_3)'$ to denote consumption in the three states.

As a point of departure, I first suppose that there are three securities available for trading, so markets are complete. These securities have current prices $S_0^1 = 1$, $S_0^2 = 2$, and $S_0^3 = 3$, and state contingent payoffs at time 1 of $S_1^1 = (1, 1, 1)$, $S_1^2 = (3, 2, 1)$, and $S_1^3 = (1, 3, 5)$. The given securities prices imply that the Arrow-Debreu state prices are $(1/3, 1/3, 1/3)$. If I require that consumption be nonnegative and let the agent have an endowment of 1 at time 0, the feasible consumption set is

$$\{c \in \Re^3 \mid c \geq 0, 1/3c_1 + 1/3c_2 + 1/3c_3 \leq 1\}.$$

This set is shown in Figure 1.

Now suppose that the third security is not available for trading. When only the first two securities are available for trading, the feasible consumption set is defined by the inequalities

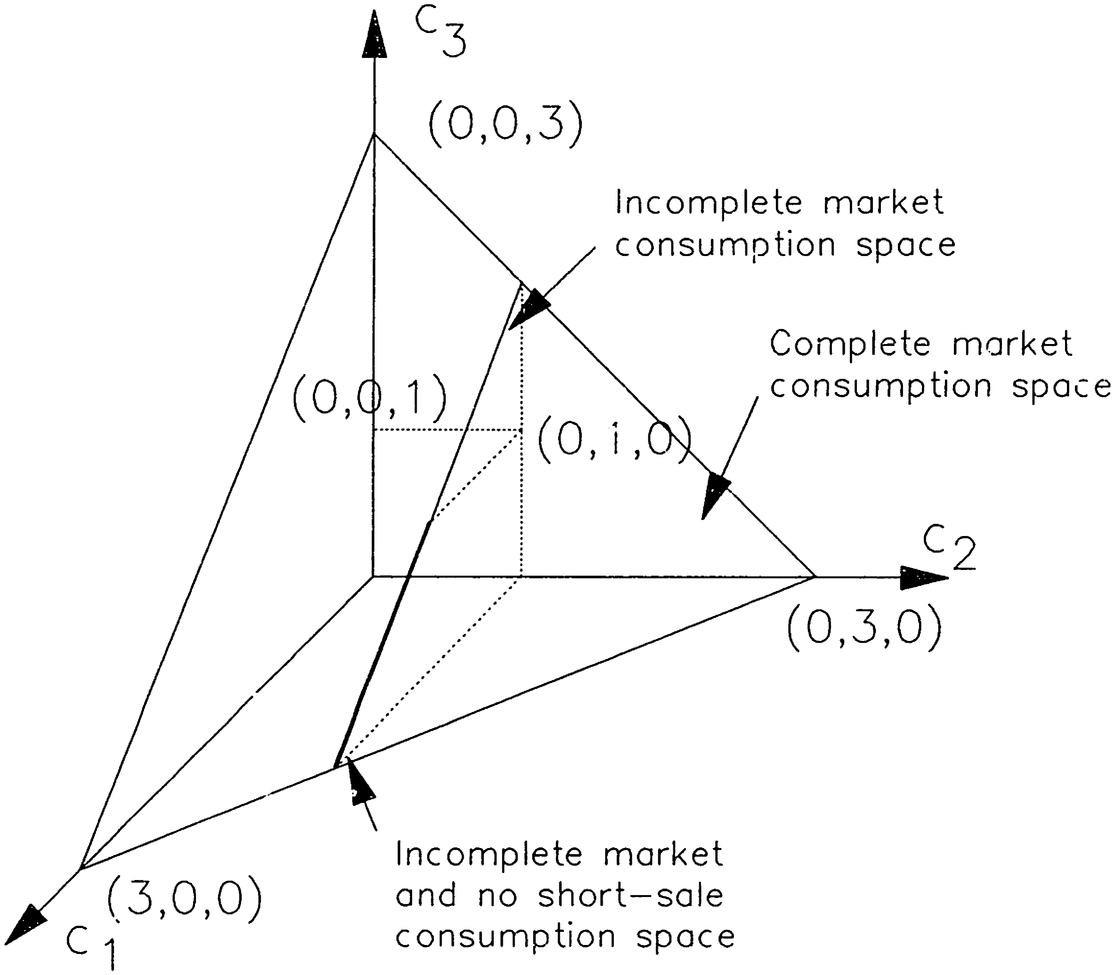


Figure 1
Graphical representation of the feasible consumption sets
The complete markets feasible consumption set is the polyhedron consisting of the convex hull of (0,0,0), (3,0,0), (0,3,0), and (0,0,3); the bundles a nonsatiated agent might choose lie on the face consisting of the convex hull of (3,0,0), (0,3,0), and (0,0,3). If the third security is not available for trading the bundles a nonsatiated agent might choose lie on the line segment connecting (0,1,2) and (2,1,0). If the third security is not available for trading and the second security may not be sold short the bundles a nonsatiated agent might choose lie in the segment connecting (1,1,1) and (2,1,0).

$$c \leq \varphi_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \varphi_2 \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}, \tag{1}$$

$$\varphi_1 + 2\varphi_2 \leq 1, \tag{2}$$

where φ_1 and φ_2 denote the number of shares of the two securities. The first inequality says essentially that the agent's consumption must be spanned by the payoffs of the two securities,[1] and the second inequality is the budget constraint.

Equations (1) and (2) imply

$$c_2 \leq 1, \tag{3}$$

$$1/2c_1 + 1/2c_3 \leq 1, \tag{4}$$

so that when only the two securities are available for trading the feasible consumption set is

$$\{c \in \Re^3 \mid c \geq 0, c_2 \leq 1, 1/2c_1 + 1/2c_3 \leq 1\}.$$

The bundles a nonsatiated agent might choose lie in the intersection of the two planes $c_2 = 1$, $1/2c_1 + 1/2c_3 = 1$, or in the segment connecting $(0,1,2)$ and $(2,1,0)$. This segment is also shown in Figure 1. It is a subset of the feasible consumption set for the complete markets case.

The "1" on the right hand side of the two inequalities (3) and (4) is the initial wealth. The coefficient "1" on c_2 in the inequality $c_2 \leq 1$ and the coefficients "1/2" and "1/2" on c_1 and c_3 in the inequality $1/2c_1 + 1/2c_3 \leq 1$ have a natural economic interpretation. These coefficients can be thought of as the prices of consumption in the two states. To see this, note that the least costly way to obtain an additional unit of consumption in the state ω_1 without decreasing consumption in any state is to take a long position of 1/2 unit of the second security and a short position of 1/2 unit of the first. The cost of this portfolio is 1/2. The least costly way to obtain an additional unit of consumption in the state ω_3 without decreasing consumption in any state is to take a long position of 3/2 units of the first security and a short position of 1/2 unit of the second security. The cost of this position is 1/2. Both of these positions yield a positive amount in the second state, which is why the coefficient on c_2 is zero in the inequality $1/2c_1 + 1/2c_3 \leq 1$. That is, the two portfolio strategies that generate consumption in the first and third states also generate consumption in the second state, so if one is undertaking either of these strategies there is a range of state ω_2 consumption for which the cost of consumption in that state is zero.

[1] The inequality appears because I have not yet excluded the possibility that the agent becomes satiated; I do this below.

An additional unit of consumption in state ω_2 can be obtained either by purchasing 1 unit of the first security or 1/2 unit of the second security. Both of these strategies have a cost of one, and both also yield positive amounts in the first and third states.

It turns out that I can make precise the interpretation of the coefficients of the two inequalities as prices. When markets are incomplete there is not a unique state price vector. Rather, there are infinitely many state price vectors consistent with the absence of arbitrage. In the example, the state prices $\pi = (\pi_1, \pi_2, \pi_3)$ consistent with the absence of arbitrage satisfy

$$\begin{aligned} \pi_1 + \pi_2 + \pi_3 &= 1, && (5)\\ 3\pi_1 + 2\pi_2 + \pi_3 &= 2, && (6)\\ \pi_1 &> 0, && (7)\\ \pi_2 &> 0, && (8)\\ \pi_3 &> 0, && (9) \end{aligned}$$

or

$$\begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} = \alpha \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \end{pmatrix},$$

where $0 < \alpha < 1$.

In the complete markets case one defines the set of feasible consumption bundles by forming a budget constraint using the single state price vector. Reasoning by analogy with the complete markets case, one might conjecture that when markets are incomplete one would define the feasible set by forming budget constraints using all of the (infinitely many) state price vectors consistent with the absence of arbitrage. This is a correct definition of the feasible set, but most of the constraints are redundant. One can define the feasible set using only the extreme points of the closure of the set of state price vectors consistent with the absence of arbitrage. These extreme points are $(0,1,0)$ and $(1/2,0,1/2)$. The components of these vectors are the coefficients on c_1, c_2, and c_3 in the inequalities (3) and (4).

This result provides a natural interpretation of the feasible consumption set in terms of prices and budget constraints. The agent's consumption in the second state, for example, is constrained by the fact the he must sacrifice one unit of his initial wealth in order to increase his state ω_2 consumption by one unit. This interpretation makes it much easier to understand the feasible consumption set. In addition, it is easy to see exactly how closing markets affects the feasible set. In the example, disallowing trading in the third security increased the dimension of the linear span of the set of state prices consistent with no arbitrage from one to two and added another plane defining the feasible consumption set.

The extreme points are relatively easy to calculate even for large problems. Given the extreme points, the agent's problem is simply to maximize utility subject to budget constraints formed using the extreme points. The problem is almost identical to the complete markets case, the only difference being that there is more than one budget constraint when markets are incomplete.

The approach also works when there are short-sale constraints. If I now suppose that the second security may not be sold short, the set of feasible consumption bundles is defined by the inequalities

$$c \leq \varphi_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \varphi_2 \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix},$$

$$\begin{aligned} \varphi_1 + 2\varphi_2 &\leq 1, \\ \varphi_2 &\geq 0. \end{aligned}$$

I obtain

$$\begin{aligned} c_2 &\leq 1, \\ 1/2c_1 + 1/2c_3 &\leq 1, \\ c_3 &\leq 1. \end{aligned}$$

The last of these inequalities comes from the short-sale constraint. The triple $(0,0,1)$, the components of which are the coefficients on c_1, c_2, and c_3 in the last inequality, is also an extreme point. The bundles a nonsatiated agent might choose lie in the segment connecting $(1,1,1)$ and $(2,1,0)$. I defer discussion of the details of short-sale constraints and simply observe that $(0,0,1)$, $(0,1,0)$ and $(1/2,0,1/2)$ are the extreme points of the set of state prices consistent with no arbitrage when the second security may not be sold short. The triples $(0,1,0)$ and $(1/2,0,1/2)$ satisfy (5)–(9) above, while $(0,0,1)$ satisfies $3\pi_1 + 2\pi_2 + \pi_3 \leq 2$ along with (5) and (7)–(9). Although the role of these relations will not be clear until Section 4, I note that the securities prices are supermartingales with respect to the measure $\pi = (0,0,1)$ and martingales with respect to $(0,1,0)$ and $(1/2,0,1/2)$.

3 The Finite Market Economy

I consider the following model of an agent's intertemporal consumption and portfolio policy under uncertainty. There is a finite number N of states of the world, indexed by $\omega \in \Omega$. There is a finite number of time periods, indexed by $t = 0, 1, 2, ..., T$. The information in the economy is exogenously specified and is represented by a sequence of partitions of Ω, $\{F_t; t = 0, \ldots, T\}$. The interpretation is that at time t the agent knows which cell of F_t contains the true state. Information increases through time; F_{t+1} is at least as fine as F_t. Without loss of generality I assume F_0 is trivial and F_T is the discrete partition (i.e., $F_0 = \Omega$ and $F_T = \{\omega | \omega \in \Omega\}$). The σ-field of events generated by F_t is denoted $\mathcal{F}_t$, and the family of σ-fields $\mathbf{F} = \{\mathcal{F}_t; t = 0, 1, \ldots, T\}$ is called the filtration generated by the sequence of partitions F_t. This information structure can be easily and intuitively represented by an event tree, and I make use of this representation when it is convenient to do so.

There is a single consumption good which the agent consumes at each date and cannot store. The agent's consumption can depend only upon the information he has at time t. I formalize this by defining a consumption bundle to be a stochastic process $c = \{c_t;\ t = 0, 1, ..., T\}$ that is adapted to $\mathbf{F}$; this means that c_t is measurable with respect to $\mathcal{F}_t$. Similarly, the agent receives a wage income stream which is an $\mathbf{F}$-adapted stochastic process $y = \{y_t;\ t = 0, 1, ..., T\}$.

At each date there are $K + 1$ securities available for trading in frictionless markets, with $K \geq 0$. It is convenient, but certainly not necessary, to assume that all $K + 1$ securities are available for trading at each date. Only trivial modifications are needed if some securities are not available for trading at some dates, and I indicate them below. I use S_t^k to denote the price of security k at time t, and use $S = \{S_t;\ t = 0, 1, ..., T\}$, where $S_t = (S_t^0, S_t^1, \ldots, S_t^K)'$, to denote the securities price process. For convenience I assume that the securities pay no dividends; this affects nothing. I assume that the securities price process S is adapted to $\mathbf{F}$. I make no further assumptions on the price system besides requiring that $|S_t^k| < +\infty$ for $t = 0, 1, ..., T$ and $k = 0, 1, 2, ..., K$. In particular, I require neither that one of the securities be a riskless bond nor that there are sufficiently many linearly independent securities that markets are (dynamically) complete.

The agent's problem is to manage a portfolio of these $K + 1$ securities in order to obtain his most preferred consumption plan. I use a vector space representation to formalize this as follows.

Let $N_t = |F_t|$, the number of cells (or nodes or events) in partition F_t at date t. Clearly $N_0 = 1$ and $N_T = N$. If I let $L = N_0 + N_1 + ... + N_T$, then L is the total number of cells from time 0 to time T. It is clear that any $\mathbf{F}$-adapted stochastic process can be characterized by its values at these L cells. This allows one to represent any $\mathbf{F}$-adapted process by an L–dimensional Euclidean vector space $\Re^L$ with its coordinates properly defined. I allocate the first coordinate to date 0's cell, the next N_1 coordinates to date 1's cells, the next N_2 coordinates to date 2's cells, etc.

Now one can see that any consumption bundle can be viewed as a vector in $\Re^L$ with each coordinate representing the consumption in some cell of some partition (in some state at some date). Similarly, a wage income stream can also be viewed as a vector in $\Re^L$. From now on I use $c, y \in \Re^L$ to denote consumption bundles and income streams respectively.

A trading strategy is a predictable $K+1$ dimensional process $\varphi = \{\varphi_t;\ t = 1, ..., T\}$, where $\varphi_t = (\varphi_t^0, \varphi_t^1, ..., \varphi_t^K)'$. The components of the trading strategy φ_t^k can be interpreted as the number of shares of security k held by the investor between $t-1$ and t. The assumption that the trading strategy is predictable means that φ_t is measurable with respect to $\mathcal{F}_{t-1}$. This is the natural information constraint.

It is possible to model the trading strategy as a vector in a Euclidean space in a fashion similar to the way I handled the consumption bundles and wage income streams. Since at each time $t < T$ the $K+1$ components of the trading strategy $\varphi_t^0, \varphi_t^1, ..., \varphi_t^K$ have to be determined at each of the N_t cells of partition F_t, the total number of values to be determined is $M = (K+1)(N_0 + N_1 + ... + N_{T-1})$. Thus a trading strategy can be represented by a vector $\varphi \in \Re^M$. Each of the coordinates of φ represents a long or short position in one of the securities at some date $t < T$ and subset of Ω. I allocate the first $K+1$ coordinates to be the investment strategies a date 0, the next $K+1$ coordinates to the investment strategies at the first cell of F_1 at date 1, the next $K+1$ to be the strategies at the second cell of F_1, etc.

The trading strategies and payoffs are related through a payoff matrix $X \in \Re^{L \times M}$. For $t = 0, 1, \ldots, T-1$, think of each security at each date and cell of F_t as a distinct investment opportunity. From the discussion above, there are $M = (K+1)(N_0 + N_1 + \cdots + N_{T-1})$ such opportunities. I represent each such investment opportunity as a vector in $\Re^L$, where I recall that $L = N_0 + N_1 + \cdots + N_T$, so that a vector in $\Re^L$ can represent the payoff of an investment in each cell of F_t for each date $t = 0, 1, \ldots, T$. As with c and y, I allocate the first coordinate to date 0's cell, the next N_1 coordinates to date 1's cells, the next N_2 coordinates to date 2's cells, etc.

I use $x^m \in \Re^L$ to denote the vector corresponding to the m-th investment opportunity and use $x^m(A_t)$ to denote the element of x^m corresponding to $A_t \in F_t$. Suppose the m-th investment opportunity x^m is the opportunity to invest in the k-th security in the event A_t. Following Breeden (1987), I represent this opportunity by setting $x^m(A_t) = -S_t^k(A_t)$, setting $x^m(A_{t+1}) = S_{t+1}^k(A_{t+1})$ if $A_{t+1} \subset A_t$, and setting $x^m(A_{t+1}) = 0$ if $A_{t+1} \not\subset A_t$. All other elements of x^m (i.e., those corresponding to other dates) are set equal to zero. For a common stock, the vector representing an investment opportunity will have one negative element corresponding to the cost of buying the stock at a particular date and state and a number of positive elements corresponding to the payoffs from closing out the position next period. The matrix X is formed by adjoining the M column vectors x^m, with the m-th column of X corresponding to the m-th investment opportunity.

Each component of the vector φ also corresponds to one investment opportunity. The consumption bundle produced by a trading strategy φ given a matrix X and income vector y is $c = X\varphi + y$. If certain securities are not available for trading in certain events at certain dates, there is no vector x^m corresponding to the missing investment opportunities and I form the matrix X by adjoining $M < (K+1)(N_0 + N_1 + \cdots + N_{T-1})$ columns x^m. Similarly, the vector φ would contain no component corresponding to the missing investment opportunity.

I use $C(y)$ to denote the set of feasible consumption bundles that can be generated by a dynamic trading strategy using income y. That is,

$$C(y) = \{c \in \Re^L \mid c \geq 0,\ \exists \varphi \in \Re^M \text{ s.t. } c = X\varphi + y\}. \tag{10}$$

The equation $c = X\varphi + y$ appears with an equality instead of an inequality because I assume below that the agent can never be satiated and therefore will never choose a consumption bundle such that $c \leq X\varphi + y$ and $c \neq X\varphi + y$.[2] Since φ represents a dynamic trading strategy, it is easy to see that $C(y)$ is identical to the set of feasible consumption bundles in a dynamic programming formulation of the consumption and portfolio problem. Were all of the elements of y save the first zero the condition $c = X\varphi + y$ would be the usual self-financing constraint. The agent chooses among various consumption bundles c.

Specifically, I consider the consumer-investor's problem

$$\max_{c \in C(y)} u(c, p), \tag{11}$$

where p is an $L-$vector of the probabilities of the various states and $u(c,p)$ represents a non-decreasing utility function. I assume that the agent's preferences are time consistent in the sense that a consumption plan that is initially optimal remains optimal at all future dates and states. Weller (1978) shows that a sufficient condition for time consistency is that the agent's preferences have an expected utility representation with respect to the uncertainty remaining at every date and state, but this condition is stronger than is necessary [Donaldson, Rossman, and Selden (1980), Johnsen and Donaldson (1985)]. Time additive functions $u(c,p) = u(c_1) + \sum_{i=2}^{L} p_i u(c_i)$ are of course included in the class of utility functions I consider, but certainly do not exhaust the class.

An example will make the formulation clear. Consider a three date economy with two securities and the following partitions of Ω: $F_0 = \Omega$, $F_1 = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$, and $F_2 = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\}$. One of the securities (security 0) is riskless and always has a price of 1. The other (security 1) is risky and has the following prices in the various cells of the

[2] The appendix also considers the case where the agent may reach satiation and the set of feasible consumption bundles is defined $C(y) = \{c \in \Re^L \mid c \geq 0,\ \exists \varphi \in \Re^M \text{ s.t. } c \leq X\varphi + y\}$ and the budget feasible set (see equation (13) below) is defined $B(y) = \{c \in \Re^L | c \geq 0, \pi(c - y) \leq 0\ \forall \pi \in \Pi\}$. There I obtain identical results using slightly more elaborate arguments.

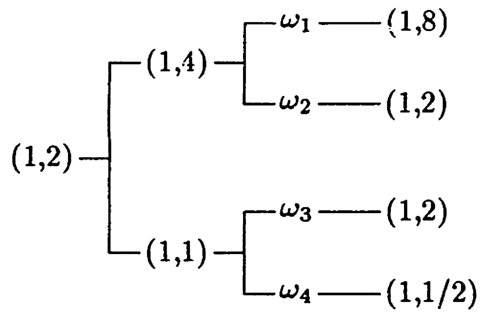


Figure 2
Event tree representation of the securities price process
The two components of the ordered pairs indicate the prices of the riskless and risky securities, respectively.

partitions of Ω: $S_0^1(\Omega) = 2$; $S_1^1(\{\omega_1, \omega_2\}) = 4$, $S_1^1(\{\omega_3, \omega_4\}) = 1$; $S_2^1(\omega_1) = 8$, $S_2^1(\omega_2) = 2$, $S_2^1(\omega_3) = 2$, and $S_2^1(\omega_4) = 1/2$. The numbers of cells at each date are $N_0 = 1$, $N_1 = 2$, and $N_2 = 4$, and I have $L = N_0 + N_1 + N_2 = 7$. Hence any consumption bundle can be represented by a vector in $\Re^7$, i.e., $c = (c(\Omega), c(\{\omega_1, \omega_2\}), c(\{\omega_3, \omega_4\}), c(\{\omega_1\}), c(\{\omega_2\}), c(\{\omega_3\}), c(\{\omega_4\}))^\top$, where the $c(\{\omega_i\})$ denote consumption at the various dates and subsets of Ω. This securities price process can be represented by the event tree in Figure 2.

There are two investment opportunities at date 0, the riskless bond and the risky security. Similarly, at date 1, there are two investment opportunities at each of the cells $\{\omega_1, \omega_2\}$ and $\{\omega_3, \omega_4\}$. Thus a trading strategy is $\varphi \in \Re^6$, and the payoff matrix X has six columns. For this example, I have

$$X = \begin{pmatrix} -1 & -2 & 0 & 0 & 0 & 0 \\ 1 & 4 & -1 & -4 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 8 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1/2 \end{pmatrix}.$$

The consumption bundle produced by a trading strategy φ given an income vector y is $c = X\varphi + y$, and $C(y) = \{c \in \Re^7 \mid c \geq 0, \exists \varphi \in \Re^6 \text{ s.t. } c = X\varphi + y\}$.

Incomplete markets can be represented quite easily in this set-up. For example, if the risky security were not available for trading at time 1, the matrix X would be missing its fourth and sixth columns.

4 Characterization of the Set of Feasible Consumption Bundles

I now make precise and prove the claims made in Section 2. In particular, in this section I show that the set of feasible consumption bundles $C(y)$ consisting of the $c \geq 0$ that can be generated by a dynamic trading strategy using income y is identical to a set $B(y)$ consisting of the $c \geq 0$ that are budget feasible with respect to all of the state prices that are consistent with the absence of arbitrage, and then show that this set $B(y)$ can be characterized using finitely many constraints. Hence the solution of (11) is identical to the solution of the static problem

$$\max_{c \in B(y)} u(c, p), \tag{12}$$

and the trading strategies that implement the solutions are identical.

Breeden (1987) uses a representation of the set of feasible consumption bundles essentially identical to (10) in advancing a numerical technique to solve the consumption-portfolio problem as a static problem. Breeden uses the equality $c = X\varphi + y$ to eliminate c and solved a static problem where the maximization was over φ. By characterizing the set of feasible consumption bundles in terms of the budget feasible set $B(y)$ I provide an alternative characterization of the set of feasible consumption bundles that both is more useful for computation and has a natural economic interpretation.

I first consider incomplete markets when short-sales of the available securities are permitted, and then turn to short-sale constraints.

4.1 Incomplete Markets

If the consumption-portfolio problem is to be well-posed I certainly do not want the securities price process to admit any arbitrage opportunities. That is, I want to exclude the possibility that something might be created from nothing, or that a trading strategy which produces positive payoffs and never requires any investment might exist. I begin by making precise this notion of an arbitrage opportunity.

Definition 1 *An arbitrage opportunity is a vector $\varphi \in \Re^M$ such that $X\varphi \geq 0$ and $X\varphi \neq 0$.*

This definition of an arbitrage opportunity is the same as the "simple free lunch" of Harrison and Kreps (1979). I assume that no such opportunities exist. That is, I assume that $X\varphi \geq 0$ implies $X\varphi = 0$.

I also define a set of state prices Π, or, after normalization by the price of any security, a set of equivalent martingale measures:

$$\Pi = \{\pi \in \Re^L \mid \pi X = 0, \pi > 0, \pi_1 = 1\}.$$

Here π_1 denotes the first component of the vector π. I explain these after I present Proposition 1.

Proposition 1 *If there are no arbitrage opportunities then* Π *is non-empty. Moreover the linear span of* Π *is identical to* $X^{\perp}$*, the largest linear subspace orthogonal to the linear subspace spanned by the column vectors of* X*. Therefore it has dimension* $L - \text{rank}(X)$*, i.e.* $\dim(\text{span}(\Pi)) = L - \text{rank}(X)$.

PROOF. A special case of Motzkin's Transposition Theorem [Schrijver (1986), Corollary 7.1k] says that there is a vector z with $z > 0$ and $zA = 0$ if and only if $Aw \geq 0$ implies $Aw = 0$. Letting $A = X$ and $w = \varphi$, there exists $\hat{\pi} > 0$ such that $\hat{\pi}X = 0$. If I normalize $\hat{\pi}$ so that its first component $\pi_1 = 1$ the normalized vector π will be an element of Π.

Since it is obvious that $\Pi \subset X^{\perp}$, all I need do to prove the second claim is show that any y satisfying $yX = 0$ can be represented as a linear combination of elements of Π. Let $w = \hat{\pi} - \epsilon y$. For sufficiently small ϵ $w > 0$ and $w/w_1 \in \Pi$, so it is easy to see that y can be represented as a linear combination of the $\hat{\pi}$ and w/w_1, i.e. $y \in \text{span}(\Pi)$. ∎

The relationship between this proposition and the securities markets is as follows. If markets are dynamically complete then at each node of the event tree describing the economy there must be available for trading as many linearly independent securities as branches leaving the node, and the dimension of the linear span of Π is one — i.e., there is a unique state price vector. Each column of X corresponds to one investment opportunity or security at one node, so if there are fewer linearly independent securities than branches at some nodes X will have fewer linearly independent columns than in the complete markets case. If I count each "missing" linearly independent security at each node as an unavailable market, the dimension of the linear span of Π, or $\dim X^{\perp}$, is equal to the number of unavailable markets plus one.

The $\pi \in \Pi$ can be interpreted as Arrow-Debreu state prices, and the requirement that $\pi_1 = 1$ says simply that the price of consumption at time 0 is one. These state prices comprise a measure on Ω such that when one integrates a security price S_t^k over the intersection of the current cell and any partition F_s $(s > t)$ with respect to this measure the integral is equal to the current security price multiplied by the price of the current state. Letting A_t denote an element of F_t,

$$\pi(A_t)S_t^k(A_t) = \sum_{A_{t+j} \in F_{t+j} \cap A_t} \pi(A_{t+j})S_{t+j}(A_{t+j}) \quad \forall t \in \{0, \ldots, T-1\},\ \forall j \in \{1, \ldots, T-t\}.$$

If the elements of X have been normalized by the price of any security, or if the rate of return on one of the securities is always zero, then the $\pi \in \Pi$ may also be interpreted as probabilities. In this case, for each $A_t \in F_t$ there is a security with unit current price and unit price in every state next period, so the π indeed are probabilities. The set Π will consist of the equivalent probability measures under which the securities prices are martingales. Letting Q_π be the new probability measure given by π, Q_π is equivalent to the original measure because $\pi > 0$ and the securities prices are martingales because

$$E^{Q_\pi}(S^{k*}(t+1) \mid \mathcal{F}_t) = \sum_{A_{t+j} \in F_{t+j} \cap A_t} \frac{\pi(A_{t+1})}{\pi(A_t)} S^*_{t+1}(A_{t+1}) = S^*_t,$$

where S_t^{*k} denotes the time t normalized price of the k-th security. Normalization is really not necessary in this model because even if one does not normalize the securities prices one can still treat π as a finite measure and define conditional measure in the way one defines conditional probability. The sum of these conditional measures may not be one, but this is the only difference.

Next, define the budget feasible set

$$B(y) = \{c \in \Re^L \mid c \geq 0, \pi(c-y) = 0 \ \forall \pi \in \Pi\}. \tag{13}$$

The following theorem shows that the budget feasible set is identical to the set of feasible consumption bundles.

Theorem 1 $B(y) = C(y)$.

PROOF. Let $D = \{c \in \Re^L \mid \exists \varphi \in \Re^M \text{ s.t. } c - y = X\varphi\}$ and let $E = \{c \in \Re^L \mid \pi(c-y) = 0 \ \forall \pi \in \Pi\}$. I first show $D = E$.

(i) $D \subset E$. Consider $c \in D$. For $\pi \in \Pi$, I have $\pi X \varphi = 0$ giving us $\pi(c-y) = 0$. Therefore $c \in E$.

(ii) $E \subset D$. From Proposition 1, $E = \{c \in \Re^L \mid \pi(c-y) = 0 \ \forall \pi \in X^\perp\}$. Thus there exists a vector $\varphi \in \Re^M$ such that $c - y = X\varphi$.

Together, (i) and (ii) yield $D = E$. Finally, I have $B(y) = E \cap \{c \in \Re^L \mid c \geq 0\} = D \cap \{c \in \Re^L \mid c \geq 0\} = C(y)$. ∎

Theorem 1 provides a complete characterization of the feasible consumption set when markets are incomplete. It allows one to see exactly how the incomplete markets case differs from the complete markets case where the feasible consumption set can be defined by a single budget constraint. This characterization of the budget feasible set is not yet useful for characterizing optimal policies or for computation because $B(y)$ is defined by infinitely many constraints. A more useful characterization of the set of feasible consumption bundles follows.

Define $\overline{\Pi} = \{\pi \in \Re^L \mid \pi X = 0, \pi \geq 0, \pi_1 = 1\} = \text{cl}(\Pi)$. The set $\overline{\Pi}$ is a closed, bounded (convex) polyhedron contained in $X^\perp$, and therefore is generated by (or is the convex hull of) finitely many extreme points. If I define $\Pi^e = \{\pi \mid \pi \text{ is an extreme point of } \overline{\Pi}\}$ it is clear from the discussion in Section 7.2 of Schrijver (1986) that the following lemma is true.

Lemma 1 $\overline{\Pi} = \text{conv.hull}(\Pi^e)$.

Define

$$B^e(y) = \{c \in \Re^L \mid c \geq 0, \pi(c-y) = 0 \ \forall \pi \in \Pi^e\}.$$

Then

Corollary 1 $B^e(y) = B(y)$.

PROOF. From lemma 1, it is easy to see that $B^e(y) = \{c \in \Re^L \mid c \geq 0, \pi(c-y) = 0 \ \forall \pi \in \overline{\Pi}\} = \{c \in \Re^L \mid c \geq 0, \pi(c-y) = 0 \ \forall \pi \in \text{cl}(\Pi)\} = B(y)$. ∎

This is one of the key results of the essay. The characterization of the set of feasible consumption bundles in terms of finitely many constraints involving the extreme points of Π allows one to attack the problem

$$\max_{c \in B^e(y)} u(c,p)$$

using the Lagrangian theory. Furthermore, the budget feasible set has a nice geometric interpretation as a polyhedron defined by budget hyperplanes formed using the extreme points of the set of state prices consistent with the absence of arbitrage. Section 6 consists of an example which illustrates this approach.

For computational purposes, if the number of extreme points, k, of $\text{cl}(\Pi)$ is greater than $\dim(X^\perp)$, then $k - \dim(X^\perp)$ of the constraints $\pi(c-y) = 0$ will be redundant, and $k - \dim(X^\perp)$ of them can be omitted. In fact, one can perform maximization using $\dim(X^\perp)$ constraints of the form $q_i(c-y) = 0$, for $i = 1, \ldots, \dim(X^\perp)$, where the q_i are any $\dim(X^\perp)$ linearly independent vectors that lie in Π or its affine hull. A set of $\dim(X^\perp)$ extreme points of Π form a particularly convenient set, but any set may be used. This may make certain problems easier computationally. Standard algorithms exist for generating a basis for the orthogonal space of a matrix, and given a basis for $X^\perp$ it is easy to find $\dim(X^\perp)$ points that lie in the affine hull of Π. For some problems it may be easier to use these algorithms than to generate the extreme points of Π.[3]

[3]If one uses the Householder method to construct the QR decomposition of X, $X = QR$, then the last $L - M$ columns of Q, say $(q^{M+1}, \ldots, q^L)$, will span the left null space of X (see, e.g., Golub and Van Loan, Section 6.2 (1983)). Given a basis for the left null space of X one can easily construct $\dim(\Pi)$ linearly independent vectors that lie in $\{\pi \in \Re^L \mid \pi X = 0,\ \pi_1 = 1\}$. Consider the q^i, $i \in \{M+1, \ldots, L\}$, such that $q_1^i \neq 0$. (Note that at least one of the q^i must have a nonzero first component, for if $q_1^i = 0\ \forall i$ then $\{\pi \in \Re^L \mid \pi X = 0,\ \pi_1 = 1\} = \emptyset$, which contradicts Proposition 1.) Normalize these so that the first components are equal to one, i.e., define $\overline{q}^i = q^i / q_1^i$. Pick any one of these $\overline{q}^i$, say $\overline{q}^k$, and construct $\widehat{q}^i = \overline{q}^k - \overline{q}^i$ for those vectors with $q_1^i \neq 0$, $i \in \{M+1, \ldots, L\}$, $i \neq k$. Then let $\widehat{q}^i = q^i$ if $q_1^i = 0$, $i \in \{M+1, \ldots, L\}$. Consider the space given by

$$\overline{q}^k + \sum_{\substack{i \in \{M+1, \ldots, L\} \\ i \neq k}} \alpha_i \widehat{q}^i,$$

where $\sum_i \alpha_i = 0$. This space is the affine hull of Π. That is, it is exactly $\{\pi \in \Re^L \mid \pi X = 0,\ \pi_1 = 1\}$.

4.2 Short-sale Constraints

It is straightforward to extend the approach to include short sale constraints on some of the securities. The arguments are either identical to or slight generalizations of those in the previous section. In this formulation short sale constraints (or restrictions on long positions) appear in the form of inequality constraints on the components of φ, i.e. $\varphi_i \geq 0$ or $\varphi_i \leq 0$. A short-sale constraint on one security will involve constraints on more than one of the elements of φ. In particular, if a security is available for trading at k different dates and states and it may never be sold short, then k of the φ_i are constrained to be greater than or equal to zero.

Rearrange the rows of X, and the components of φ, c and y so that

$$\varphi = \begin{pmatrix} \varphi_r \\ \varphi_s \\ \varphi_t \end{pmatrix},$$

$$X = (X_r \quad X_s \quad X_t),$$

with $\varphi_r \geq 0$, $\varphi_s \leq 0$, and φ_t unconstrained. Define

$$C_1(y) = \{c \in \Re^L \mid c \geq 0, \exists\varphi \text{ s.t. } c = X\varphi + y, \varphi_r \geq 0, \varphi_s \leq 0\}.$$

Also define

$$\Pi_1 = \{\pi \in \Re^L \mid \pi > 0, \pi X_r \leq 0, \pi X_s \geq 0, \pi X_t = 0, \pi_1 = 1\}$$

and

$$B_1(y) = \{c \in \Re^L \mid c \geq 0, \pi(c-y) = 0 \;\forall\pi \in \Pi,\; \pi(c-y) \leq 0 \;\forall\pi \in \Pi_1 \setminus \Pi\}.$$

Similar to the intepretation of Π, one can establish the following relationship between the lack of arbitrage opportunites and Π_1.

Proposition 2 *If there are no arbitrage opportunities for dynamic trading strategies satisfying $\varphi_r \geq 0$ and $\varphi_s \leq 0$ then Π_1 is non-empty.*

PROOF. Motzkin's Transposition Theorem [Schrijver (1986), Corollary 7.1k] implies that there is a vector $\hat{\pi}$ with $\hat{\pi} > 0$, $\hat{\pi}X_r \leq 0$, $\hat{\pi}X_s \geq 0$ and $\hat{\pi}X_t = 0$ if and only if $X\varphi \geq 0$ implies $X\varphi = 0$, where $\varphi_r \geq 0$ and $\varphi_s \leq 0$. If one normalizes $\hat{\pi}$ so that its first component $\pi_1 = 1$ the normalized vector π will be an element of Π. ∎

Similar to the incomplete markets case, I want to show that $C_1(y) = B_1(y)$.

Theorem 2 $B_1(y) = C_1(y)$.

PROOF. Let $D = \{c \in \Re^L \mid \exists\ \varphi \text{ s.t. } \varphi_r \geq 0, \varphi_s \leq 0, c = X\varphi + y\}$ and let $E = \{c \in \Re^L \mid \pi(c-y) = 0\ \forall \pi \in \Pi,\ \pi(c-y) \leq 0\ \forall \pi \in \Pi_1 \setminus \Pi\}$. Similar to the proof of Theorem 1, I only need to show $D = E$. $D \subset E$ is obvious, so I only have to show $E \subset D$.

Since $\Pi \subset \Pi_1$, Theorem 1 implies that there exists a vector φ such that $c - y = X\varphi$. I first claim that $\varphi_r \geq 0$. If this is not true, one may assume for simplicity that the first component of φ_r, φ_{r1}, is strictly negative. Find a vector z such that $zX_1 \neq 0$ and $zX^1 = 0$, where X_1 is the first column of X and X^1 is the submatrix of X formed by removing X_1 from X. (Such a vector z always exists because the lack of arbitrage opportunities implies that Π is non-empty.) Now, choose a vector $\pi \in \Pi$ and a sufficiently small real number ϵ such that $\pi_\epsilon = \pi + \epsilon z > 0$ and $\pi_\epsilon X_1 < 0$ and $\pi_\epsilon X^1 = 0$. Let $\hat{\pi}_\epsilon$ be a normalization of π_ϵ such that $\hat{\pi}_\epsilon = \pi_\epsilon / \pi_{\epsilon 1} \in \Pi_1$. I find that $\hat{\pi}_\epsilon(c-y) = \hat{\pi}_\epsilon X_1 \varphi_{t1} > 0$, a clear contradiction. Therefore $\varphi_r \geq 0$, and similarly $\varphi_s \leq 0$. ∎

What is interesting here is that with short-sale restrictions, the set Π_1 consists not only of the equivalent measures under which securities prices are martingales, but also those under which the prices of the securities for which short-sale is restricted are supermartingales. This is similar to the supermartingale result of Dybvig and Ross (1986). An intuition for the result is as follows. If a security price follows a supermartingale with non-zero expected change, then that security is "overpriced," and offers an apparent arbitrage opportunity. However, if short-sales are not possible the opportunity cannot be exploited and nothing precludes the security price from following a supermartingale.

Next define

$$\overline{\Pi}_1 = \{\pi \in \Re^L \mid \pi \geq 0, \pi X_r \leq 0, \pi X_s \geq 0, \pi X_t = 0, \pi_1 = 1\}.$$

The set $\overline{\Pi}_1$ is a closed, bounded (convex) polyhedron and is generated by finitely many extreme points. Let Π_1^e denote the set of extreme points of $\overline{\Pi}_1$. Define

$$B_1^e(y) = \{c \in \Re^L \mid c \geq 0, \pi(c-y) = 0\ \forall \pi \in \Pi^e,\ \pi(c-y) \leq 0\ \forall \pi \in \Pi_1^e\}.$$

Then, similar to the incomplete markets case, I have

Corollary 2 $B_1(y) = B_1^e(y)$.

PROOF. Identical to the proof of Corollary 1. ∎

This corollary allows us to attack the problem

$$\max_{c \in B_1^e(y)} u(c,p)$$

using the Lagrangian theory. Theorems 1 and 2 and Corollaries 1 and 2 combine incomplete markets and short-sale restrictions in a very general way and show how a dynamic consumption and portfolio problem can be solved as a static maximization problem with finitely many equality and inequality constraints.

4.3 Generating Extreme Points

As the last step in characterizing the feasible consumption set I consider how to generate the extreme points of $\overline{\Pi}$ and $\overline{\Pi}_1$ in order to form the budget constraints.[4] In general it can be difficult to generate the extreme points of an arbitrary polyhedron. However, the event tree information structure gives the payoff matrix X a special structure that considerably simplifies the task of generating the extreme points. I need the following proposition which follows from Theorem 8.4 in Schrijver (1986) and the discussion immediately following that theorem.

Proposition 3 *Let π be a vector in $\Re^L$, let A and B be matrices, let b be a vector, and let P be the bounded (convex) polyhedron defined by the system $\pi A \geq 0$, $\pi B = b$. A point π^e is an extreme point of P if and only if $\pi^e \in P$ and π^e satisfies $\pi B = b$ along with $L - \text{rank}(B)$ linearly independent equations from the subsystem $\pi A = 0$.*

Let

$$J = \begin{pmatrix} 0 \\ I \end{pmatrix},$$

where J is $L \times (L-1)$ and I is an $(L-1) \times (L-1)$ identity matrix, and let $\underline{X} = (\, z \quad X \,)$, where $z^\top = (\, 1 \quad 0 \quad \ldots \quad 0 \,)$. One can see that $\overline{\Pi}$ is defined by

$$\begin{aligned} \pi J &\geq 0, \\ \pi \underline{X} &= z^\top. \end{aligned}$$

Set up a system of the form

$$\pi Z = (z^\top, 0),$$

where Z is a matrix formed from the columns of $\underline{X}$ and $L-\text{rank}(\underline{X})$ columns from J. This is essentially constraining $L-\text{rank}(\underline{X})$ of the elements of π to equal zero and solving for the remaining elements such that $\pi X = 0$. Applying the above proposition with $A = J$, $B = \underline{X}$, and $b = z^\top$, one sees that any solution $\hat{\pi}$ of the above system is a candidate extreme point. If

[4] As indicated in Section 4.1, when there are no short-sale constraints one can form the budget constraints by using $\dim(X^\perp)$ vectors that lie in the affine hull of Π.

$\hat{\pi} \in \Pi$ then it is an extreme point; otherwise it is not and it is discarded. One can continue this procedure until one has formed all possible matrices Z.

One could apply this proposition directly to the entire system in order to generate the extreme points of $\overline{\Pi}$. However, this would not exploit the special structure of the event tree and the payoff matrix X. Instead, I use this proposition to generate the extreme points of the sets of conditional state prices at the price subsystems at each node of the event tree, and then combine these conditional state prices in the obvious fashion. The following proposition shows how this can be done.

Proposition 4 *The set of extreme points for the whole price system Π^e can be generated through multiplications of the extreme points of the price subsystems at each node of the event tree. That is, the extreme points of the price subsystems can be interpreted as conditional probabilities (or measures), and they can be used to define the probabilities (or measures) for each node using the rule $P(A \cap B) = P(B|A)P(A)$.*

PROOF. I only sketch the basic idea. Suppose one has an extreme point π for the whole price system. Take any sub-tree, and project π to the price subsystem associated with this sub-tree. One obtains either a zero probability at the node that generates this tree or a positive probability. In the latter case, I claim that the projection defines an extreme point for the sub-tree. If this is not true, the projection can be written as a convex combination of two different points of the subsystem. This in turn defines two different points in Π which have the property that π can be represented as a convex combination of these two points. This contradicts the assumption that π is an extreme point.

Conversely, any point π generated from the extreme points of the subsystems through multiplications must be an extreme point of the whole system, for otherwise it could be represented as a convex combination of two different points in Π. The projection of π on any sub-tree (which is either zero or an extreme point for the subsystem by assumption) could be written as a convex combination of the projection of the other two points. The projection of the other two points must differ on at least one sub-tree. This violates the assumption that π is generated from extreme points of the subsystems through multiplications. ∎

With this procedure, the problem of finding the extreme points of a polyhedron is reduced to a series of comparatively simple problems that involve finding the extreme points at each node of the event tree. The solution for π can then be obtained by the appropriate multiplications of the solutions of these smaller problems. This procedure is illustrated in the example in Section 6.

A similar procedure can be employed for the extreme points of $\overline{\Pi}_1$. The polyhedron $\overline{\Pi}_1$ is defined by the systems

$$\begin{aligned}\pi J &\geq 0,\\ \pi X_r &\geq 0,\\ \pi X_s &\leq 0,\\ \pi \underline{X_t} &= z^{\mathsf{T}},\end{aligned}$$

where J, $\underline{X_t}$ and z correspond to the J, $\underline{X}$, and z defined earlier. By Proposition 3 one can find the extreme points of this polyhedron by forming a system

$$\pi Z = 0,$$

where Z is an $L \times L$ matrix formed from the columns of $\underline{X_t}$ and $L-\text{rank}(\underline{X_t})$ linearly independent columns taken from J, X_r and X_s. The solution $\hat{\pi}$ of this system is a candidate extreme point. If $\hat{\pi} \in \overline{\Pi}_1$, $\hat{\pi}$ is an extreme point, otherwise one discards it. Again, one continues this procedure until one has formed all possible matrices Z.

Just as before, the actual computation of the extreme points is simplified by the fact that the matrix X has a special structure and the extreme points of the whole system can be obtained through multiplications of the extreme points of the price subsystems at each node.

5 Solution of the Optimization Problem

Given the characterization of the feasible set, the agent's problem (11) becomes

$$\max_c u(c, p) \tag{14}$$

subject to the constraints

$$\begin{aligned} c &\geq 0, && && (15)\\ \pi(c-y) &= 0 && \forall \pi \in \Pi^e, && (16)\\ \pi(c-y) &\leq 0 && \forall \pi \in \Pi_1^e. && (17)\end{aligned}$$

The $\pi \in \Pi_1^e$ and the $\pi \in \Pi^e$ are the extreme points of the sets

$$\overline{\Pi}_1 = \{\, \pi \in \Re^L \mid \pi \geq 0,\ \pi X_r \leq 0,\ \pi X_s \geq 0,\ \pi X_t = 0,\ \pi_1 = 1 \,\}$$

and

$$\overline{\Pi} = \{\, \pi \in \Re^L \mid \pi \geq 0,\ \pi X = 0,\ \pi_1 = 1 \,\}.$$

The problem (14)–(17) is a nonlinear program subject to linear inequality and equality constraints and can be solved using standard algorithms (see, for example, Gill, Murray, and Wright (1981) and Fletcher (1981)).

A useful characterization of the solution of the static problem (14)–(17) can be obtained from a simple manipulation of the first order conditions. The solution of (14)–(17) will satisfy a system of equations

$$\begin{aligned} u_{c_i}(c,p) - \eta_i &= \sum_{r=1}^{R} \lambda^r \pi_i^r \qquad i = 1,\ldots,L, &(18)\\ \pi^r(c-y) &= 0 \qquad r = 1,\ldots,\bar{r}, &(19)\\ \pi^r(c-y) &\leq 0 \qquad r = \bar{r}+1,\ldots,R, &(20)\\ c_i &\geq 0 \qquad i = 1,\ldots,L, &(21)\end{aligned}$$

where the extreme points associated with the constraints (16) and (17) are indexed using the integers $1,\ldots,R$, the η_i are the Lagrange multipliers of the constraints $c_i \geq 0$, and the λ^r are the Lagrange multipliers of the constraints $\pi^r(c-y) \leq 0$. We can rewrite (18) as

$$\frac{\partial}{\partial c_i} u(c,p) - \eta_i = \lambda \sum_{r=1}^{R} \alpha_r \pi_i^r \qquad i = 1,\ldots,L, \tag{22}$$

where $\sum_{r=1}^{R} \alpha_r = 1$, $\alpha_r \geq 0$. Examining (22), one can see that this is one of the first order conditions of a problem with one budget constraint formed using a convex combination of the extreme points. The particular convex combination depends upon the utility function. That is, at the solution it is as though there were complete markets, where the Arrow-Debreu state price vector depends on the utility function.

This characterization of the solution is very simple. The unique state price vector is just a separating hyperplane. (The budget feasible set $B(y)$ and the upper contour sets of the utility function are convex sets.) However, the characterization, while not very deep, is useful.[5]

Associated with this single Arrow-Debreu state price are, at every node, conditional state prices for consumption at the nodes next period. For utility functions of the HARA class, the conditional state prices are the same at every node when the uncertainty in the prices of the securities is multiplicative, there is a bond, and the investment opportunity set is constant (and there are no binding short-sale constraints or non-negativity constraints on consumption). Further, the single Arrow-Debreu price for the whole system can be recovered by multiplying together the conditional state prices as in Proposition 4. Since the conditional state prices are the same at every node, one can recover the price for the whole system by solving only a

[5]This characterization in terms of a single state price also motivated some of the research in the continuous time case [He and Pearson (1989)].

one-period problem. Once one has obtained the single Arrow-Debreu price for the whole system it is easy to calculate the optimal consumption and portfolio policies.

A direct proof of the claim that the conditional state prices are the same at all nodes is both cumbersome and tedious, for it involves writing out all of the first order conditions for the static problem with budget constraints formed using the elements of Π^e and recognizing certain symmetries. However, it is easy to see why the claim should be true. It is well known that in the setting described above the proportion of the agent's risky asset portfolio invested in each of the risky securities is independent of wealth (e.g., Hakansson (1970); see also Merton (1971)). Hence the optimal proportions will not depend on the current node. But the portfolio choices can be the same at all nodes only if the conditional state prices are the same at all nodes.

5.1 Implementing an optimal solution.

Once an optimal consumption bundle is found, the remaining issue is how to implement it through a dynamic trading strategy, i.e., how to solve explicitly for φ. From the equality $c - y = X\varphi$ it is clear that one can solve for φ in terms of $c - y$ and the elements of X. This turns out to be fairly simple to do. Working backwards, given the optimal consumption pattern at date T one determines the trading strategy for each node at time $T - 1$ and the wealth level required to carry out this strategy. This requires only that one solve a system of linear equations. After finding the wealth required at each node at time $T - 1$ and adding the time $T - 1$ consumption at that node one can determine the trading strategies for each of the nodes at $T - 2$ and the wealth levels required at these nodes. Repeating the procedure T times gives the trading strategies for all dates and nodes. This procedure is illustrated in the example in Section 6.

5.2 Computational considerations

In formulation the agent's original problem in this paper (11) becomes the static problem (14)–(17), which can be solved using standard nonlinear programming algorithms. An alternative formulation of the problem is due to Breeden (1987). Breeden also considers the intertemporal consumption and portfolio problem (11), uses the relationship $c = X\varphi + y$ to eliminate c, and solves

$$\max_{\varphi} u(X\varphi + y, p) \tag{23}$$

subject to

$$\varphi \geq 0, \tag{24}$$

$$X\varphi + y \geq 0. \tag{25}$$

This is also a nonlinear program subject to linear inequality and equality constraints and can be solved using standard algorithms. Dynamic programming is yet another alternative for solving the original problem if the agent's preferences have an expected utility representation.

Breeden (1987) discusses why his approach is better for computation than dynamic programming when the agent has a time additive state independent utility function, and his discussion also applies to the static problem (14)–(17). The computational inefficiency of dynamic programming stems from the fact that computing optimal policies using dynamic programming requires that at each node one compute the optimal policies for all possible levels of wealth, including those that will never be realized at the node. The relative superiority of the static formulations is even greater when the agent's utility function is not time additive (or more generally of the linear Koopmans aggregator type discussed by Bergman (1985)). For utility functions in this class optimal policies depend upon wealth but not upon past consumption histories. In general optimal policies depend upon past consumption histories as well as wealth, and computing optimal policies using dynamic programming requires that at each node one compute the optimal policies for all possible levels of wealth and past consumption histories. This may involve computing optimal policies for all possible realizations of a high dimensional random variable, and Breeden's discussion of the computational inefficiency of dynamic programming applies with even more force.

If the agent's preferences do not have an expected utility representation dynamic programming cannot be used to compute optimal policies and the static formulations of this essay and Breeden are the only way to compute optimal intertemporal consumption and portfolio policies.

In general, there is little that can be said about the relative merits of the two static formulations, Breeden's (23)–(25) and the one of this paper (14)–(17). Computational efficiency is likely to be problem, data, and algorithm-dependent. However, the formulation in this paper is superior for additively separable utility functions, a class which includes the utility functions most commonly used in finance.

If the agent's utility is additively separable, the objective function (14) will be of the form

$$u(c,p) = u(c_1) + \sum_{i=2}^{L} p_i u(c_i), \tag{26}$$

which is additively separable in the variables c_i. The objective function in Breeden's formulation is

$$u(X\varphi + y, p) = u(x_1\varphi + y_1) + \sum_{i=2}^{L} p_i u(x_i\varphi + y_i), \tag{27}$$

where x_i denotes the i-th row of the matrix X and y_i denotes the i-th component of the vector y. Even though the agent's utility is additively separable in consumption, the objective function is not additively separable in the variables φ_i over which maximization is to be performed. This lack of additive separability, compared to the additive separability in the c_i in the formulation in this paper, makes the latter approach preferred for computation.

The problem (14)–(17) with an additively separable utility function is a *separable programming problem* (see, e.g., Walsh (1975), p. 5). General nonlinear programming algorithms tend to be more efficient when applied to such problems.[6] More importantly, special methods are available for the solution of such problems (see, e.g., Hadley (1964) and Beale (1970); a more recent algorithm suited for such problems is described in Fourer (1986)). In essence, these methods reduce the nonlinear programming problem to an approximating linear programming problem via piecewise linear approximations of the nonlinear functions. One can see the basic idea with a simple example.

Take the objective function (26), and construct a piecewise linear approximation of each of the L component functions $u(c_i)$.[7] This is shown in Figure 3, where each of the linear segments has the indicated slope d_i^j. The agent's problem (14)–(17) becomes

$$\max_{e_i^1, e_i^2, e_i^3} \left[(d_1^1 e_1^1 + d_1^2 e_1^2 + d_1^3 e_1^3) + \sum_{i=2}^{L} p_i (d_i^1 e_i^1 + d_i^2 e_i^2 + d_i^3 e_i^3) \right]$$

s.t.

$$\begin{aligned}
c_i &= e_i^1 + e_i^2 + e_i^3, \quad i = 1, \ldots, L, \\
0 &\leq e_i^1 \leq M_i^1 \quad i = 1, \ldots, L, \\
0 &\leq e_i^2 \leq M_i^2 \quad i = 1, \ldots, L, \\
0 &\leq e_i^3 \leq M_i^3 \quad i = 1, \ldots, L, \\
c &\geq 0, \\
\pi(c - y) &\leq 0 \quad \forall \pi \in \Pi_1^e, \\
\pi(c - y) &= 0 \quad \forall \pi \in \Pi^e.
\end{aligned}$$

Since each of the $u(c_i)$ is concave, one has $d_i^1 > d_i^2 > d_i^3$, implying that the variable e_i^1 is preferred to e_i^2 and the variable e_i^2 is preferred to e_i^3. Thus, at optimality one is assured that $e_i^2 > 0$ implies $e_i^1 = M_i^1$ and $e_i^3 > 0$ implies $e_i^2 = M_i^2$.

[6] A limited set of computational experiments using an algorithm for general nonlinear programming problems, subroutine E04VCF from the NAG FORTRAN Library – Mark 11 [Numerical Algorithms Group Limited (1982)], confirmed that this is true for power and exponential utility functions in a simple example.

[7] This discussion is drawn from an example in Shapiro (1979).

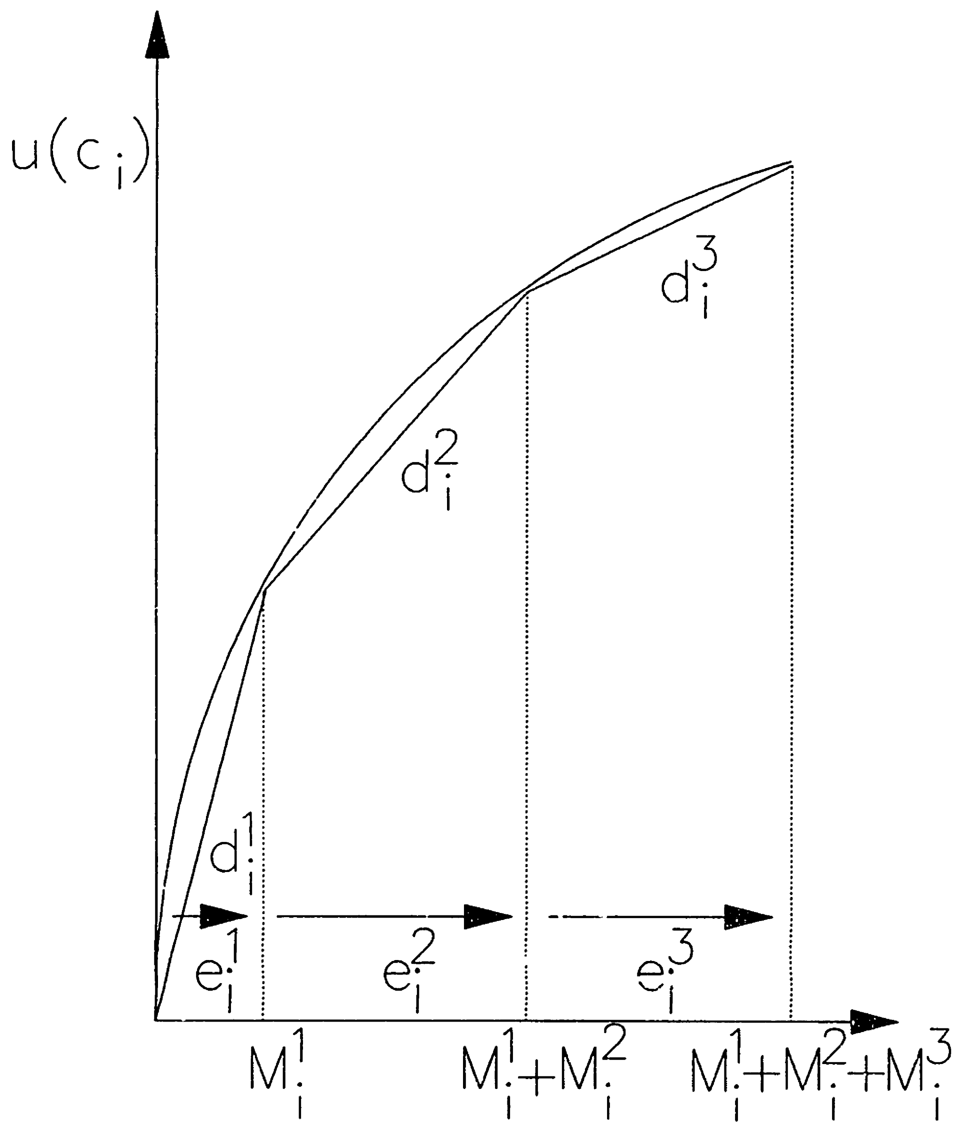


Figure 3
Approximation of the objective function
Each component of the objective function in an additively separable programming problem may be approximated by a piecewise linear function.

This approximating linear program is much easier to solve than the original nonlinear program. While it only yields approximate solutions, more accurate approximations can be obtained by repeated solution using finer approximations in the regions surrounding the solutions of the previous approximations. The relative efficiency of this approach is of no importance in small examples, but is likely to be of importance in problems of large enough size to be of practical interest.

6 An Example

The economy has three dates $t = 0, 1, 2$ and the following partitions of Ω: $F_0 = \Omega$, $F_1 = \{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5, \omega_6\}, \{\omega_7, \omega_8, \omega_9\}\}$, and $F_2 = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9\}$. There are two securities, a riskless bond with price $S_t^0 = 1$ for $t = 0, 1, 2$, and a risky security with prices $S_0^1(\Omega) = 1$; $S_1^1(\{\omega_1, \omega_2, \omega_3\}) = 2$, $S_1^1(\{\omega_4, \omega_5, \omega_6\}) = 1$, $S_1^1(\{\omega_7, \omega_8, \omega_9\}) = 1/2$; $S_2^1(\omega_1) = 4$, $S_2^1(\omega_2) = 2$, $S_2^1(\omega_3) = 1$, $S_2^1(\omega_4) = 2$, $S_2^1(\omega_5) = 1$, $S_2^1(\omega_6) = 1/2$, $S_2^1(\omega_7) = 1$, $S_2^1(\omega_8) = 1/2$, and $S_2^1(\omega_9) = 1/4$. This securities price process can be represented by the event tree shown in Figure 4.

Using the convention for allocating the components of π, c, and y to the various events,

$$\begin{aligned}\pi = (&\pi(\Omega), \pi(\{\omega_1, \omega_2, \omega_3\}), \pi(\{\omega_4, \omega_5, \omega_6\}), \pi(\{\omega_7, \omega_8, \omega_9\}),\\ &\pi(\omega_1), \pi(\omega_2), \pi(\omega_3), \pi(\omega_4), \pi(\omega_5), \pi(\omega_6), \pi(\omega_7), \pi(\omega_8), \pi(\omega_9)),\end{aligned}$$

$$\begin{aligned}c = (&c(\Omega), c(\{\omega_1, \omega_2, \omega_3\}), c(\{\omega_4, \omega_5, \omega_6\}), c(\{\omega_7, \omega_8, \omega_9\}),\\ &c(\omega_1), c(\omega_2), c(\omega_3), c(\omega_4), c(\omega_5), c(\omega_6), c(\omega_7), c(\omega_8), c(\omega_9))',\end{aligned}$$

and similarly for y. The matrix X is

$$X = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ .5 & 1 & 0 & 0 & 0 & 0 & -.5 & 1 \\ 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & .5 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & .25 & 1 \end{pmatrix}.$$

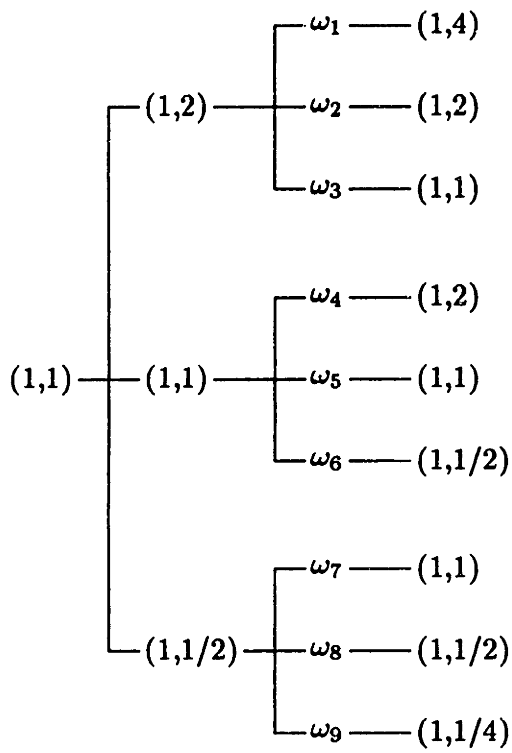


Figure 4
Event tree representation of the securities price process
The two components of the ordered pairs indicate the prices of the riskless and risky securities, respectively.

The first step is to find the extreme points of $\overline{\Pi}$.

The matrix Z consists of X along with four columns taken from J. The columns from J have the effect of constraining certain elements of π to equal zero. Any π such that $\pi X = 0$ and $\pi_1 = 1$ can be written

$$\pi = \{1, \pi_2, \pi_3, \pi_4, \pi_2\frac{\pi_5}{\pi_2}, \pi_2\frac{\pi_6}{\pi_2}, \pi_2\frac{\pi_7}{\pi_2}, \pi_3\frac{\pi_8}{\pi_3}, \pi_3\frac{\pi_9}{\pi_3}, \pi_3\frac{\pi_{10}}{\pi_3}, \pi_4\frac{\pi_{11}}{\pi_4}, \pi_4\frac{\pi_{12}}{\pi_4}, \pi_4\frac{\pi_{13}}{\pi_4}\}. \qquad (28)$$

The components π_2, π_3, π_4 satisfy the equations

$$\begin{pmatrix} 1 & \pi_2 & \pi_3 & \pi_4 \\ 1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 2 & 1 \\ 1 & 1 \\ 0.5 & 1 \end{pmatrix} = (0 \quad 0), \qquad (29)$$

along with a constraint requiring that one of the π_i equal zero arising from one of the columns of J. The components π_5/π_2, π_6/π_2, π_7/π_2 satisfy the equations

$$\begin{pmatrix} 1 & \pi_5/\pi_2 & \pi_6/\pi_2 & \pi_7/\pi_2 \\ 1 & \pi_5/\pi_2 & \pi_6/\pi_2 & \pi_7/\pi_2 \end{pmatrix} \begin{pmatrix} -2 & -1 \\ 4 & 1 \\ 2 & 1 \\ 1 & 1 \end{pmatrix} = (0 \quad 0), \tag{30}$$

along with a constraint requiring that one of the terms $\frac{\pi_i}{\pi_2}$ equal zero. The values of π_5, π_6, and π_7 are obtained from the obvious multiplications. Similar equations hold for the other components of π.

Solving these equations amounts to finding the extreme points for the price subsystems at each node of the event tree in Figure 4. It is easy to see that $(0,1,1)$ and $(1/3,0,2/3)$ are the only two extreme points that satisfy (29). Similarly, $(0,1,0)$ and $(1/3,0,2/3)$ are the only two extreme points that satisfy (30), and one finds the same two extreme points for the other two subsystems.

Now one finds the elements of Π^e, the extreme points of the closure of the set of equivalent measures for the whole price system, by multiplying together the extreme points of the subsystems as one would usually do for conditional probabilities; how to do this is clear from equation (28). One obtains

$$\Pi^e = \left\{ \begin{pmatrix} 1\\0\\1\\0\\0\\0\\0\\0\\1\\0\\0\\0\\0 \end{pmatrix}, \begin{pmatrix} 1\\0\\1\\0\\0\\0\\0\\1/3\\0\\2/3\\0\\0\\0 \end{pmatrix}, \begin{pmatrix} 1\\1/3\\0\\2/3\\0\\1/3\\0\\0\\0\\0\\0\\2/3\\0 \end{pmatrix}, \begin{pmatrix} 1\\1/3\\0\\2/3\\1/9\\0\\2/9\\0\\0\\0\\2/9\\0\\4/9 \end{pmatrix}, \begin{pmatrix} 1\\1/3\\0\\2/3\\1/9\\0\\2/9\\0\\0\\0\\0\\2/3\\0 \end{pmatrix}, \begin{pmatrix} 1\\1/3\\0\\2/3\\0\\1/3\\0\\0\\0\\0\\2/9\\0\\4/9 \end{pmatrix} \right\}.$$

These six extreme points are not linearly independent and any five of them may be used to define the constraints in the maximization problem.

The nonzero components of these extreme points may be interpreted as the prices of consumption in the various states. For example, the least costly way to obtain an additional unit of consumption in the state $\{\omega_1,\omega_2,\omega_3\}$ without decreasing consumption in any state is to take a long position in 2/3 share of the risky security and a short position in 1/3 unit of the riskless bond. This position yields 1 unit of consumption in the state $\{\omega_1,\omega_2,\omega_3\}$ and 1/2 unit of consumption in the state $\{\omega_1,\omega_2,\omega_3\}$. Entering this position costs 1/3 unit of current consumption, and $\pi_2^r = \pi^r(\{\omega_1,\omega_2,\omega_3\}) = 1/3$ for the extreme points with $\pi_2^r \neq 0$ (here π^r denotes the r-th element of Π^e). This set of transactions is also the cheapest way to obtain consumption in the state $\{\omega_7, \omega_8, \omega_9\}$, and $\pi_4^r = \pi^r(\{\omega_7, \omega_8, \omega_9\}) = 2/3$ for the extreme points with $\pi_4^r \neq 0$.

With the extreme points one can now solve the static utility maximization problem and determine the optimal consumption bundle. If all outcomes are equally likely and that the consumer has log utility and an endowment of 10 now and zero in both future periods the problem is

$$\max_c \left[\log c_1 + 1/3 \sum_{i=2}^{4} \log c_i + 1/9 \sum_{i=5}^{13} \log c_i\right]$$

subject to

$$\pi^r c = 10 \qquad \text{for } r = 1, \ldots, 5.$$

Here the first five elements of Π^e are used to define $B(y)$. The optimal consumption plan is

$$c = (3\frac{1}{3}, 5, 3\frac{1}{3}, 2\frac{1}{2}, 7\frac{1}{2}, 5, 3\frac{3}{4}, 5, 3\frac{1}{3}, 2\frac{1}{2}, 3\frac{3}{4}, 2\frac{1}{2}, 1\frac{7}{8}).$$

Given the optimal consumption plan one works backward to determine the trading strategies. For example, at the node $\{\omega_1, \omega_2, \omega_3\}$, the trading strategy satisfies the (overdetermined) system

$$\begin{aligned}
\varphi^0(\{\omega_1, \omega_2, \omega_3\}) + 4\varphi^1(\{\omega_1, \omega_2, \omega_3\}) &= 7\frac{1}{2}, \\
\varphi^0(\{\omega_1, \omega_2, \omega_3\}) + 2\varphi^1(\{\omega_1, \omega_2, \omega_3\}) &= 5, \\
\varphi^0(\{\omega_1, \omega_2, \omega_3\}) + \varphi^1(\{\omega_1, \omega_2, \omega_3\}) &= 3\frac{3}{4},
\end{aligned}$$

where $\varphi^1(\{\omega_1, \omega_2, \omega_3\})$ denotes the number of units of the risky security to hold at node $\{\omega_1, \omega_2, \omega_3\}$ and $\varphi^0(\{\omega_1, \omega_2, \omega_3\})$ denotes the number of units of the bond. Solving these,

$$\begin{aligned}
\varphi^0(\{\omega_1, \omega_2, \omega_3\}) &= 2\frac{1}{2}, \\
\varphi^1(\{\omega_1, \omega_2, \omega_3\}) &= 1\frac{1}{4}.
\end{aligned}$$

The total cost at $\{\omega_1, \omega_2, \omega_3\}$ is then $\varphi^0(\{\omega_1, \omega_2, \omega_3\}) + 2\varphi^1(\{\omega_1, \omega_2, \omega_3\}) = 5$. Performing the calculations at the other nodes in the obvious way yields

$$
\begin{aligned}
\varphi^0 &= (3\frac{1}{3}, 2\frac{1}{2}, 1\frac{2}{3}, 1\frac{1}{4}), \\
\varphi^1 &= (1\frac{2}{3}, 1\frac{1}{4}, 1\frac{2}{3}, 2\frac{1}{2}),
\end{aligned}
$$

where $\varphi^0 \equiv (\varphi^0(\Omega), \varphi^0(\{\omega_1, \omega_2, \omega_3\}), \varphi^0(\{\omega_4, \omega_5, \omega_6\}), \varphi^0(\{\omega_7, \omega_8, \omega_9\}))$ and φ^1 is defined analogously.

7 Conclusion

In this essay I show how the martingale approach can be used to determine optimal intertemporal consumption and portfolio policies in a general finite dimensional economy with incomplete markets and short-sale constraints. When markets are incomplete the feasible consumption bundles satisfy budget constraints formed using all Arrow-Debreu state prices consistent with the absence of arbitrage. The set of feasible consumption bundles is identical to the budget feasible set defined by the budget constraints formed using the extreme points of the closure of the set of state prices consistent with no arbitrage. This result makes the martingale approach useful even when markets are incomplete and there are infinitely many Arrow-Debreu state prices or equivalent measures consistent with the absence of arbitrage.

When there are short-sale constraints securities prices must be super-martingales under the set of Arrow-Debreu state prices or equivalent measures consistent with the absence of arbitrage, and the feasible consumption bundles must satisfy budget constraints formed using all of these state prices. In this case I also show that this set of feasible consumption bundles may be defined using only budget constraints formed from the extreme points of the set of state prices consistent with no arbitrage.

This approach to the consumption-portfolio problem is very convenient for computation. It also provides an alternative approach to obtain the optimal consumption policies for time-additive, state-independent, utility functions of the HARA class (with no nonnegativity restrictions on consumption) when there is a constant investment opportunity set and there are no short-sale constraints.

All of the results of course apply only to the discrete time finite dimensional case. How the martingale approach can be useful in the infinite dimensional continuous time case with incomplete markets and short-sale constraints is the subject of another paper, He and Pearson (1989).

Appendix

This appendix studies the budget feasible set when the agent may discard the consumption good. I start with incomplete markets and no restrictions on short-sales, and then consider the problem when short-sales are restricted.

Incomplete markets

I again use $C(y)$ to denote the set of feasible consumption bundles, and have

$$C(y) = \{c \in \Re^L \mid c \geq 0, \exists \varphi \in \Re^M \text{ s.t. } c \leq X\varphi + y\}.$$

Again, it is easy to see that $C(y)$ is identical to the set of feasible consumption bundles in a dynamic programming formulation of the consumption and portfolio problem.

In this case I define the budget feasible set

$$B(y) = \{c \in \Re^L \mid c \geq 0, \pi(c - y) \leq 0 \ \forall \pi \in \Pi\}.$$

I need the following variant of Farkas' Lemma to show that $B(y) = C(y)$.

Lemma 2 *Let A be a $m \times n$ matrix and let b be an $m-$vector. Then the system of linear inequalities $Az \leq b$ has a solution z iff $wb \geq 0$ for each row vector $w \geq 0$ with $wA = 0$.*

PROOF. Schrijver, Corollary 7.1e. ∎

With this lemma I can now prove a theorem that corresponds to Theorem (2) above, i.e. I can prove that the budget feasible set is identical to the set of feasible consumption bundles.

Theorem 3 $B(y) = C(y)$.

PROOF. Let $D = \{c \mid \exists \varphi \in \Re^M \text{ s.t. } c - y \leq X\varphi\}$ and let $E = \{c \mid \pi(c - y) \leq 0, \ \forall \pi \in \Pi\}$. I first show $D = E$.

(i) $D \subset E$. Consider $c \in D$. For $\pi \in \Pi$, I have $\pi X \varphi = 0$ giving us $\pi(c - y) \leq 0$. Therefore $c \in E$.

(ii) $E \subset D$. Consider $c \in E$. Clearly $\pi(y - c) \geq 0 \ \forall \pi \in \Pi$. Using the above lemma with $-X = A$, $\pi = w$, and $y - c = b$ I have that $-X\varphi \leq y - c$ has a solution for φ, i.e. $c - y \leq X\varphi$ has a solution φ. Therefore $c \in D$.

Together, (i) and (ii) yield $D = E$. Finally, I have $B(y) = E \cap \{c \mid c \geq 0\} = D \cap \{c \mid c \geq 0\} = C(y)$. ∎

I also provide a more useful characterization of the set of feasible consumption bundles corresponding to Corollary 1. First define $\overline{\Pi} \equiv \{\pi \geq 0 \mid \pi X = 0, \pi \geq 0, \pi_1 = 1\} = \text{cl}(\Pi)$. As above, conv.hull$(\Pi^e) = \overline{\Pi}$.

Corollary 3 *Defining $B^e(y) = \{c \geq 0 \mid \pi(c - y) \leq 0 \ \forall \pi \in \Pi^e\}$, then $B^e(y) = B(y)$.*

PROOF. It is easy to see that $B^e(y) = \{c \geq 0 \mid \pi(c-y) \leq 0 \ \forall \pi \in \overline{\Pi}\} = \{c \geq 0 \mid \ \pi(c-y) \leq 0 \ \forall \pi \in \mathrm{cl}(\Pi)\} = B(y)$. ▮

Short-sale constraints

Define

$$C_1(y) = \{c \geq 0 \mid \exists \varphi \text{ s.t. } c \leq X\varphi + y, \varphi_r \geq 0, \varphi_s \leq 0\}.$$

Also define

$$\Pi_1 = \{\varphi \in \Re^L \mid \pi > 0, \pi X_r \leq 0, \pi X_s \geq 0, \pi X_t = 0, \pi_1 = 1\}$$

and

$$B_1(y) = \{c \geq 0 | \ \pi(c-y) \leq 0 \ \forall \pi \in \Pi_1\}.$$

Similar to the incomplete markets case, I want to show that $C_1(y) = B_1(y)$. To do this I need the following lemma.

Lemma 3 *Let A be a matrix, w be a row vector, and z and b be column vectors. Partition A so that $A = (\, A_r \quad A_s \quad A_t \,)$ and $z = \begin{pmatrix} z_r \\ z_s \\ z_t \end{pmatrix}$ The system $Az \leq b$ has a solution z with $z_r \geq 0$, $z_s \leq 0$ if and only if $wb \geq 0$ for each row vector $w \geq 0$ with $wA_r \geq 0$, $wA_s \leq 0$, $wA_t = 0$.*

PROOF. Define $A' = (\, I \quad A_r \quad -A_s \quad A_t \quad -A_t \,)$ and accordingly z', a column vector . Then $Az \leq b$ has a solution z with the desired properties iff $A'z' = 0$ has a solution $z' \geq 0$. $A'z' = b$ has a solution $z' \geq 0$ iff $wb \geq 0$ for each row vector w with $wA' \geq 0$ [Schrijver, Corollary 7.1d]. But $wA' \geq 0$ implies $w \geq 0$, $wA_r \geq 0$, $-wA_s \geq 0$, $wA_t \geq 0$ and $-wA_t \geq 0$, which is equivalent to $w \geq 0$, $wA_r \geq 0$, $wA_s \leq$ and $wA_t = 0$. ▮

I am now in a position to prove the equivalence of $C_1(y)$ and $B_1(y)$.

Theorem 4 $B_1(y) = C_1(y)$.

PROOF. Let $D = \{c \mid \exists \ \varphi \text{ s.t. } \varphi_r \geq 0, \varphi_s \leq 0, c \leq X\varphi + y\}$ and $E = \{c \mid \pi(c-y) \leq 0 \ \forall \pi \in \Pi_1\}$. Similar to the proof of Theorem 3, I only need to show $D = E$. $D \subset E$ is obvious, so I only have to show $E \subset D$.

Consider $c \in E$. Clearly, $\pi(y-c) \geq 0$ for all $\pi \in \Pi_1$. Using the above lemma with $A = -X$, $w = \pi$, $b = y - c$ and $z = \varphi$, I have that $-X\varphi \leq y - c$ has a solution φ with $\varphi_r \geq 0$, $\varphi_s \leq 0$, or equivalently, $c \leq X\varphi + y$ has a solution for φ with the desired properties. Therefore $c \in D$. ▮

Next define

$$\overline{\Pi}_1 \equiv \{\pi \geq 0 \mid \pi X_r \leq 0, \pi X_s \geq 0, \pi X_t = 0, \pi_1 = 1\}.$$

Let Π_1^e denote the set of extreme points of $\overline{\Pi}_1$. Define

$$B_1^e(y) = \{c \geq 0 \mid \ \pi(c - y) \leq 0 \ \forall \pi \in \Pi_1^e\}.$$

Then similar to the incomplete markets case,

Corollary 4 $B_1(y) = B_1^e(y)$.

PROOF. Identical to the proof of Corollary 3. ∎

As above, this corollary allows one to attack the problem

$$\max_{c \in B_1^e(y)} u(c, p)$$

using the Lagrangian theory, and the set of extreme points may be found as in Section 4.

References

Beale, E. M. L., 1970, "Advanced Algorithmic Features for General Mathematical Programming Systems," in J. Abadie (ed.), *Integer and Nonlinear Programming,* North-Holland, Amsterdam.

Bergman, Y. Z., 1985, "Time Preference and Capital Asset Pricing Models," *Journal of Financial Economics,* 14, 1985, 145–160.

Breeden, D. T., 1987, "Optimal Dynamic Trading Strategies: A Simple Approach," Working Paper 87-107, Fuqua School of Business, Duke University, August.

Chamberlain, G., 1988, "Asset Pricing in Multiperiod Securities Markets," *Econometrica,* 56, 1283–1300.

Cox, J. C., and C.-f. Huang, 1989a, "A Variational Problem Arising in Financial Economics," working paper, Sloan School of Management, Massachusetts Institute of Technology, November; forthcoming in *Journal of Mathematical Economics.*

Cox, J. C., and C.-f. Huang, 1989b, "Optimal Consumption and Portfolio Policies When Asset Prices Follow a Diffusion Process," working paper, Sloan School of Management, Massachusetts Institute of Technology, December; forthcoming in *Journal of Economic Theory.*

Donaldson, J., M. Rossman, and L. Selden, 1980, "On the Need to Revise Dynamic Consumption Strategies in the Presence of Unchanging Arrow-Debreu Preferences," working paper, Columbia University, June.

Duffie, D., and C.-f. Huang, 1985, "Implementing Arrow-Debreu Equilibria by Continuous Trading of a Few Long-lived Securities," *Econometrica,* 53, 1337–1356.

Dybvig, P. H., and S. A. Ross, 1986, "Tax Clienteles and Asset Pricing," *Journal of Finance,* XLI, 751–762.

Fletcher, R., 1981, *Practical Methods of Optimization, Vol. 2: Constrained Optimization,* John Wiley & Sons, Chichester, U.K.

Fourer, R., 1985, "A Simplex Algorithm for Piecewise-linear Programming I: Derivation and Proof," *Mathematical Programming*, 33, 204–233.

Gill, P. E., W. Murray, and M. Wright, 1981, *Practical Optimization*, Academic Press, London.

Golub, G. H., and C. F. Van Loan, 1983, *Matrix Computations*, The Johns Hopkins University Press, Baltimore.

Hadley, G., 1964, *Nonlinear and Dynamic Programming*, Addison-Wesley, Reading, Massachusetts.

Hakansson, N. H., 1970, "Optimal Investment and Consumption Strategies Under Risk for a Class of Utility Functions," *Econometrica*, 38, 587–607.

Harrison, J. M., and D. M. Kreps, 1979, "Martingales and Multiperiod Securities Markets," *Journal of Economic Theory*, 20, 381–408.

Harrison, J. M., and S. R. Pliska, "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and Their Applications*, 11, 1981, 215–260.

He, H., and N. D. Pearson, 1989, "Consumption and Portfolio Policies With Incomplete Markets and Short-sale Constraints: The Infinite Dimensional Case, working paper, University of California, Berkeley, October.

Huang, C.-f., 1987, "An Intertemporal General Equilibrium Asset Pricing Model: The Case of Diffusion Information," *Econometrica*, 55, 117–142.

Johnsen, T. H., and J. B. Donaldson, 1985, "The Structure of Intertemporal Preferences Under Uncertainty and Time Consistent Plans," *Econometrica*, 53, 1451–1458.

Merton, R. C., 1969, "Lifetime Portfolio Selection Under Uncertainty: The Continuous Time Case," *Review of Economics and Statistics*, LI, 247–257.

Merton, R. C., 1971, "Optimum Consumption and Portfolio Rules in a Continuous Time Model," *Journal of Economic Theory*, 3, 373–413.

Mossin, J., 1968, "Optimal Multiperiod Portfolio Policies," *Journal of Business*, 41, 215–229.

Numerical Algorithms Group Limited, 1982, *The NAG FORTRAN Library Manual - Mark 11*, Numerical Algorithms Group Limited.

Pagès, H., 1987, "Optimal Consumption and Portfolio Policies When Markets are Incomplete," working paper, Massachusetts Institute of Technology, February.

Pliska, S. R., 1982, "A Discrete Time Stochastic Decision Model," in W. Fleming and L. Gorostiza (eds.), *Advances in Filtering and Optimal Stochastic Control: Proceedings of the IFIP-WG 7/1 Working Conference, Cocoyoc, Mexico, February 1-6, 1982*, Springer-Verlag, Berlin.

Pliska, S. R., 1986, "A Stochastic Calculus Model of Continuous Trading: Optimal Portfolios," *Mathematics of Operations Research*, 11, 1986, 371–382.

Samuelson, P. A., 1969, "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *Review of Economics and Statistics*, LI, 239–246.

Schrijver, A., 1986, *Theory of Linear and Integer Programming*, John Wiley & Sons, New York.

Shapiro, J. F., 1979, *Mathematical Programming: Structures and Algorithms*, John Wiley & Sons, New York.

Walsh, G. R., 1975, *Methods of Optimization*, John Wiley & Sons, New York.

Weller, P., 1978, "Consistent Planning Under Uncertainty," *Review of Economic Studies*, 45, 263–266.

CHAPTER 2
AN EMPIRICAL EXAMINATION OF THE COX, INGERSOLL, ROSS MODEL OF THE TERM STRUCTURE OF INTEREST RATES USING THE METHOD OF MAXIMUM LIKELIHOOD

Abstract We estimate the parameters of one and two-factor models of the term structure of interest rates due to Cox, Ingersoll, and Ross (1985b) using the method of maximum likelihood and present tests of the models. We recover the unobservable state variables and exploit their conditional density in estimation and testing, and use both the time-series and cross-sectional information in a sample of bonds that includes coupon bonds. In an initial set of tests, we compare the basic CIR one-factor model of the term structure (interpreted as a model of nominal bond prices) to one of their two-factor models, and compare these to extended (translated) models. We reject the one-factor model in favor of its translated variant, and also reject the CIR one-factor models in favor of the translated two-factor model. However, the fit of the models for long term bonds is relatively poor. We also nest these models within a more general Markov model for yields and reject the restrictions implied by the bond pricing models.

1 Introduction

The term structure of interest rates, the function mapping time to maturity to the prices (or yields) of default-free bonds, has long been of interest to economists. Our contribution to this voluminous literature is to propose and implement a method to estimate and test two of the term structure models of Cox, Ingersoll, and Ross (CIR) (1985b) using the method of maximum likelihood. The key to our approach is a method to recover the unobservable state variables. This enables us to exploit their conditional density in estimation and testing, and to use both the time-series and cross-sectional information in a sample of bonds that includes coupon bonds. We compare the basic CIR one-factor model of the term structure (interpreted as a model of nominal bond prices) to one of their two-factor models, and also compare these to extended (translated) one and two-factor models. We also nest the bond pricing models within a general Markov model for yields and test the restrictions implied by the bond pricing model. Our approach can also be used to estimate and test a number of the "arbitrage" models of the term structure that have appeared in the literature.

There are two motivations for this work. The first of these is very practical. The one and two-factor CIR term structure models are potentially useful for practical purposes such as pricing options on bonds, hedging bond portfolios, and formulating dynamic trading strategies.[1] To use the CIR models for these purposes, one needs reasonably precise estimates of the parameters. The first goal of this work is to provide an estimation technique that exploits both the information in the conditional density of the state variables and the information in the prices of long term (coupon) bonds. The second goal is to assess the performance of the models against a more general alternative. In particular, we find that when we nest the bond pricing models within a general Markov model for yields of discount bonds we reject the restrictions implied by the bond pricing models.

The appeal of the CIR (1985b) models is that they provide closed form expressions for bond prices and endogenous interest rate dynamics (and hence endogenous bond price dynamics) that are supported by an underlying economic equilibrium. Also, interest rates in the CIR models are guaranteed to be non-negative. An alternative "arbitrage" approach of Brennan and Schwartz (1979, 1980, 1982), Dothan (1978), Oldfield and Rogalski (1987), Richard (1978), Schaefer and Schwartz (1984), and Vasichek (1977) involves closing the bond pricing model with an arbitrary assumption about the form of the risk premium investors require for bearing interest rate risk. As pointed out by CIR (1985b), this arbitrage approach provides no way of guaranteeing that the term structure so obtained is supported by any underlying economic equilibrium.[2]

A difficulty in testing the underlying real model is that the underlying state variable, the instantaneous riskless rate of interest, is not observable. This unobservability of the rate of interest makes it difficult to exploit its known conditional density in estimation and testing. For example, Gibbons and Ramaswamy (1986) used the steady state density of the interest rate in their test of the CIR model, while Brown and Dybvig (1986) use neither the conditional nor steady state densities in estimating the parameters of the model (interpreted as a model of nominal bond prices). Similarly, in the two-factor nominal models the drift of the instantaneous rate of inflation is also not observable.

In this paper we use the method of maximum likelihood to estimate one of the nominal models in CIR (1985b) and perform several tests of the model. Our procedure allows us to recover the unobservable state variables, and hence enables us to exploit the known conditional density of the state variables in estimation and testing. We also use the information contained in the prices of a cross-section of bonds observed at each point in time. Our use of the conditional density to extract information from a time series of bond prices combined with the cross-

[1] Of course, to use it for these purposes the one-factor model must be interpreted as a model of the prices of nominal bonds.

[2] This is not to say that there is no such equilibrium. For example, Marsh (1980), Campbell (1986), and we show that the arbitrage model of Vasicek (1977) is consistent with the CIR (1985a) framework. Also, while he does not develop it in an equilibrium framework, the example for which Richard gives a closed form solution is identical in most essentials to one of the two-factor models presented in CIR (1985b).

sectional information in the prices of bonds observed at the same time is a strength of our approach. Our procedure allows us separately to identify *all* of the parameters of the model, and also to use the information in the prices of coupon bonds. Our use of the information contained in the prices of coupon bonds potentially permits powerful tests of the model, because the prices of long term bonds are relatively sensitive to the model parameters. We also apply our procedure to the basic CIR one factor model of the term structure, but interpreted as a model of nominal bond prices.

The tests of the model we offer are of two forms. We begin by nesting the one-factor CIR (1985b) model within an extended (translated) model discussed in CIR (1985b) and Marsh (1980). The extension is simply a translation of the interest rate process (corresponding to a translation of the rates of return on the productive processes), and hence is of limited theoretical interest. However, it provides a more flexible specification of a restrictive feature of the basic CIR model, and hence admits a meaningful test of the model. The inflation rate process can be extended in an identical fashion, providing a bivariate translated process and an extended two-factor model.

Moreover, Ornstein-Uhlenbeck processes[3] can be obtained as special cases of the translated processes. These processes (including their special cases) have served as the workhorse models in the term structure literature,[4] and the bivariate translated process nests a number of the stochastic processes for which closed form bond pricing formulas have been obtained.[5] For example, the extended model includes the models of Vasichek (1977), Schaefer and Schwartz (1984), and Merton (1970), and all of the models in Oldfield and Rogalski (1987), as special cases.

The first set of tests we perform involve using the likelihood ratio to test restrictions on the translated two-factor model, i.e., to select among models for which tractable closed-form solutions are available. We compare the one-factor models to the two-factor models, and compare the basic CIR models to their translated variants. This set of tests addresses the question of which of the tractable models best fits the data.

The set of tractable models does not exhaust the set of all possible models. Our second set of tests addresses the overall performance of the CIR models by examining whether the restrictions on the time series behavior of yields implied by the models are satisfied. In particular, we nest the bond pricing models within a more general Markov model for yields on discount bonds and test whether the restrictions implied by the model are satisfied.

[3]See Feller (1970), p. 335.

[4]See, e.g., Vasichek (1977), Schaefer and Schwartz (1984), Merton (1970), Marsh (1980), Ingersoll, Skelton, and Weil (1978), and Oldfield and Rogalski (1987)).

[5]Exceptions include the two-factor nominal model in CIR (1985b) using the model of the price level that they term "model one," models in which multiple state variables follow correlated Ornstein-Uhlenbeck processes, and the nonlinear general equilibrium model of Longstaff (1989).

A recent test of certain implications of the CIR model is performed by Stambaugh (1988).[6] One implication of the CIR model is that conditional expected holding period returns of discount bonds should be related linearly to forward rates. If bond prices are described by a k-factor model, conditional expected returns should be explained by k forward rates. This leads to a straightforward test of the model when the data set includes enough discount bonds to compute more than k implied forward rates. Stambaugh finds only weak evidence for more than two factors. While Stambaugh's approach is appealling and leads to a straightforward test of the model, limitations are that it may only be applied to discount bonds, the underlying state variables are neither identified nor recovered, and estimates of the parameters of the CIR model are not obtained.

Estimates of (certain combinations of) the parameters of the basic one-factor CIR (1985b) model, interpreted as a model of the prices of nominal bonds, are presented in Brown and Dybvig (1986). Their approach is as follows. They first select a sample of U.S. Treasury issues trading at a number of dates. The prices of these securities are functions of their maturity, their coupon rates and dates, the parameters of the model, and the current instantaneous rate of interest r. Let $\tilde{P}(\tau)$ denote the observed price of a bond with maturity τ, and let $P(\phi, r, \tau)$ denote the price as given by the CIR model when the interest rate is r, where $\phi = (\phi_1, \phi_2, \phi_3)'$, $\phi_1 \equiv \sqrt{(\kappa+\lambda)^2 + 2\sigma^2}$, $\phi_2 \equiv (\kappa + \lambda + \phi_1)/2$, and $\phi_3 \equiv 2\kappa\theta/\sigma^2$. Brown and Dybvig write $\tilde{P}(\tau) = P(\phi, r, \tau) + \epsilon$ and estimate the parameter vector $\phi_1, \phi_2, \phi_3, r$ using nonlinear regression techniques. Brown and Dybvig do not estimate the parameter vector $\kappa, \sigma^2, \theta, \lambda$ of the CIR model because the the parameters κ and λ cannot be identified separately using their procedure. The source of the problem is that the risk premium λ can only separately be identified using time-series information, and Brown and Dybvig do not exploit this information.[7] Our approach to estimation makes use of both the time-series information contained in the distribution of the riskless rate and the information in cross-sectional observations on bond prices, and permits us separately to identify all of the parameters of the models.

Brown and Schaefer (1988) apply the approach of Brown and Dybvig to estimate the model using U.K. government index linked securities.

Gibbons and Ramaswamy (1986) present a test of the CIR model of the prices of real bonds that is robust to the specification of the price level process. They first calculate the ex-post real yields to maturity of a sample of Treasury bills. Given these, the assumption in CIR

[6] Mellino (1986) reviews empirical work on traditional hypotheses about the term structure in the light of current models. Campbell (1986), Mankiw (1986), and Fama (1984a, 1984b) are recent papers which present empirical evidence bearing on the traditional theories. Roll (1970) has tested the pure expectations, liquidity preference, and market segmentation theories, while Roll (1971) has constructed a mean-variance model in an effort to explain the determinants of term premia.

[7] This observation is exactly analogous to the observation that the expected rate of return on a common stock cannot be estimated from the prices of options observed at a single point in time.

(1985b) that the price level process is independent of the real economy allows Gibbons and Ramaswamy to estimate certain of the moments of the distribution of real yields on discount bonds, and obtain the parameters of the underlying real model, using the generalized method of moments (Hansen (1982)). The advantage of this approach is that it permits estimation of the real model without any assumption about the price level process besides the independence assumption made in CIR (1985b). A disadvantage is that this approach allows them to make use only of moments of the steady state density of the interest rate. Further, their procedure requires that they possess a sequence of observations on bonds with the same terms, i.e., the same time to maturity and coupon rate. In practice, this restriction permits them to use only the information contained in the prices of short term discount bonds. While Treasury notes and bonds currently are issued on a regular cycle, different issues have different coupon rates, only short term discount bonds have been widely traded for any length of time, and there are not enough coupon bonds to construct the implied prices of pure discount bonds without interpolation or extrapolation. The ability to use only short term bonds is a serious drawback, because compared to longer term bonds, the prices of such bonds are relatively insensitive to the model parameters.

Heston (1988) employs a similar estimation and testing strategy using the real holding period returns of a sample of Treasury bills.

The relative merits of our appoach and that of Gibbons and Ramaswamy (1986) depend upon how one views the tradeoff between efficiency and specification errors. Our approach exploits the conditional density of the underlying state variables and makes use of the information in the prices of coupon bonds. The tradeoff is that we need to model explicitly the price level process. This requirement may impede testing of the underlying real model. However, we accept this possibility in exchange for the advantages that our approach provides. Moreover, for practical purposes such as hedging bond portfolios and formulating dynamic trading strategies it is necessary to model explicitly the price level process.

The balance of the paper is organized as follows. In the next section we list the assumptions and certain results of the CIR model, and derive the translation of the basic CIR model discussed in CIR (1985b) and Marsh (1980). We present the derivation of the translated model because it appears that the existence of the extension is not well known.[8] With this, we have a model that nests a number of the closed form bond pricing formulas in the literature. Section 3 describes the estimation methodology and the data set used in estimating the parameters of the models. Our first set of tests in which we select among the tractable models is described in Section 4, while we present the results bearing upon the overall performance of the models in Section 5.

[8]See, for example, Campbell (1986), where it is shown that the Vasichek (1977) model is consistent with a special case of the CIR (1985a) asset pricing model. The Vasichek model is a special case of the translated model discussed in Marsh (1980) and CIR (1985b) and derived here.

We conclude in Section 6.

2 The Cox, Ingersoll, Ross Model

In this section we list the assumptions and certain results of the CIR model of the term structure of interest rates, and present a simple translation of the basic CIR (1985b) model discussed in CIR and Marsh (1980). The extension is useful in testing the model, and may be of value in applications. Furthermore, it nests as special cases a number of the bond pricing models for which tractable closed form solutions for the bond price are known.[9]

2.1 The One-Factor Model

In developing a general theory of asset pricing with diffusion information, CIR (1985a) make (among others) the following assumptions:

A1. There is a single physical good which may be allocated to consumption or investment.

A2. Production possibilities consist of a set of n linear activities. The returns to investments η in these activities are given by a system of stochastic differential equations of the form

$$d\eta(t) = I_\eta \alpha(Y,t)dt + I_\eta G(Y,t)dW(t), \tag{1}$$

where $W(t)$ is an $(n+k)$-dimensional Brownian motion in $\Re^{n+k}$, Y is a k-dimensional vector of state variables, I_η is a $n \times n$ diagonal matrix valued function of η whose i-th diagonal element is the i-th component of η, $\alpha(Y,t)$ is an n-dimensional vector valued function of Y and t, and $G(Y,t)$ is an $n \times (n+k)$ matrix valued function of Y and t. The covariance matrix of physical rates of return GG' is positive definite.

A3. The movement of the k-dimensional vector of state variables, Y, is given by a system of stochastic differential equations of the form

$$dY(t) = \mu(Y,t)dt + S(Y,t)dW(t),$$

where $\mu(Y,t)$ is a k-dimensional vector and $S(Y,t)$ is a $k \times (n+k)$ dimensional matrix. The covariance matrix of changes in the state variables, SS', is non-negative definite.

A4. There is a market for instantaneous borrowing and lending at an interest rate r. The market clearing interest rate, as a function of underlying variables, is determined as part of the competitive equilibrium of the economy.

[9] Some models it does not subsume are listed in footnote 5.

A5. There are a fixed number of identical individuals, each of whom seeks to maximize an objective function of the form $E\int_t^{t'} U(C(s),Y(s),s)ds$, where U is a von Neumann-Morgenstern utility function.

A6. Physical investment and trading in claims takes place continuously in time with no adjustment or transactions costs.

CIR (1985a) show that the equilibrium interest rate can be written

$$r = a^{*\prime}\alpha + a^{*\prime}GG'a^{*}W\left(\frac{J_{WW}}{J_W}\right) + a^{*\prime}GS'a^{*}W\left(\frac{J_{WY}}{J_W}\right),$$

where α is the vector of the expected rates of return on the n linear activities, GG' is the covariance matrix of the rates of return, GS' gives the covariances among the rates of return on the linear activities and the changes in the state variables, $J(W,Y)$ is the indirect utility function, and a^* represents the optimal proportion of wealth W to be invested in each of the productive processes. CIR show

$$a^* = (GG')^{-1}\alpha + \left(\frac{1 - 1'(GG')^{-1}\alpha}{1'(GG')^{-1}1}\right)(GG')^{-1}1.$$

CIR also show that the price of any contingent claim satisfies a particular partial differential equation. We record a special case of this equation below.

CIR (1985b) specialize the model by assuming log utility, which gives $\frac{WJ_{WW}}{J_W} = -1$ and $\frac{WJ_{WY}}{J_W} = 0$, so $r = a^{*\prime}\alpha - a^{*\prime}GG'a^*$. The partial differential equation satisfied by the price F of any contingent claim the payouts of which do not depend on wealth becomes

$$\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\text{cov}(Y_i,Y_j)F_{Y_iY_j} + \sum_{i=1}^{k}F_{Y_i}(\mu_i + \frac{1}{W}\text{cov}(W,Y_i)) + F_t - rF + \delta = 0,$$

where δ is the payout rate of the claim and $\text{cov}(W,Y_i)$ represents the local covariance of wealth W and the state variable Y_i.

CIR also assume that the means and covariances of the rates of return on the productive processes are proportional to a single state variable Y. The development of the state variable Y is described by a stochastic differential equation

$$dY(t) = (\xi Y(t) + \zeta)dt + \nu\sqrt{Y(t)}dW(t). \tag{2}$$

where ν is a $1 \times (n+1)$-vector. Write $\Omega Y = GG'$ and $\hat{\alpha} Y = \alpha$. Then $a^* = \Omega^{-1}\hat{\alpha} + \left(\frac{1-1'\Omega^{-1}\hat{\alpha}}{1'\Omega^{-1}1}\right)\Omega^{-1}1$ and

$$\begin{aligned} r(t) &= (a^{*\prime}\hat{\alpha} - a^{*\prime}\Omega^{-1}a^*)Y(t) \\ &= \left(\frac{1'\Omega^{-1}\hat{\alpha} - 1}{1'\Omega^{-1}1}\right)Y(t). \end{aligned} \tag{3}$$

From (2) and (3), the interest rate r follows a "square root" diffusion process

$$dr(t) = \kappa(\theta - r(t))dt + \sigma\sqrt{r(t)}dZ_1(t). \tag{4}$$

where $Z_1(t)$ is a linear combination of components of $W(t)$. The factor risk premium is $a^{*\prime}GS' = a^{*\prime}\Sigma Y \equiv \lambda r$, where $GS' = \Sigma Y$.

Given this, the fundamental equation for the value of a discount bond with maturity τ, $P(\tau)$, becomes

$$1/2\sigma^2 rP_{rr} + \kappa(\theta - r)P_r - \lambda rP_r - P_\tau - rP = 0$$

with boundary condition $P(0) = 1$. The solution to this partial differential equation with the given boundary condition is given in CIR (1985b) as

$$P(\tau) = A(\tau)e^{-B(\tau)r}, \tag{5}$$

where

$$A(\tau) \equiv \left[\frac{2\gamma e^{(\gamma+\lambda+\kappa)\tau/2}}{(\gamma+\lambda+\kappa)(e^{\gamma\tau}-1)+2\gamma}\right]^{2\kappa\theta/\sigma^2}, \tag{6}$$

$$B(\tau) \equiv \frac{2(e^{\gamma\tau}-1)}{(\gamma+\lambda+\kappa)(e^{\gamma\tau}-1)+2\gamma}, \tag{7}$$

$$\gamma \equiv ((\kappa+\lambda)^2 + 2\sigma^2)^{1/2}. \tag{8}$$

The extension of the model is to assume that the means of the rates of return on the productive processes are linear functions of a single state variable Y'; i.e., they are proportional to Y' plus a constant $\overline{Y}$. The movement of the state variable Y' is described by the stochastic differential equation

$$dY'(t) = (\xi Y'(t) + \zeta)dt + \nu\sqrt{Y'(t)}dW(t). \tag{9}$$

Now writing $\Omega Y' = GG'$ and $\hat{\alpha}(Y' + \overline{Y}) = \alpha$, we obtain

$$r(t) = \left(\frac{1'\Omega^{-1}\hat{\alpha} - 1}{1'\Omega^{-1}1}\right)Y'(t) + \left(\frac{1'\Omega^{-1}\hat{\alpha}}{1'\Omega^{-1}1}\right)\overline{Y}. \tag{10}$$

Following CIR (1985b), we assume that $1'\Omega^{-1}\hat{\alpha} > 1$. The interest rate r can be written $r = r' + \bar{r}$, where $\bar{r} \equiv \left(\frac{1'\Omega^{-1}\hat{\alpha}}{1'\Omega^{-1}1}\right)\overline{Y}$ and r' follows a square root process

$$dr'(t) = \kappa(\theta - r'(t))dt + \sigma\sqrt{r'(t)}dZ_1(t). \tag{11}$$

If we maintain the assumption that the covariances among the rates of return on the n linear activities and the state variable Y' are proportional to Y', and further assume that[10]

$$\hat{\alpha}'\Omega^{-1} - \frac{1'\Omega^{-1}\hat{\alpha}'\Omega^{-1}1}{1'\Omega^{-1}1} = 0, \tag{12}$$

we again obtain a constant factor risk premium $a^{*\prime}GS' = a^{*\prime}\Sigma Y' \equiv \lambda r'$. Given this, the equation for the value of a discount bond becomes

$$1/2\sigma^2 r' P_{r'r'} + \kappa(\theta - r')P_{r'} - \lambda r' P_{r'} - P_\tau - (r' + \bar{r})P = 0$$

with boundary condition $P(0) = 1$, and solution

$$P(\tau) = A(\tau)e^{-B(\tau)r' - \bar{r}\tau}, \tag{13}$$

where

$$A(\tau) \equiv \left[\frac{2\gamma e^{(\gamma+\lambda+\kappa)\tau/2}}{(\gamma+\lambda+\kappa)(e^{\gamma\tau}-1)+2\gamma}\right]^{2\kappa\theta/\sigma^2}, \tag{14}$$

$$B(\tau) \equiv \frac{2(e^{\gamma\tau}-1)}{(\gamma+\lambda+\kappa)(e^{\gamma\tau}-1)+2\gamma}, \tag{15}$$

$$\gamma \equiv ((\kappa+\lambda)^2 + 2\sigma^2)^{1/2}. \tag{16}$$

The formula is more flexible than the basic CIR model (5)–(8), and includes the basic CIR formula as a special case when $\bar{r} = 0$.[11]

2.2 The Two-Factor Model

CIR (1985b) extend their model to the pricing of nominal bonds in two ways, one of which (the model of the price level they term "model 2") is as follows. They suppose that the expected rate of inflation y follows a process

$$dy(t) = \kappa_2(\theta_2 - y(t))dt + \sigma_2\sqrt{y(t)}dZ_2(t),$$

and the price level follows the process

$$dp(t) = y(t)p(t)dt + \sigma_p p(t)\sqrt{y(t)}dZ_3(t),$$

[10] In this setting, equation (12) is equivalent to the assumption that the set of active technologies does not change. It is admittedly a strong restriction on $\hat{\alpha}$ and Ω, and is discussed in the appendix.

[11] The extended one-factor model can also be viewed as the original CIR one-factor model combined with a constant inflation rate $\bar{r}$.

with $\text{cov}(y,p) = \rho\sigma_2\sigma_p yp$.[12] With this model of the price level the partial differential equation for the real price $N(\tau)$ of a nominal bond is

$$1/2\sigma^2 r N_{rr} + 1/2\sigma_2^2 y N_{yy} + \rho\sigma_2\sigma_p yp N_{yp} + \sigma_p^2 yp^2 N_{pp}$$
$$+(\kappa\theta - (\kappa+\lambda)r)N_r + \kappa_2(\theta_2 - y)N_y + ypN_p - N_\tau - rN = 0$$

with boundary condition $N(0) = 1/p(T)$, where $p(T)$ denotes the price level at the maturity of the bond. The solution is (CIR (1985b))

$$N(\tau) = C(\tau)e^{-D(\tau)y}P(\tau)/p(t), \tag{17}$$

where

$$C(\tau) \equiv \left[\frac{2\xi e^{(\xi+\kappa_2+\rho\sigma_2\sigma_p)\tau/2}}{(\xi+\kappa_2+\rho\sigma_2\sigma_p)(e^{\xi\tau}-1)+2\xi}\right]^{2\kappa_2\theta_2/\sigma_2^2}, \tag{18}$$

$$D(\tau) \equiv \frac{2(e^{\xi\tau}-1)(1-\sigma_p^2)}{(\xi+\kappa_2+\rho\sigma_2\sigma_p)(e^{\xi\tau}-1)+2\xi}, \tag{19}$$

$$\xi \equiv ((\kappa_2+\rho\sigma_2\sigma_p)^2 + 2\sigma_2^2(1-\sigma_p^2))^{1/2}. \tag{20}$$

We also use $P(\tau)$ to denote the nominal bond price, and have

$$P(\tau) = N(\tau)p(t). \tag{21}$$

We can also translate the inflation rate process in the same way we translated the real model, but do not do so here, or report results for a model with both state variable processes translated, because the data do not permit us to distinguish the model with a translated inflation rate from the model with only the real interest rate process translated.

The extended model, even though it admits negative nominal interest rates when $\bar{r} < 0$, has a particularly interesting form. A restrictive feature of the basic one and two-factor CIR models is that the local variance of the instantaneous real riskless interest rate is proportional to its level, i.e., $\text{var}(r) = \sigma^2 r$. The extended model weakens this assumption by allowing the local variance of the instantaneous riskless rate of interest to be a linear function of the instantaneous rate of interest, i.e., $\text{var}(r) = \sigma^2 r' = \sigma^2 r - \sigma^2 \bar{r}$. With the extended models, $\bar{r}$, rather than zero, provides a lower bound for the first state variable.

To see why the translated model is of interest, consider the CIR square root process (4) for r, and suppose that κ is large. In this case, values of r close to zero are extremely unlikely; as

[12] In the results reported below we set $\sigma_p = \rho = 0$ because this restriction is never rejected by the data.

r approaches zero, the diffusion coefficient $\sigma\sqrt{r}$ approaches zero, and there is a strong upward drift. In the translated process with $\bar{r} < 0$, while there is still a strong upward drift when r approaches zero, the diffusion coefficient does not approach zero. This makes small values of r much more likely. Put differently, for a given value of r, the conditional density for r in the basic CIR model with $\bar{r} = 0$ assigns less probability to values of r near zero than does the density in the translated model, holding constant κ, the steady state mean $\theta + \bar{r}$, and the variance of the steady state distribution of the interest rate.

This feature of the translated process turns out to be empirically relevant. In these models, steep[13] upward sloping yield curves on some days, together with steep downward sloping yield curves on others, are possible only if κ (or κ_2) is large. Steeply upward sloping yield curves occur when r is considerably below its long run mean $\theta + \bar{r}$, while downward sloping yield curves occur when r sufficiently above its long run mean. If $\bar{r} = 0$, the extent to which r can be below a fixed mean $\theta + \bar{r}$ is limited by the fact that 0 is a lower bound for r, and therefore the steepness of the yield curve is limited. Moreover, values of r near zero are unlikely. If $\bar{r} < 0$, it is possible to have steeper upward sloping yield curves for the same value of $\theta + \bar{r}$, and a curve of a given steepness is much more likely. We find in the data a number of steep upward sloping yield curves along with downward sloping yield curves, and when we constrain $\bar{r} = 0$ we cannot find a set of parameter values such that $r(t)$ and $y(t)$ are both greater than zero for all times t.

In short, we advance the extended model as an alternative not because we necessarily believe that the variance of the instantaneous rate of interest is related linearly to the level of the rate, though of course it may be, but because the extended model may provide a better approximation to the true (presumably nonlinear) function, and a better approximation to the conditional density of the state variables.

Morever, our translated model for the interest rate process includes Ornstein-Uhlenbeck processes of the form

$$dr(t) = \kappa(\hat{\theta} - r(t))dt + \hat{\sigma} dZ_1(t), \tag{22}$$

as special cases. To see this, note that the process for the interest rate r in the translated model can be written

$$dr(t) = dr'(t) = \kappa(\hat{\theta} - r(t))dt + \sqrt{(\sigma^2 r(t) + \hat{\sigma}^2)} dZ_1(t),$$

where $\hat{\theta} \equiv \theta + \bar{r}$ and $\hat{\sigma}^2 \equiv -\sigma^2 \bar{r}$. Letting $\sigma \to 0$, $\bar{r} \to -\infty$, and $\theta \to \infty$ at rates such that $\hat{\theta} = \theta + \bar{r}$ and $\hat{\sigma}^2 = -\sigma^2 \bar{r}$ are constant, in the limit we obtain an Ornstein-Uhlenbeck process for the real interest rate r.

[13] For the purposes of this discussion, the "steepness" of the yield curve is measured by the difference in yields of 3 and 6 month Treasury bills.

To see what this limiting process implies about the underlying equilibrium model, write (1) and (9) as

$$
\begin{aligned}
d\eta(t) &= I_\eta \hat{\alpha} Y(t)dt + I_\eta H\sqrt{Y(t) - \overline{Y}(t)}dW(t), \\
dY(t) = dY'(t) &= \xi(Y(t) - (\overline{Y} - \zeta/\xi))dt + \nu\sqrt{(Y(t) - \overline{Y}(t))}dW(t),
\end{aligned}
$$

where $Y(t) = Y'(t) + \overline{Y}$ and $H = \{h_{ij}\}$ is an $n \times (n+k)$ matrix such that $H\sqrt{Y'} = G$. Letting the $h_{ij} \to 0$, $\overline{Y} \to -\infty$, $\nu_j \to 0$, and $\zeta \to \infty$ at rates such that $-HH'\overline{Y}$, $-\nu\nu'\overline{Y}$, and $(\overline{Y} - \zeta/\xi) \equiv \hat{\zeta}$ are constant, the processes become

$$
\begin{aligned}
d\eta(t) &= I_\eta \hat{\alpha} Y(t)dt + I_\eta H\sqrt{-\overline{Y}(t)}dW(t), \\
dY(t) &= \xi(Y(t) - \hat{\zeta})dt + \nu\sqrt{-\overline{Y}(t)}dW(t).
\end{aligned}
$$

The equalities $r = a^{*\prime}\alpha - a^{*\prime}GG'a^*$ and $a^* = (GG')^{-1} + \left(\frac{1 - 1'(GG')^{-1}1}{1'(GG')^{-1}1}\right)(GG')^{-1}1$ yield (22), and the factor risk premium is constant as in Vasichek's model if a condition analogous to (12) is imposed.

Hence, although it is not obvious from simple comparison of the bond pricing formulas, the Vasichek (1977) model in which the instantaneous rate of interest follows an Ornstein-Uhlenbeck process is a special case of the translated CIR model. Schaefer and Schwartz (1984) have a model in which the two state variables follow an Ornstein-Uhlenbeck process and a square-root process, respectively. While the state variables in their model have different interpretations than the corresponding state variables in the CIR model, their approximate analytical solution is identical to the translated two-factor bond pricing formula. Also, all of the interest rate processes considered in Oldfield and Rogalski (1987) are special cases of the translated processes here.

3 Estimation Approach and Data

We first present our estimation approach in the context of the basic one-factor CIR model (interpreted as a model of nominal bond prices), and then extend our approach to the estimation of the two-factor model. Once we have presented the approach for the two-factor CIR model, the adaptation of the approach to the translated one and two-factor models, and for other models of bond prices, should be clear.

When estimating the two-factor model, we do not explicitly link the expected rate of inflation y and observations on the price level. That is, we do not make use of data on the price level (CPI) and the conditional density for the price level in estimating the model. One reason for not using price level data is that some of the price level data are collected in the middle of the month, and are therefore not aligned with the bond price data, which are from the last business day of the month. Appending an appropriate measurement error model to the price level data might allow one to interpret the actual CPI data, measured at mid-month, as the end-of-month price level plus a measurement error. However, appending a measurement error model to the price level process effectively prevents one from learning about the price level process from the CPI data.

Even were the data alignment not an issue, no analytic expression for the conditional joint density of y and p is known. Obtaining the density via numerical solution of the backward equation would involve a prohibitive amount of computation, as the backward equation is in two space variables.

Due to the fact that our estimation procedure does not explicitly associate y with the price level, we interpret our tests as tests of a "necessary condition" for the CIR model. That is, if the CIR model is to hold when y and p are explicity associated with price level data, it must of necessity hold when we do not link y and p with the price level data. Of course, the two-factor model (17)–(20) may fit the bond price data, even though the estimated process for the price level does not adequately describe the price level data.

Below we find it convenient to work with (negative of) the natural logarithms of the bond prices instead of the prices themselves, and define $Y(\tau_j) \equiv -\ln P(\tau_j)$. Also, throughout most of this section, the exposition is as though the data consists of the prices of bonds for cash settlement. In actuality, the prices in our data set are quoted on the basis of settlement in two business days ("skip-day" settlement). We indicate the appropriate adjustment for the settlement terms at the end of the section.

3.1 The One-Factor Models

Our approach is constructed from two components. The first is the known conditional density for the instantaneous rate of interest r. Using r_t as an alternative notation for $r(t)$, the density for r_s conditional on r_t, for $s > t$, is (CIR (1985b))

$$f(r_s \mid r_t) = ce^{-u-v}\left(\frac{v}{u}\right)^{q/2} I_q(2(uv)^{1/2}), \tag{23}$$

where

$$\begin{aligned} c &\equiv \frac{2\kappa}{\sigma^2(1-e^{-\kappa(s-t)})}, \\ u &\equiv cr_t e^{-\kappa(s-t)}, \\ v &\equiv cr_s, \end{aligned}$$

$$q \equiv \frac{2\kappa\theta}{\sigma^2} - 1,$$

and $I_q(\cdot)$ denotes the modified Bessel function of the first kind of order q.[14]

Given this conditional density for r_s and the bond pricing function (5)–(8), it is straightforward to obtain the density of the (negative) log price of a discount bond with maturity τ_1 at time s, $Y_s(\tau_1)$, conditional on the log price $Y_t(\tau_1)$ of a discount bond with maturity τ_1 at time t. We can recover the interest rate r_s from the price of the bond, obtaining

$$\begin{aligned} r_s &= \frac{\ln A(\tau_1) - \ln P_s(\tau_1)}{B(\tau_1)} \\ &= \frac{\ln A(\tau_1) + Y_s(\tau_1)}{B(\tau_1)}. \end{aligned}$$

The density for $Y_s(\tau_1)$ conditional on $Y_t(\tau_1)$ is

$$f(Y_s(\tau_1) \mid Y_t(\tau_1)) = ce^{-u-v}\left(\frac{v}{u}\right)^{q/2} I_q(2(uv)^{1/2})\frac{1}{B(\tau_1)}, \tag{24}$$

where

$$\begin{aligned} c &\equiv \frac{2\kappa}{\sigma^2(1-e^{-\kappa(s-t)})}, \\ u &\equiv cr_t e^{-\kappa(s-t)}, \\ v &\equiv cr_s, \\ q &\equiv \frac{2\kappa\theta}{\sigma^2} - 1, \\ r_t &= \frac{\ln A(\tau_1) + Y_t(\tau_1)}{B(\tau_1)}, \\ r_s &= \frac{\ln A(\tau_1) + Y_s(\tau_1)}{B(\tau_1)}. \end{aligned}$$

The term $\frac{1}{B(\tau_1)}$ in (24) is the absolute value of the determinant of the Jacobian of the transformation from $Y_s(\tau_1)$ to r_s.

We can now use the method of maximum likelihood to take full advantage of the probability distribution of the instantaneous interest rate r. As long as we are willing to interpret the basic CIR one-factor model as a model of the prices of nominal bonds, the conditional density (24) enables us to apply the method of maximum likelihood to a time series of observations $\{Y_s(\tau_1)\}$.

[14]See Oliver (1965) for the properties of the modified Bessel function.

If we suppose that there are N observations available at times $t_1, t_2, \dots, t_N$, the logarithm of the likelihood function is

$$\begin{aligned}\mathcal{L}(\kappa, \sigma^2, \theta, \lambda) &\equiv \ln f(Y_{t_2}(\tau_1), Y_{t_3}(\tau_1), \dots, Y_{t_N}(\tau_1) \mid Y_{t_1}(\tau_1)) \\ &= \ln \prod_{i=2}^{N} f(Y_{t_i}(\tau_1) \mid Y_{t_{i-1}}(\tau_1)) \\ &= \sum_{i=2}^{N} \ln f(Y_{t_i}(\tau_1) \mid Y_{t_{i-1}}(\tau_1)),\end{aligned}$$

where the second equality follows from the fact that the process for the interest rate r is Markov. However, at any instant of time this procedure makes use of only one point on the yield curve. The information contained in the time series of this point is limited, and it seems unlikely that it contains enough information to provide parameter estimates consistent with the entire yield curve.

However, we can incorporate the information in a cross-section of bond prices observed at each point in time using a statistical model for the prices of bonds. The second building block of our approach is just such a statistical model. Suppose that at each point in time we observe the prices of a sample of l discount bonds and $m-l$ coupon bonds. Letting $\widetilde{P}_{t_i}(\tau_j)$ denote the observed price of a bond with maturity τ_j at the i-th time t_i and $P_{t_i}(\tau_j)$ denote the price predicted by the CIR model, we write the observed prices of the discount bonds

$$\widetilde{P}_{t_i}(\tau_j) = P_{t_i}(\tau_j) e^{-\epsilon_{j,t_i}}, \quad j = 1, \dots, l, \quad i = 1, \dots, N. \tag{25}$$

The properties of the residuals ϵ_{j,t_i} are specified below.

The price of a coupon bond is simply the price of a portfolio of discount bonds, i.e.,

$$Q(\tau_j) = \sum_{h=1}^{H_j} c(\tau_h) P(\tau_h) + P(\tau_j), \quad j = l+1, \dots, m,$$

where $Q(\tau_j)$ denotes the price of the coupon bond with maturity τ_j, the τ_h, $h = 1, 2, \dots, H_j$ are the times until the coupon payment dates, and $c(\tau_h)$ is the coupon paid at τ_h. We suppose that the observed prices $\widetilde{Q}(\tau_j)$ of the $m-l$ coupon bonds satisfy a statistical model

$$\widetilde{Q}_{t_i}(\tau_j) = Q_{t_i}(\tau_j) e^{-\epsilon_{j,t_i}}, \quad j = l+1, \dots, m, \quad i = 1, \dots, N. \tag{26}$$

Letting $\widetilde{Y}_{t_i}(\tau_j) \equiv -\ln \widetilde{P}_{t_i}(\tau_j)$ for $j = 1, \dots, l$ and $\widetilde{Y}_{t_i}(\tau_j) \equiv -\ln \widetilde{Q}_{t_i}(\tau_j)$ for $j = l+1, \dots, m$, we have

$$\widetilde{Y}_{t_i}(\tau_j) = Y_{t_i}(\tau_j) + \epsilon_{j,t_i} \quad j = 1, \dots, m, \quad i = 1, \dots, N.$$

Some such statistical model for the prices of bonds is clearly needed, because the CIR model implies an exact linear relation among the yields on discount bonds with different maturities, and no exact linear relation appears in the data. In other words, a strict interpretation of the CIR model is that the residuals ϵ_{j,t_i} are identically zero. Based on such a strict interpretation of the CIR model, our need to rely on the statistical model above would lead to an immediate rejection of the model. However, it seems unreasonable to reject the model on these grounds. As argued by Stambaugh (1988), possible quotation errors, the averaging of bid and ask prices, and other possible problems in the data, would alone lead to a rejection of the CIR model based on its strictest interpretation. Even if the model were literally true as a description of equilibrium prices, anything that caused observed prices to deviate from equilibrium prices would lead to an immediate rejection of the model.

More importantly, the CIR model is only a model. Being so, it is surely misspecified, even as a description of equilibrium prices, and it is unreasonable to expect that the data satisfy its strictest implications. With this, the ϵ_j are to be interpreted as due to omitted factors, perhaps including measurement error. We assume that the ϵ_j are conditionally normally distributed,[15] and we also assume that the ϵ_j are independent. The normality assumption is for statistical convenience. We maintain the independence assumption because this is in some sense the "closest" we can get to the strict implications of the CIR model. That is, the strongest assumption that is not obviously violated is that the omitted factors are independent. Among other things, this excludes the presence of additional state variables.

With the statistical model above, we cannot proceed with estimation without some further restriction on the ϵ_j. More precisely, we need to impose a linear restriction on the ϵ_j in order to recover the unobservable interest rate r. Any linear restriction on the ϵ_j's will allow us to do so. The simplest approach is to assume that one of the ϵ_j is identically zero, i.e., that the observed price of one of the bonds is equal to its equilibrium price. This is perfectly reasonable, provided that we exercise some care in selecting the bond. We include in our sample observations on the (just auctioned) 13-week bill and the (just auctioned) 26-week bill. These bills are actively traded, or traded "on-the-run." When estimating the one-factor models, we assume that the observed price of the 13-week bill is equal to its equilibrium price, while in estimating two-factor models we assume that the observed prices of both the 13 and 26-week bills are equal to the predicted prices. Given the trading activity, number of participants in the market for these securities, and low costs of transacting in these bills, it seems reasonable to assume that the observed prices are adequate approximations of the equilibrium prices that would be observed in a market without transactions costs.[16]

[15] A similar assumption is made by Brown and Dybvig (1986). They append an additive normally distributed error to the bond pricing model.

[16] This assumption is also made by Gibbons and Ramaswamy (1986). Morever, if the implications of equilibrium

For $j = 2, \ldots, m$ and $i = 1, \ldots, N$, we assume that the errors are independent of the yield $\widetilde{Y}_{t_i}(\tau_1)$ on the 13-week bill with maturity τ_1, and joint normal, and we denote their conditional density $f(\epsilon_{2,t_i}, \ldots, \epsilon_{m,t_i} \mid \epsilon_{2,t_{i-1}}, \ldots, \epsilon_{m,t_{i-1}})$. With this stochastic specification of the errors, we can apply the method of maximum likelihood to the m time series of different yields at the same time.

We let $\underline{\widetilde{Y}}_{t_i}$ denote the vector of yields $(\widetilde{Y}_{t_i}(\tau_1), \widetilde{Y}_{t_i}(\tau_2), \ldots, \widetilde{Y}_{t_i}(\tau_m),)$, and let ϵ_{t_i} denote the vector of errors $(\epsilon_{2,t_i}, \epsilon_{3,t_i}, \ldots, \epsilon_{m,t_i})$. The joint density of $\underline{\widetilde{Y}}_{t_i}$ conditional on $\underline{\widetilde{Y}}_{t_{i-1}}$ is

$$\begin{aligned} f(\underline{\widetilde{Y}}_{t_i} \mid \underline{\widetilde{Y}}_{t_{i-1}}) &= f(\underline{\widetilde{Y}}_{t_i} \mid \widetilde{Y}_{t_{i-1}}(\tau_1), \epsilon_{t_{i-1}}) \\ &= f(\widetilde{Y}_{t_i}(\tau_1), \epsilon_{t_i} \mid \widetilde{Y}_{t_{i-1}}(\tau_1), \epsilon_{t_{i-1}}) \times \left| \frac{\partial(\widetilde{Y}_{t_i}(\tau_1), \epsilon_{t_i})}{\partial \underline{\widetilde{Y}}_{t_i}} \right| \\ &= f(\widetilde{Y}_{t_i}(\tau_1), \epsilon_{t_i} \mid \widetilde{Y}_{t_{i-1}}(\tau_1), \epsilon_{t_{i-1}}) \\ &= f(\widetilde{Y}_{t_i}(\tau_1), \mid \widetilde{Y}_{t_{i-1}}(\tau_1),) \times f(\epsilon_{t_i} \mid \epsilon_{t_{i-1}}). \end{aligned}$$

The last equality follows from the assumption that the errors are independent of the yield on the 13-week bill, and the third follows from

$$\begin{aligned} \left| \frac{\partial(\widetilde{Y}_{t_i}(\tau_1), \epsilon_{t_i})}{\partial \underline{\widetilde{Y}}_{t_i}} \right| &= \begin{vmatrix} 1 & 0 & \ldots & 0 \\ -\beta_2 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\beta_m & 0 & \ldots & 1 \end{vmatrix} \\ &= 1, \end{aligned}$$

where $\beta_j \equiv -B(\tau_j)/B(\tau_1)$.

If the errors have mean zero and covariance matrix Σ_ϵ, the logarithm of the likelihood function of the $\widetilde{Y}$ is

$$\begin{aligned} \mathcal{L}(\kappa, \sigma^2, \theta, \lambda, \Sigma_\epsilon) &\equiv \ln f(\underline{\widetilde{Y}}_{t_2}, \underline{\widetilde{Y}}_{t_3}, \ldots, \underline{\widetilde{Y}}_{t_N} \mid \underline{\widetilde{Y}}_{t_1}) \\ &= \ln \prod_{i=2}^{N} f(\underline{\widetilde{Y}}_{t_i} \mid \underline{\widetilde{Y}}_{t_{i-1}}) \\ &= \sum_{i=2}^{N} \ln f(\underline{\widetilde{Y}}_{t_i} \mid \underline{\widetilde{Y}}_{t_{i-1}}) \\ &= \sum_{i=2}^{N} \ln f(\widetilde{Y}_{t_i}(\tau_1) \mid \widetilde{Y}_{t_{i-1}}(\tau_1)) + \sum_{i=2}^{N} \ln f(\epsilon_{t_i} \mid \epsilon_{t_{i-1}}). \end{aligned}$$

asset pricing theories do not apply to the observed prices of the most actively traded securities, it is not clear that they apply to any observables.

The maximization problem is not as difficult as it might seem, because it is possible to concentrate the likelihood function on the parameters κ, σ^2, θ and λ.

3.2 The Two-Factor Model

Our approach to estimating the two-factor model is a simple extension of the the approach we use to estimate the one-factor model. In estimating the two-factor model we also use the conditional density for the instantaneous drift in the price level. The density for y_s conditional on y_t has the same form as the density for r_s conditional on r_t, and is

$$f(y_s \mid y_t) = c_2 e^{-u_2 - v_2} \left(\frac{v_2}{u_2}\right)^{q_2/2} I_{q_2}(2(u_2 v_2)^{1/2}) \tag{27}$$

where

$$\begin{aligned} c_2 &\equiv \frac{2\kappa_2}{\sigma_2^2(1 - e^{-\kappa_2(s-t)})}, \\ u_2 &\equiv c_2 y_t e^{-\kappa_2(s-t)}, \\ v_2 &\equiv c_2 y_s, \\ q_2 &\equiv \frac{2\kappa_2\theta_2}{\sigma_2^2} - 1. \end{aligned}$$

This, together with the density for the instantaneous interest rate r given by (23) and the statistical model for bond prices (25) and (26) are the building blocks of our approach to estimating the two factor model.

As we now want to recover two unobservable state variables, r_{t_i} and y_{t_i}, from the observed prices of the bonds for each time t_i, we must impose two linear restrictions on the errors ϵ_{j,t_i}. As discussed above, we recover the interest rate r_{t_i} and the drift in the price level y_{t_i} by assuming that the observed prices of the 13 and 26-week bills $\widetilde{P}_{t_i}(\tau_1)$ and $\widetilde{P}_{t_i}(\tau_2)$ are equal to their equilibrium prices. With this assumption, we can obtain the interest rate and the drift in the price level

$$\begin{aligned} r_{t_i} &= \left(D(\tau_2)\widetilde{Y}_{t_i}(\tau_1) - D(\tau_1)\widetilde{Y}_{t_i}(\tau_2) + D(\tau_2)\ln(A(\tau_1)C(\tau_1)) - D(\tau_1)\ln(A(\tau_2)C(\tau_2))\right)/F, \\ y_{t_i} &= \left(-B(\tau_2)\widetilde{Y}_{t_i}(\tau_1) + B(\tau_1)\widetilde{Y}_{t_i}(\tau_2) - B(\tau_2)\ln(A(\tau_1)C(\tau_1)) + B(\tau_1)\ln(A(\tau_2)C(\tau_2))\right)/F, \end{aligned}$$

where $F \equiv B(\tau_1)D(\tau_2) - D(\tau_1)B(\tau_2)$. Also, we have

$$\epsilon_{j,t_i} = \widetilde{Y}_{t_i}(\tau_j) + \ln(A(\tau_j)C(\tau_j)) - B(\tau_j)r_{t_i} - D(\tau_j)y_{t_i}, \qquad j = 3, \ldots, m.$$

Letting ϵ_{t_i} now denote the vector of residuals $\epsilon_{3,t_i}, \ldots, \epsilon_{m,t_i}$ and $\epsilon_{t_{i-1}}$ denote the vector of residuals $\epsilon_{3,t_{i-1}}, \ldots, \epsilon_{m,t_{i-1}}$, we denote that the density of the errors ϵ_{t_i} conditional on $\epsilon_{t_{i-1}}$ by

$f(\epsilon_{t_i} \mid \epsilon_{t_{i-1}})$. The joint density for the observations $\underline{\widetilde{Y}}_{t_i}$ conditional on $\underline{\widetilde{Y}}_{t_{i-1}}$, for $j = 1, 2, \ldots, m$, is

$$f(\underline{\widetilde{Y}}_{t_i} \mid \underline{\widetilde{Y}}_{t_{i-1}}) = f(r_{t_i} \mid r_{t_{i-1}}) f(y_{t_i} \mid y_{t_{i-1}}) f(\epsilon_{t_i} \mid \epsilon_{t_{i-1}}) J,$$

where

$$J \equiv \text{abs} \begin{vmatrix} \frac{D(\tau_2)}{F} & \frac{-D(\tau_1)}{F} & 0 & \ldots & 0 \\ -\frac{B(\tau_2)}{F} & \frac{B(\tau_1)}{F} & 0 & \ldots & 0 \\ -\beta_3^1 & -\beta_3^2 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\beta_m^1 & -\beta_m^2 & 0 & \ldots & 1 \end{vmatrix}$$

$$= \frac{1}{F},$$

and

$$\begin{aligned} \beta_j^1 &\equiv -B(\tau_j)\frac{D(\tau_2) - B(\tau_2)}{F}, \\ \beta_j^2 &\equiv -B(\tau_j)\frac{B(\tau_1) - D(\tau_1)}{F}. \end{aligned}$$

The logarithm of the likelihood function is

$$\begin{aligned} \mathcal{L}(\kappa, \sigma^2, \theta, \lambda, \kappa_2, \sigma_2, \theta_2, \sigma_p, \rho, \Sigma_\epsilon) &\equiv \ln f(\ln \underline{\widetilde{Y}}_{t_2}, \ldots, \ln \underline{\widetilde{Y}}_{t_N} \mid \ln \underline{\widetilde{Y}}_{t_1}) \\ &= \ln \prod_{i=2}^{N} f(\ln \underline{\widetilde{Y}}_{t_i} \mid \ln \underline{\widetilde{Y}}_{t_{i-1}}) \qquad (29) \\ &= \sum_{i=2}^{N} \ln f(\ln \underline{\widetilde{Y}}_{t_i} \mid \ln \underline{\widetilde{Y}}_{t_{i-1}}). \qquad (30) \end{aligned}$$

Again, with the exception of a modification due to the settlement terms, this is the likelihood function we use in estimating the two factor model. One can obtain the likelihood function for the extended two-factor model by mimicking the steps above. Given this, we can select among models (e.g., test hypotheses of the form $\bar{r} = 0$) using likelihood ratio tests.

3.3 Description of the Data

The data set we use to estimate and select among the tractable models consists of monthly observations on the prices of 10 different bills, notes, and bonds drawn from the Government Bond Master File prepared by the Center for Research in Security Prices. Each month between December 1971 and December 1986 we took from the file the price and other data for the (just auctioned) 13 and 26-week bills, the longest maturity bill available,[17] the most recently issued (current) 2, 3, 4, 5, 7, and 10 year notes (when available), and the most recently issued non-callable, non-convertible bond with no special tax status. In the early part of the sample period, current notes of all maturities were not available. For example, the Treasury first issued 10 year notes in 1976. When current notes were not available, we selected the most recently issued note or bond with the appropriate maturity. For example, when a 10 year note was not available, we used the most recently issued non-callable, non-convertible bond with no special tax status and a remaining maturity of between 7 and 10 years; when a 4 year note was not available, we used the most recently issued note with a remaining maturity of between 3 and 4 years. We began the sample in December 1971 because, beginning with this month, all of the maturities are continuously available.[18] For all securities, the prices used consist of the means of the bid and ask prices on the last business day of the month, plus the accrued interest to the settlement date.

The sample selection procedure produces a sample that covers the entire yield curve and consists predominantly of current, actively traded securities. The sample consists solely of such securities when they are available, and the sample selection procedure picks recently issued securities when they are not. By including the (just auctioned) 13 and 26-week bills, the sample includes the most actively traded discount bonds. As argued above, it seems reasonable to treat the observed prices of these securities as reasonable approximations of equilibrium prices. We restricted our attention to recently issued securities because the institutional literature (e.g., Stigum (1983), Fabozzi and Pollack (1987)) indicates that government securities issued some time in the past are not actively traded, and suggests that the quoted prices of such securities may not be reasonable approximations of equilibrium prices.

3.4 Adjustment for Settlement Terms

An inconvenience, though in this case not a serious one, is created by the fact that the prices reported in the CRSP Government Bond Files are not quoted on a cash basis. Rather, the prices are typically quoted on the basis of settlement in two business days, or "skip-day" settlement. (That is, bonds and cash are exchanged two business days after the quotation date.) Hence the reported prices are actually not bond prices, but rather the forward prices of very short term forward contracts.

Fortunately, given a model of bond prices, it is trivial to obtain formulas for the forward prices. Letting $P_{t_q}(\tau)$ denote the price on the quotation date t_q of a bond with maturity τ and $P_{t_q}(t_s - t_q)$ denote the price of a bond that matures on the settlement date t_s, the forward price

[17] The bill with the longest maturity has a maturity of one year or slightly less.

[18] Immediately prior to December 1971, there was no non-callable, non-convertible bond with no special tax status and a remaining maturity greater than ten years.

of a forward contract expiring at t_s on the bond with maturity τ is

$$F_{t_q}(P_{t_q}(\tau), t_s - t_q) = P_{t_q}(\tau)/P_{t_q}(t_s - t_q).$$

The data actually consist of observations on these forward prices.

With the CIR one-factor model, we have

$$F_{t_q}(P_{t_q}(\tau), t_s - t_q) = \frac{A(\tau)}{A(t_s - t_q)} e^{-(B(\tau) - B(t_s - t_q))r},$$

where $A(\cdot)$ and $B(\cdot)$ are defined by (6) and (7). With the CIR two factor model,

$$F_{t_q}(\widehat{N}_{t_q}(\tau), t_s - t_q) = \frac{A(\tau)}{A(t_s - t_q)} \frac{C(\tau)}{C(t_s - t_q)} e^{-(B(\tau) - B(t_s - t_q))r} e^{-(D(\tau) - D(t_s - t_q))y},$$

where $C(\cdot)$ and $D(\cdot)$ are defined by (18) and (19). All of our computations are done using these two formulas. That is, although the discussion of the methodology in Section 3 was presented in terms of bond prices and not forward prices, in actual computation we used these formulas. This procedure is exactly correct.

We note that only the existence of a closed-form bond pricing formula enables us to be exactly correct. While the relationship between the prices of bonds for cash settlement and the forward prices is straightforward, we can not recover the prices for cash settlement via some transformation of the data because we do not observe prices for cash settlement of bonds which mature on the settlement date.

4 Selecting Among the Tractable Models

The basic one-factor and two-factor models of CIR, and their extended variants presented above, comprise a set of four neatly nested economic models: the one-factor model with $\bar{r} = 0$ and $\bar{r}$ unconstrained, and the two-factor model with $\bar{r} = 0$ and $\bar{r}$ unconstrained. If, in the two-factor CIR model, we let $\theta_2 = 0$, fix $\sigma_2 > 0$, and consider the limit as $\kappa_2 \to \infty$, we obtain the one factor model.[19] In addition, we can obtain the basic CIR one and two-factor models from the extended variants by letting $\bar{r} = 0$. Given our maximum likelihood approach, if the statistical models were also so nicely nested we could select among the models simply by comparing the logarithms of the likelihood functions, i.e. by using the likelihood ratio test.

Unfortunately, the statistical models are not so nicely nested. In the one-factor models, we impose one linear restriction on the errors in the statistical model of bond prices in order to recover the interest rate r, while in the two-factor models we must impose two linear restrictions

[19]We can also obtain the one-factor model from the two-factor model by setting $\sigma_p = 1$ (see equation (17)). If we did this and interpretted the one-factor model as a special case of the two-factor model with $\sigma_p = 1$ the test statistics below comparing the one and two-factor models would be distributed χ^2 with one instead of two degrees of freedom and our rejections of the one-factor model would be even more resounding.

to recover r and y. Our statistical models are not nested, because we cannot get from the one-factor model to the two-factor model simply by relaxing restrictions. It turns out that this is not a great cause of concern because the translated two-factor model dominates the one-factor models in spite of the stronger restriction imposed on the errors. Although the statistical models for the one and two-factor bond pricing models are not nested, the likelihood ratio statistic provides an informal way of comparing the one and two-factor models,[20] and is the appropriate test statistic for the nested statistical models. Therefore we focus on it in the balance of this section.

In all of the results we set $\sigma_p = \rho = 0$. This restriction is never rejected by the data; in fact, imposing this restriction affects the value of the likelihood function trivially, or not at all. Also, in estimating the two-factor models we impose the restriction $\bar{r} = -10$ (-1000 percent). This restriction is imposed because, for $\bar{r}$ less than about -1 (-100 percent), the value of the likelihood function, while generally increasing in $-\bar{r}$, is insensitive to the value of $\bar{r}$. Specifically, for each value of $\bar{r}$ a constrained maximum of the likelihood function can be found. These maxima are almost imperceptibly increasing in $-\bar{r}$, and, as $-\bar{r}$ is increased, the estimate of θ increases in such a way that $\theta + \bar{r}$ is constant, while the estimate of σ^2 steadily decreases at a rate such that $-\sigma^2\bar{r}$ is approximately constant. The somewhat arbitrary value of $\bar{r} = -10$ was chosen because the rate of increase in the value of the likelihood function is very small by this point. With $\bar{r} = -10$, the behavior of r is very similar to the behavior of an Ornstein-Uhlenbeck process,[21] which corresponds to $\bar{r} = -\infty$, $\theta = \infty$, and $\sigma^2 = 0$, and the bond prices are essentially identical. In the tables below results with $\bar{r} = -10$ are labelled as results with "$\bar{r}$ unconstrained," and the degrees of freedom given for the test statistics are correct if the value of the likelihood function with $\bar{r} = -10$ is equal to the unconstrained value. If the value of the likelihood function with $\bar{r} = -10$ is not approximately equal to the unconstrained value then the rejections we obtain would be even more resounding.

In our initial set of computations[22] we used a data set including observations on bonds of all 10 maturities, and estimated the models for the entire time period January 1972–December 1986, and two subperiods January 1972–September 1979 and October 1979–December 1986. The choice of the subperiods was dictated by the fact that the Federal Reserve changed policies in October 1979, and it is possible that the behavior of interest rates differs in the periods before and after this date.

[20] Formal tests of non-nested hypotheses are available (see, for example Cox (1961, 1962) and White (1983)). However, the translated two-factor model so dominates the one-factor models that simply comparing the logarithms of the likelihood functions seems unlikely to cause one to draw an inappropriate inference.

[21] With the obvious qualification that $\bar{r}$ is a lower bound on the interest rate process in the translated model.

[22] The function $e^{-z}I_q(z)$ which appears in the conditional densities was evaluated using the backward recursion in Section 19.4.2 of Luke (1977), except when either z or q was large. For large z and q the bessel function was evaluated using the asymptotic expansions in Oliver (1965).

4.1 Overview of the Results

We start with a broad overview of the results in Table 1, which displays the values of the logarithms of the likelihood functions for the various models estimated using the entire sample and the two subperiods along with the likelihood ratio test statistics. For example, the value of the likelihood function for the one-factor model under the restriction $\bar{r} = 0$ is $\mathcal{L} = 4.50182 \times 10^3$, and the value of the likelihood function when $\bar{r}$ is unrestricted is $\mathcal{L} = 4.50738 \times 10^3$. The notation "no feasible parameter values" for the two-factor models with $\bar{r} = 0$ indicates that no parameter values for which r_{t_i} and y_{t_i} are nonegative for every time t_i could be found.

Likelihood ratio tests of particular restrictions can be calculated by taking twice the difference of the log-likelihoods. The test statistic L for the likelihood ratio test of the hypothesis $\bar{r} = 0$ is equal to twice the difference of the log-likelihoods, or $L = 2 \times (4.50738 - 4.50182) \times 10^3 = 11.12$. The test statistic is distributed χ^2 with degrees of freedom equal to the number of restrictions being tested. In this case only one restriction is being tested, so the test statistic is distributed χ_1^2. If we use the customary significance level of one percent, the critical value is 6.6349, and we reject the restriction that $\bar{r} = 0$.

Looking at the likelihood ratio test statistics to see which restrictions on the translated two-factor model may be rejected, we reject the one-factor model in favor of its translated variant, and reject both the basic and translated one-factor models in favor of the translated two-factor model. Moreover, the data seems to be inconsistent with the untranslated two-factor model because of our inability to find feasible parameter values. In the succeeding subsections we present further analyses.

4.2 The One-Factor Model

The results for the basic CIR one-factor model and the translated model using a data set that includes all of the bonds are shown in Table 2 for the entire period and the two subperiods. To get a feel for the parameter estimates, consider the model with $\bar{r} = 0$ estimated using the entire sample period January 1972–December 1986. The long run mean of the interest rate is given by the estimate of θ, .08903. The estimate of κ of .16939 implies relatively weak mean reversion in the interest rate process. The "half-life" of the process, i.e. the future time when the expectation of the interest rate has a value halfway between between the current level and the long run mean, is given by $\frac{\ln 2}{\kappa}$, or, with $\kappa = .16939$, 4.09 years. The variance of the steady state distribution of the interest rate is $\frac{\sigma^2 \theta}{2\kappa}$, or 0.00275.

The first observation we make is that the estimated parameter values are significantly different in the two subsamples. A test of this can be performed by computing a likelihood ratio test statistic equal to twice the log-likelihood for the entire period less twice the sum of the log-likelihoods for the two subperiods. The test statistics for the test of the equality of the parameters across the two time periods for both the basic CIR one-factor model and the translated model are shown in the lower panel of Table 2, along with the critical values, which are distributed χ^2 with degrees of freedom equal to the number of parameter restrictions being tested. Examing the test statistics, we see that the hypothesis that the parameters are equal in the different time periods can be rejected at any reasonable level of significance.

Another question of interest is whether the basic one-factor model can be rejected in favor of the translated model for the entire period and for both subperiods, i.e., whether we can reject the hypotheses that $\bar{r} = 0$ for the several periods. Examing the test statistics reported in the first panel of Table 2, we see that tests of size .01 reject the restrictions $\bar{r} = 0$ both for the entire sample period and the second subperiod, and a test of size .05 rejects the restriction for the first subperiod.

We can obtain some insight into the reasons for the rejection by examining how well predicted bond prices fit actual prices for both the basic and translated models. One measure of this fit is provided by the estimated standard deviations of the residuals in the statistical model, which are shown in Table 3.[23] Examining these estimates of the standard deviations, it seems that the translated model fits long term bond prices no better than the basic CIR one-factor model, suggesting that the rejection of the basic model may be due to the fact that the density for the translated model better explains the time series of recovered state variables r.

We can test the conjecture that the rejections are not due to the pricing of the long term bonds by reestimating the model using only the Treasury bills. The results using only the bills, along with the test statistics for the hypotheses that $\bar{r} = 0$ and that the parameters are equal in different time periods, are shown in Table 4. Examing these results, we still reject the hypotheses that $\bar{r} = 0$. This suggests that the advantage of the translated one-factor model lies in the fact that it provides a better approximation of the distribution of the interest rate. In addition, we reject the hypotheses that the parameters are equal in different time periods when we estimate the model using only the Treasury bills.

This ability to obtain these rejections using only the Treasury bills suggests the potential of our approach of recovering the state variable r and making use of its conditional density in estimation. We return to this point below in examining the results of the two-factor model, where we also find that the results seem to be driven by the Treasury bills.

4.3 The Two-Factor Models

Parameter estimates for the two-factor models estimated using all 10 bonds for the entire sample period and both subperiods are shown in Table 5, and estimates of the standard deviations of

[23]These estimates of the standard deviations of the residuals were calculated under the assumption that the means of the errors are zero, so they are also mean-squared errors.

the residuals in the statistical model of bond prices are shown in Table 6.[24] As indicated above, we were not able to find feasible starting values for the basic CIR model with $\bar{r} = 0$. Setting this aside for the moment, the point estimate of κ for the translated models is 7.45 for the entire sample period, 14.51 for the subperiod 1972–1979, and 5.11 for the subperiod 1979–1986. The point estimate of κ_2, the mean reversion parameter for the second state variable process, is about .08 for the entire sample period, .46 for the first subperiod, and .13 for the second subperiod. These values of the speed of adjustment parameter κ are large enough to allow relatively steeply rising yield curves for Treasury bills on days when r is below its long run mean $\theta + \bar{r}$, combined with relatively steeply falling yield curves when r is above its long run mean.[25] The values of κ_2 are small enough (i.e., the mean reversion is weak enough) to be consistent with shifts in the level of the short end of the yield curve.

Specifically, in order to have steeply rising yield curves on some days, together with falling yield curves on others, there must be at least one state variable process for which the speed of adjustment parameter is large. The value of the state variable must be small on days when the yield curve is rising, and large when it is falling. In order for the level of the yield curve to shift from day to day, there must be at least one state variable for which the speed of adjustment parameter is small. If the speed of adjustment parameters were large for all state variables, the yields of longer term bonds would be approximately constant across days regardless of the current value of the state variable. We observe relatively steeply rising yield curves on some days, relatively steeply falling ones on others, and also observe that the level of the yield curve shifts over time. These observations together cannot be reconciled with a one-factor model, but can be explained by a two factor model in which the speed of adjustment parameter is large for one state variable and small for the other.

One check on the values of the parameters is provided by examining the long run means of the state variable processes. The long run mean of the instantaneous real interest rate r is $\theta + \bar{r}$, while the long-run mean of the expected drift in the price level is θ_2. The point estimate of $\theta + \bar{r}$ is 0.026 (standard error .005) for the entire sample period, -0.031 (standard error .063) for the subperiod 1972–1979, and 0.041 (standard error .006) for the subperiod 1979–1986. The point estimate of θ_2, the long run mean for the expected rate of inflation, is about 0.093 for the entire sample period, 0.093 for the first subperiod, and 0.1014 for the second subperiod.

We can obtain confidence in our interpretation of the process for r as the real interest

[24] We found two local maxima of essentially the same magnitude, one with $\kappa > \kappa_2$ and one with $\kappa < \kappa_2$, for each of the models for which we are able to find feasible parameter values. We present only estimates from the local maxima with $\kappa > \kappa_2$, i.e., the results we present associate the state variable with the larger speed of adjustment parameter with the real model. We do this for the reasons below that lead us to believe that our parameter estimates are plausible.

[25] In this discussion the steepness of the short end of the yield curve is measured by the difference in the yield on the three and six month Treasury bills, the two shortest maturity bills in the sample.

rate process by comparing our results to those of Gibbons and Ramaswamy (1986). These authors estimated a one-factor model for real yields using Treasury bill data and obtained point estimates of κ of 3.9, 6.9, and 2.4 for the three periods 1964–1983, 1964–1979, and 1979–1983. Gibbons and Ramaswamy obtained point estimates of 5.4, 2.2, and 3.2 for σ^2, and -4.6, -8.5, and -3.0 for λ. These estimates, however, may not be directly compared to our estimates from the translated model because both the local variance of the interest rate and the risk premium depend upon the level of the interest rate. In the translated model the local variance of the interest rate is given by var(r) $= \sigma^2(r - \bar{r})$ and the risk premium is $\lambda r' = \lambda(r - \bar{r})$, while in the basic CIR model we have var(r) $= \sigma^2 r$, and the risk premium is λr. Using the point estimates of Gibbons and Ramaswamy to calculate the local variance of the interest rate and the risk premium at the long-run mean of the interest rate, we obtain var(r) = 0.043, 0.004, and 0.096 and risk premiums of -0.036, -0.017, and -0.090 for their three sample periods. For our translated model with $\bar{r}$ unconstrained, performing similar calculations at the long run mean of r gives us local variances of 0.004, 0.010, and 0.004 and risk premiums of -0.048, -0.221, and -0.009 for our three sample periods.

These estimates of Gibbons and Ramaswamy, combined with our estimates of the long run means of the state variable processes, lead us to interpret the first state variable as the real interest rate and the second state variable as the expected inflation rate. While the estimates of Gibbons and Ramaswamy differ from ours, some of the differences must be attributed to the differing sample periods and sampling variation. Regardless, we interpret our estimates as broadly consistent with those of Gibbons and Ramaswamy in that our estimates also imply relatively strong mean reversion in the real interest rate.

Despite the reasonableness of the parameter estimates for the translated models, our inability to find feasible parameters for the basic CIR two-factor model is troubling. The difficulty arises from the fact that in the basic CIR model, both r and y must be nonnegative. If the portion of the yield curve that we use to recover the state variables takes on many different shapes during the sample period, it may be impossible to find a set of parameter values for which r and y are nonnegative for every observation.

We can get some insight into the source of the problem by examining the time series of recovered state variables for one of the translated models. Figures 1 and 2 display the time series of recovered state variables r' and y for the translated model with $\bar{r}$ unconstrained estimated using the entire sample. For this model, examination of Figure 1 suggests that the source of the problem with estimating the basic CIR two-factor model might lie in the observations for September 1974 and July, August, and September 1982. The values of the interest rate r for these dates, -0.0292, -0.0293, -0.0379, and -0.0270 are more than .05 less than the long run mean of 0.0264. These observations correspond to steeply rising yield curves; when r' is relatively small, the yields on shorter term bonds are small, but the yields on longer term bonds are higher because of the strong mean reversion. When r' is near its long run mean the yield curve is relatively flat.

This seems to provide the explanation of why we cannot find feasible parameter values for the basic CIR two-factor model. It seems that in order to explain the steepness of the yield curves for September 1974 and July, August, and September 1982, r must be about .05 below its long run mean. If r must also be non-negative, the long run mean must be at least .05. However, to have a relatively flat yield curve, r must be near its long run mean. If we observe a flat yield curve at a level of less than .05, then the long run mean of r must be less than .05. In short, the basic CIR model cannot simultaneously explain both steeply rising yield curves and flat yield curves at a low level of interest rates.[26]

As mentioned above, examination of the log-likelihoods and test statistics shown in Table 1 indicates that the translated two-factor models dominate the one-factor models. It is possible to get some insight into why this occurs.

Echoing a similar observation made above when comparing the basic and translated CIR one-factor models, the rejection of the one-factor models in favor of the translated two-factor model does not seem to be due to better pricing of the long term bonds. Table 6 displays the estimated standard deviations of the errors in the statistical model for the prices of bonds for the two-factor models. Comparing these to the estimated standard deviations of the error model for the one-factor model reported in Table 3, we see that the estimated standard deviations for the two-factor model are smaller than the estimated standard deviations for the one-factor model, but not dramatically so. This observation, while only suggestive, again leads one to the conjecture that the dominance of the two-factor model is not due to the better pricing of long-term bonds.

Again, we can test this conjecture by reestimating the models using only the 13-week, 26-week, and one-year Treasury bills. Table 7 displays the values of the likelihood functions and the likelihood ratio test statistics comparing the basic and translated one and two-factor models for the entire sample period and both subperiods for the models estimated using only the Treasury bills. The point estimates of the parameters of the one-factor models were previously presented in Table 3, while those for the two-factor model are shown in Table 8. Examing these tables, we see that we can reject the one-factor models in favor of the translated two-factor models using the bills alone.

A tentative conclusion is that the translated two-factor model dominates the one-factor

[26] We checked the conjecture that the problem lies in the observations for September 1974 and July, August, and September 1982 by reestimating the two-factor models after dropping these observations from the sample. Dropping the three observations allows us to obtain at least somewhat plausible parameter estimates for the basic CIR two-factor model, and does not significantly affect the parameter estimates for the translated models.

models, and of course also dominates the basic two-factor model for which feasible parameters could not be found. However, recalling the estimated standard deviations from the statistical model in Table 6, it does not seem to provide a good fit to the prices of long term bonds. This suggests that the two-factor model does not provide a good description of the data. In the next section we bring additional information to bear on the issue.

5 Tests Against a General Markov Model for Yields

The results above indicate that the two-factor model fits the data better than the one-factor model. However, we have not yet compared the two-factor model to more general models. If we restrict our attention to the three Treasury bills, we can nest the two-factor model in a somewhat more general Markov model of yields as follows.

Defining $Y_{t_i}^*(\tau_j) \equiv \frac{\widetilde{Y}_{t_i}(\tau_j)}{\tau_j}$, from the bond pricing model (17)-(21) and the statistical model (25)-(26), we have

$$Y_{t_i}^*(\tau_1) = -\frac{\ln(A(\tau_1)C(\tau_1))}{\tau_1} + \bar{r} + \frac{B(\tau_1)}{\tau_1} r'_{t_i} + \frac{D(\tau_1)}{\tau_1} y_{t_i}, \tag{31}$$

$$Y_{t_i}^*(\tau_2) = -\frac{\ln(A(\tau_2)C(\tau_2))}{\tau_2} + \bar{r} + \frac{B(\tau_2)}{\tau_2} r'_{t_i} + \frac{D(\tau_2)}{\tau_2} y_{t_i}, \tag{32}$$

$$Y_{t_i}^*(\tau_3) = -\frac{\ln(A(\tau_3)C(\tau_3))}{\tau_3} + \bar{r} + \frac{B(\tau_3)}{\tau_3} r'_{t_i} + \frac{D(\tau_3)}{\tau_3} y_{t_i} + \epsilon_{3,t_i}. \tag{33}$$

Write

$$Y_{t_i}^*(\tau_1) = \alpha_{10} + \alpha_{11} r'_{t_i} + \alpha_{12} y_{t_i}, \tag{34}$$

$$Y_{t_i}^*(\tau_2) = \alpha_{20} + \alpha_{21} r'_{t_i} + \alpha_{22} y_{t_i}, \tag{35}$$

$$Y_{t_i}^*(\tau_3) = \alpha_{30} + \alpha_{31} r'_{t_i} + \alpha_{32} y_{t_i} + \epsilon_{3,t_i}. \tag{36}$$

If the translated CIR two-factor model is true, then

$$\alpha_{j0} = -\ln(A(\tau_j)C(\tau_j))/\tau_j + \bar{r}, \tag{37}$$

$$\alpha_{j1} = B(\tau_j)/\tau_j, \tag{38}$$

$$\alpha_{j2} = D(\tau_j)/\tau_j, \tag{39}$$

for $j = 1,2,3$.

It is straightforward to see that the CIR model (31)-(33) imposes five restrictions on the more general model (34)-(36). If $\alpha_{11} > 0$, then equation (15) implies that λ can be written as a function of α_{11}, κ and σ i.e., $\lambda = \lambda(\alpha_{11}, \kappa, \sigma)$. Also, using equations (37), (14), and (18), we can solve for θ, θ_2, and $\bar{r}$ in terms of α_{10}, α_{20}, α_{30}, κ, σ, κ_2, and σ_2. Assuming $\alpha_{11} > 0$, doing this, and substituting into (31)–(33), we see that the CIR model (31)–(33) can be written as

$$Y^*_{t_i}(\tau_1) = \alpha_{10} + \alpha_{11} r'_{t_i} + \frac{D(\tau_1)}{\tau_1} y_{t_i}, \tag{40}$$

$$Y^*_{t_i}(\tau_2) = \alpha_{20} + \frac{B(\tau_2)}{\tau_2} r'_{t_i} + \frac{D(\tau_2)}{\tau_2} y_{t_i}, \tag{41}$$

$$Y^*_{t_i}(\tau_3) = \alpha_{30} + \frac{B(\tau_3)}{\tau_3} r'_{t_i} + \frac{D(\tau_3)}{\tau_3} y_{t_i} + \epsilon_{3,t_i}. \tag{42}$$

Comparing (34)–(36) and (40)–(42), we see that the CIR bond pricing model gives five restrictions on the parameters of the general model (34)–(36).[27] The conditional densities of r' and y are given by (23) and (27) with θ and θ_2 given by the appropriate functions of the α_{ji} and κ, σ, κ_2, and σ_2.

If we neglect the small differences in the times to maturity τ_j and time-steps $t_i - t_{i-1}$ across observations stemming from the different number of days in different months, the coefficients α_{ji} will be constant. Below we estimate the model (34)–(36) and test the restrictions is (40)–(42).

To see that (34)–(36) are equivalent to a particular Markov model for yields, and therefore that a test of the restrictions in (40)–(42) is equivalent to a test of the CIR model versus a general Markov model of yields, rewrite (34)–(36) to obtain

$$\begin{aligned}
Y^*_{t_i}(\tau_1) &= \alpha_{10} + \alpha_{11}\theta(1 - e^{-\kappa(t_i - t_{i-1})}) + \alpha_{12}\theta_2(1 - e^{-\kappa_2(t_i - t_{i-1})}) \\
&\quad + \alpha_{11} e^{-\kappa(t_i - t_{i-1})} r'_{t_{i-1}} + \alpha_{12} e^{-\kappa_2(t_i - t_{i-1})} y_{t_{i-1}} + \eta_{1,t_i}, \\
Y^*_{t_i}(\tau_2) &= \alpha_{20} + \alpha_{21}\theta(1 - e^{-\kappa(t_i - t_{i-1})}) + \alpha_{22}\theta_2(1 - e^{-\kappa_2(t_i - t_{i-1})}) \\
&\quad + \alpha_{21} e^{-\kappa(t_i - t_{i-1})} r'_{t_{i-1}} + \alpha_{22} e^{-\kappa_2(t_i - t_{i-1})} y_{t_{i-1}} + \eta_{2,t_i}, \\
Y^*_{t_i}(\tau_3) &= \alpha_{30} + \alpha_{31}\theta(1 - e^{-\kappa(t_i - t_{i-1})}) + \alpha_{32}\theta_2(1 - e^{-\kappa_2(t_i - t_{i-1})}) \\
&\quad + \alpha_{31} e^{-\kappa(t_i - t_{i-1})} r'_{t_{i-1}} + \alpha_{32} e^{-\kappa_2(t_i - t_{i-1})} y_{t_{i-1}} + \eta_{3,t_i},
\end{aligned}$$

where

$$\begin{aligned}
\eta_{1,t_i} &\equiv \alpha_{11}(r'_{t_i} - e^{-\kappa(t_i - t_{i-1})} r'_{t_{i-1}} - \theta(1 - e^{-\kappa(t_i - t_{i-1})})) \\
&\quad + \alpha_{12}(y_{t_i} - e^{-\kappa(t_i - t_{i-1})} y_{t_{i-1}} - \theta_2(1 - e^{-\kappa_2(t_i - t_{i-1})})), \\
\eta_{2,t_i} &\equiv \alpha_{21}(r'_{t_i} - e^{-\kappa(t_i - t_{i-1})} r'_{t_{i-1}} - \theta(1 - e^{-\kappa(t_i - t_{i-1})})) \\
&\quad + \alpha_{22}(y_{t_i} - e^{-\kappa(t_i - t_{i-1})} y_{t_{i-1}} - \theta_2(1 - e^{-\kappa_2(t_i - t_{i-1})})),
\end{aligned}$$

[27] If $\alpha_{11} \leq 0$ there is a sixth restriction $\alpha_{11} = B(\tau_1)/\tau_1$.

$$\eta_{3,t_i} \equiv \alpha_{31}(r'_{t_i} - e^{-\kappa(t_i-t_{i-1})} r'_{t_{i-1}} - \theta(1 - e^{-\kappa(t_i-t_{i-1})})) + \alpha_{32}(y_{t_i} - e^{-\kappa(t_i-t_{i-1})} y_{t_{i-1}} - \theta_3(1 - e^{-\kappa_2(t_i-t_{i-1})})) + \epsilon_{3,t_i}.$$

The variables η_{1,t_i} and η_{2,t_i} are linear combinations of "demeaned" non-central χ^2 random variables and η_{3,t_i} is a linear combination of two "demeaned" non-central χ^2 random variables and a normal random variable.

Solving (34) and (35) for $r'_{t_{i-1}}$ and $y_{t_{i-1}}$,

$$r'_{t_{i-1}} = (\alpha_{22} Y^*_{t_{i-1}}(\tau_1) - \alpha_{12} Y^*_{t_{i-1}}(\tau_2) - \alpha_{22}\alpha_{10} + \alpha_{12}\alpha_{20})/F, \quad (43)$$

$$y_{t_{i-1}} = (-\alpha_{21} Y^*_{t_{i-1}}(\tau_1) + \alpha_{11} Y^*_{t_{i-1}}(\tau_2) + \alpha_{21}\alpha_{10} - \alpha_{11}\alpha_{20})/F, \quad (44)$$

where

$$F = \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}.$$

Substituting into (43)–(43),

$$Y^*_{t_i}(\tau_1) = \beta_{10} + \beta_{11} Y^*_{t_{i-1}}(\tau_1) + \beta_{12} Y^*_{t_{i-1}}(\tau_2) + \eta_{1,t_i}, \quad (45)$$

$$Y^*_{t_i}(\tau_2) = \beta_{20} + \beta_{21} Y^*_{t_{i-1}}(\tau_1) + \beta_{22} Y^*_{t_{i-1}}(\tau_2) + \eta_{2,t_i}, \quad (46)$$

$$Y^*_{t_i}(\tau_3) = \beta_{30} + \beta_{31} Y^*_{t_{i-1}}(\tau_1) + \beta_{32} Y^*_{t_{i-1}}(\tau_2) + \eta_{3,t_i}, \quad (47)$$

where the β_{ji} are the appropriate functions of the α_{ji}, κ, σ, κ_2, and σ_2. (Recall that θ and θ_2 can be expressed as functions of these variables.)

If we continue to neglect the small differences in the times to maturity τ_j and time-steps $t_i - t_{i-1}$ across observations stemming from the different number of days in different months, the coefficients β_{ji} will be constant. Except for the special structure on the errors η_{j,t_i}, equations (45)–(47), which are equivalent to (34)–(36), comprise a regression model with nonlinear restrictions on the parameters given by the bond pricing model (17)–(21). Therefore tests of the restrictions in (40)–(42) can be interpreted as tests of the CIR two-factor model versus a more general Markov model for discount yields.

The conditional density for the alternative Markov model of yields can be obtained in a straightforward fashion, and is given by equation (56) of the appendix. Letting $f(\underline{Y}^*_{t_i}|\underline{Y}^*_{t_{i-1}})$ denote this density, the conditional density for the logs of prices $\underline{Y}_{t_i}$ is $f(\underline{Y}^*_{t_i}|\underline{Y}^*_{t_{i-1}})\tau_1\tau_2\tau_3$. The likelihood function can be obtained in the obvious fashion. Given this, the restrictions can be tested using likelihood ratio test statistics. The likelihood function is written in terms of the nine parameters α_{ij} that appear in equations (31)–(33) rather the the nine parameters β_{ij}, but this does not affect the values of the test statistic.

5.1 Results

The parametrization of the likelihood function in terms of the α_{ji} is arbitrary and the parameter estimates are difficult to interpret. Accordingly we present the results in terms of the β_{ji}. Table 9 displays the estimates of the coefficients β_{ji} for the sample period 1972–1986 and the two subperiods 1972–1979 and 1979–1986 calculated from the restricted and unrestricted parameter estimates. The table also displays the values of the likelihood ratio test statistics for tests of the restrictions implied by the bond pricing model. These statistics are considerably larger than the critical values for tests of size .01 and indicate that the performance of the general Markov model for yields is superior to the performance of the CIR two-factor model.

The distinctive feature of these results is that in the CIR model the estimates of the parameters β_{j1}, which are the "regression coefficients" for the three month bill yield, are negative, while the estimates of the parameters β_{j2}, which are the "regression coefficients" for the six month bill yield, are positive. The sums $\beta_{j1} + \beta_{j2}$ are approximately equal to one for $j = 1, 2, 3$. The interpretation of these estimates is that in the CIR model yields are predicted to decrease sharply when the three month bill yield is larger than the yield on the six month bill, and predicted to increase when the yield on the three month bill is below than the yield on the six month bill.

In the general Markov model, the coefficients β_{j1} are either positive, or negative and smaller in absolute magnitude than those in the CIR model, and the coefficients β_{j2} are smaller. With these unrestricted coefficients there is a weaker relationship between the slope of the short end of the yield curve and predicted future yields. This suggests that the deficiency of the CIR model is that it requires a relationship between the slope of the yield curve and future yields that is not found in the data.

This rejection of the bond pricing model can be further understood by examining the mean squared forecast errors of the predicted yields on the sample of discount bonds. That is, we take the difference between actual and predicted yields on the sample of Treasury bills, and look at the magnitude of the forecast errors. Table 10 displays the mean squared forecast errors of the holding period returns of the 3 different Treasury bill maturities for the entire sample period 1972–1986 and both subperiods. The predicted yields used to form the forecast errors in the first three columns were calculated using the parameter estimates from both the translated one and two-factor models and the general Markov model for yields. The last column shows the forecast errors from a naive "martingale model" in which the predicted yield of the 3, 6, and 12-month bills are assumed to be equal to their current yields. We find that the mean squared errors when the (estimates of the) expected yields are calculated using the restricted two-factor model are larger than the mean squared errors when the expected yields are calculated using either the general Markov model or the naive martingale model.[28]

One possible explanation for these results is as follows. The estimated two-factor model implies a particular relationship between the state variables and expected holding period returns. We recover the state variables from the prices of the 13 and 26-week bills through the assumption that the observed prices of these bills are equilibrium prices. In using the model to compute expected holding period returns, we use the model to infer expected holding period returns from the prices of the 13 and 26-week bills, i.e. from the level and slope of the short end of the yield curve. If the two-factor model is not a reasonable approximation of the "true" model, there is little reason to expect that the expected holding period returns calculated using it will provide good predictions of actual holding period returns.

Our estimation procedure essentially forces the one-factor model to capture the level of the short end of the yield curve, and forces the two-factor model to capture both the level and the slope of the short end of the yield curve.[29] If the relationship between the slope of the yield curve and the expected holding period returns is not well approximated by the estimated two-factor model, one might expect the general Markov and martingale models of yields to provide better predictions of holding period returns than the two-factor model. Regardless, the evidence suggests that the particular two-factor CIR model estimated here fails to provide a good description of holding period returns.

6 Conclusion

We have developed an approach for estimating both one and two-factor term structure models due to CIR using the method of maximum likelihood. Our procedure involves recovering the unobservable state variables. This enables us to exploit their conditional density in estimation and hypothesis testing, and to make use of both the time-series and cross-sectional information in a sample of bonds that includes coupon bonds. The time series of state variables that we recover enables us to examine whether factors other than the recovered state variables are useful in explaining conditional expected returns.

We were unable to find feasible parameter values for the basic (untranslated) CIR two-factor model. The results of a set of tests that select among the tractable models suggest that the one-factor model should be rejected in favor of its translated variant, and that both the basic and translated one-factor models should be rejected in favor of the translated two-factor model. However, none of the tractable models provide a good fit for the prices of long term coupon

[28] That the performance of the "martingale model" seems to be superior to that of the "general Markov model" is not evidence of a computational error. While the coefficients β_{ji} of the general Markov model are unrestricted, the errors are constrained to be highly correlated.

[29] More precisely, in the one-factor model the fitted price of the 13-week bill will equal the observed price, while in the two-factor model the fitted prices of the 13 and 26-week bills will equal their observed prices.

bonds. Moreover, in further empirical analysis we are able to reject the restrictions on a Markov model of yields implied by the bond pricing model.

Although we do not do so in this paper, our approach may also be used to estimate the models of Vasichek (1978), Schaefer and Schwartz (1984), and Oldfield and Rogalski (1987) in which one or more of the underlying state variables follow an Ornstein-Uhlenbeck process. More generally, our approach may be applied to any asset pricing model with unobservable state variables in which the pricing functions are invertible and the density of the state variables may be evaluated without excessive difficulty.

Appendix

A The restriction to obtain a constant factor risk premium

Equation (12) is a strong restriction on $\hat{\alpha}$ and Ω. One indication of the strength of the restriction is that it implies that the proportion of wealth invested in each of the technologies does not depend upon $\hat{\alpha}$. Sufficient conditions for equation (12) to hold are that there is only one productive technology or that the rates on return on the productive technologies are equal.

Once we translate the physical rates of return α (i.e., once we write $\alpha = \hat{\alpha}(Y' + \overline{Y})$), condition (12) is needed for the set of active technologies to remain unchanged when $\overline{Y} \neq 0$. To see this, suppose $\hat{\alpha}'\Omega^{-1} - \frac{1'\Omega^{-1}\hat{\alpha}'\Omega^{-1}1}{1'\Omega^{-1}1} \neq 0$ and $n > 1$. From the definition of a^*,

$$a^{*\prime}1 = \left[\frac{1'\Omega^{-1}}{1'\Omega^{-1}1} + \frac{Y' + \overline{Y}}{Y'}\left(\hat{\alpha}'\Omega^{-1} - \frac{1'\Omega^{-1}\hat{\alpha}'\Omega^{-1}1}{1'\Omega^{-1}1}\right)\right]1 = 1.$$

Also, $\frac{1'\Omega^{-1}1}{1'\Omega^{-1}1} = 1$, implying $\frac{Y'+\overline{Y}}{Y'}\left(\hat{\alpha}'\Omega^{-1} - \frac{1'\Omega^{-1}\hat{\alpha}'\Omega^{-1}1}{1'\Omega^{-1}1}\right)1 = 0$. But we have supposed that $\hat{\alpha}'\Omega^{-1} - \frac{1'\Omega^{-1}\hat{\alpha}'\Omega^{-1}1}{1'\Omega^{-1}1} \neq 0$, so at least one element must be greater than zero and at least one must be less. If $\overline{Y} < 0$ then $(Y' + \overline{Y})/Y'$ takes on values in the interval $(-\infty, 1)$, while if $\overline{Y} > 0$ then $(Y' + \overline{Y})/Y'$ takes on values in the interval $(1, \infty)$. In either case, if we require $a* \geq 0$, the set of active technologies must change (the signs of some elements of a^* must change).

If we do not require that $a^* \geq 0$ then we do not need to assume that (12) is satisfied, the factor risk premium $a^{*\prime}GS'$ has the form $\lambda r' + \varphi$, the partial differential equation satisfied by the bond price is

$$1/2\sigma^2 r' P_{r'r'} + \kappa(\theta - r')P_{r'} - (\lambda r' + \varphi)P_{r'} - (r' + \bar{r})P = 0$$

and the solution is $P(\tau) = A(\tau)^{\frac{\kappa\theta-\varphi}{\kappa\theta}} e^{-B(\tau)r' - \bar{r}\tau}$. However, we do not use this formula because we are not aware of a reasonable interpretation of negative elements of a^*.

It is our view that the strength of the restriction (12) lessens the theoretical interest of the extension, but does not detract from its empirical utility.

B The likelihood function for the general model

To obtain the general model for the yields of the three discount bonds we suppose that the observed yields are linearly related to the two state variables r' and y, and that the observed yield of the third bond also depends upon a third random variable ϵ_{3,t_i}:

$$Y^*_{t_i}(\tau_1) = \alpha_{10} + \alpha_{11} r'_{t_i} + \alpha_{12} y_{t_i}, \tag{48}$$

$$Y^*_{t_i}(\tau_2) = \alpha_{20} + \alpha_{21} r'_{t_i} + \alpha_{22} y_{t_i}, \tag{49}$$

$$Y^*_{t_i}(\tau_3) = \alpha_{30} + \alpha_{31} r'_{t_i} + \alpha_{32} y_{t_i} + \epsilon_{3,t_i}. \tag{50}$$

The parameters α_{ij}, for $i = 1,2,3$ and $j = 0,1,2$, are unrestricted. The densities of r'_{t_i} and y_{t_i} are

$$f(r'_{t_i} \mid r'_{t_{i-1}}) = ce^{-u-v}\left(\frac{v}{u}\right)^{q/2} I_q(2(uv)^{1/2}), \tag{51}$$

where

$$\begin{aligned} c &\equiv \frac{2\kappa}{\sigma^2(1-e^{-\kappa(t_i-t_{i-1})})}, \\ u &\equiv c r'_{t_{i-1}} e^{-\kappa(t_i-t_{i-1})}, \\ v &\equiv c r'_{t_i}, \\ q &\equiv \frac{2\kappa\theta}{\sigma^2} - 1, \end{aligned}$$

and

$$f(y_{t_i} \mid y_{t_{i-1}}) = c_2 e^{-u_2-v_2}\left(\frac{v_2}{u_2}\right)^{q_2/2} I_{q_2}(2(u_2 v_2)^{1/2}), \tag{52}$$

where

$$\begin{aligned} c_2 &\equiv \frac{2\kappa_2}{\sigma_2^2(1-e^{-\kappa_2(t_i-t_{i-1})})}, \\ u_2 &\equiv c_2 y_{t_{i-1}} e^{-\kappa_2(t_i-t_{i-1})}, \\ v_2 &\equiv c_2 y_{t_i}, \\ q_2 &\equiv \frac{2\kappa_2\theta_2}{\sigma_2^2} - 1. \end{aligned}$$

Also, $\epsilon_{3,t_i} \sim N(0, \sigma_\epsilon^2)$.

From (48)-(50),

$$
\begin{aligned}
r'_{t_i} &= (\alpha_{22}Y^*_{t_i}(\tau_1) - \alpha_{12}Y^*_{t_i}(\tau_2) + \alpha_{22}\alpha_{10} - \alpha_{12}\alpha_{20})/F, & (53)\\
y_{t_i} &= (-\alpha_{21}Y^*_{t_i}(\tau_1) + \alpha_{11}Y^*_{t_{i-1}}(\tau_2) - \alpha_{21}\alpha_{10} + \alpha_{11}\alpha_{20})/F, & (54)\\
\epsilon_{3,t_i} &= -\Big(\alpha_{30} + \alpha_{31}(\alpha_{22} - \alpha_{12})Y^*_{t_i}(\tau_1) + \alpha_{32}(\alpha_{11} - \alpha_{21})Y^*_{t_i}(\tau_2) & (55)\\
&+ \alpha_{30}(\alpha_{22}\alpha_{10} - \alpha_{12}\alpha_{20} - \alpha_{21}\alpha_{10} + \alpha_{11}\alpha_{20}) + Y_{t_i}(\tau_3)\Big)/F.
\end{aligned}
$$

where $F = \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}$.

Let $\underline{Y}^*_{t_{i-1}}$ denote the vector of yields $(Y^*_{t_{i-1}}(\tau_1), Y^*_{t_{i-1}}(\tau_2), Y^*_{t_{i-1}}(\tau_3))$. The joint density for the observations $\underline{Y}^*_{t_i}$ conditional on $\underline{Y}^*_{t_{i-1}}$ is

$$
f(\underline{Y}^*_{t_i} \mid \underline{Y}^*_{t_{i-1}}) = f(r'_{t_i} \mid r'_{t_{i-1}}) f(y_{t_i} \mid y_{t_{i-1}}) f(\epsilon_{3,t_i} \mid \epsilon_{3,t_{i-1}})/F. \qquad (56)
$$

Examining the density function (51) along with (48), one can see that the parameters α_{1i}, κ, θ, and σ are not separately identified. For convenience, we chose the independent parameters to be α_{1i}, $i = 1, 2, 3$, κ and σ, and set $\theta = \bar{\theta}$, an arbitrary constant.

Similarly, the parameters α_{2i}, $i = 1, 2, 3$, κ_2, θ_2, and σ_2 are not separately identified. We choose the independent parameters to be α_{2i}, κ_2 and σ_2, and set $\theta_2 = \bar{\theta}_2$, a constant. With these choices, there are 14 independent parameters: the α_{ji}, for $j = 1, 2, 3$ and $i = 0, 1, 2$, and κ, σ, κ_2, σ_2, and σ_ϵ. The logarithm of the likelihood function is

$$
\begin{aligned}
\mathcal{L}(\alpha_{10}, \ldots, \alpha_{32}, \kappa, \sigma, \kappa_2, \sigma_2, \sigma_\epsilon) &\equiv \ln f(\ln \underline{\widetilde{Y}}_{t_2}, \ldots, \ln \underline{\widetilde{Y}}_{t_N} \mid \ln \underline{\widetilde{Y}}_{t_1})\\
&= \ln \prod_{i=2}^{N} f(\ln \underline{\widetilde{Y}}_{t_i} \mid \ln \underline{\widetilde{Y}}_{t_{i-1}})\\
&= \sum_{i=2}^{N} \ln f(\ln \underline{\widetilde{Y}}_{t_i} \mid \ln \underline{\widetilde{Y}}_{t_{i-1}}).
\end{aligned}
$$

where now

$$
f(r'_{t_i} \mid r'_{t_{i-1}}) = c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q(2(uv)^{1/2}),
$$

with

$$
\begin{aligned}
c &\equiv \frac{2\kappa}{\sigma^2(1 - e^{-\kappa(t_i - t_{i-1})})},\\
u &\equiv c r'_{t_{i-1}} e^{-\kappa(t_i - t_{i-1})},\\
v &\equiv c r'_{t_i},\\
q &\equiv \frac{2\kappa\bar{\theta}}{\sigma^2} - 1,
\end{aligned}
$$

and

$$f(y_{t_i} \mid y_{t_{i-1}}) = c_2 e^{-u_2 - v_2} \left(\frac{v_2}{u_2} \right)^{q_2/2} I_{q_2}(2(u_2 v_2)^{1/2}),$$

where

$$\begin{aligned} c_2 &\equiv \frac{2\kappa_2}{\sigma_2^2(1 - e^{-\kappa_2(t_i - t_{i-1})})}, \\ u_2 &\equiv c_2 y_{t_{i-1}} e^{-\kappa_2(t_i - t_{i-1})}, \\ v_2 &\equiv c_2 y_{t_i}, \\ q_2 &\equiv \frac{2\kappa_2 \bar{\theta}_2}{\sigma_2^2} - 1, \end{aligned}$$

$\epsilon_{3,t_i} \sim N(0, \sigma_\epsilon^2)$, and r'_{t_i}, y_{t_i}, and ϵ_{3,t_i} are given by (53), (54), and (55).

References

Brennan, M. J. and E. S. Schwartz, 1979, "A Continous Time Approach to the Pricing of Bonds," *Journal of Banking and Finance* 3, 133–155.

Brennan, M. J. and E. S. Schwartz, 1980, "Conditional Predictions of Bond Prices and Returns," *Journal of Finance* 35, 405–416.

Brennan, M. J. and E. S. Schwartz, 1982, "An Equilibrium Model of Bond Pricing and a Test of Market Efficiency," *Journal of Financial and Quantitative Analysis* 17, 301–329.

Brown, R. H. and S. M. Schaefer, 1988, "Testing the CIR Model on British Government Index Linked Securities," working paper, Institute of Finance and Accounting, London Business School.

Campbell, J. Y., 1986, "A Defense of Traditional Hypotheses About the Term Structure of Interest Rates, *Journal of Finance* 41, 183–194.

Cox, D. R., 1961, "Tests of Separate Families of Hypotheses," in J. Neyman, ed., *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1, Berkeley, University of California Press, 105–123.

Cox, D. R., 1962, "Further Results on Tests of Separate Families of Hypotheses," *Journal of the Royal Statistical Society* ser. B, 406–424.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross, 1985a, "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica* 53, 363–384.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross, 1985b, "A Theory of the Term Structure of Interest Rates," *Econometrica* 53, 385–408.

Dothan, L. U., 1978, "On the Term Structure of Interest Rates," *Journal of Financial Economics* 6, 59–69.

Fabozzi, F. J. and I. F. Pollack (eds.), 1987, *The Handbook of Fixed Income Securities*, Dow Jones-Irwin, Homewood, Illinois.

Fama, E. F., 1984a, "Information in the Term Structure," *Journal of Financial Economics* 13, 509–528.

Fama, E. F., 1984b, "Term Premiums in Bond Returns, *Journal of Financial Economics* 13, 529–546.

Feller, W., 1970, *An Introduction to Probability Theory and its Applications,* vol. II, New York, John Wiley & Sons.

Gibbons M. R., and K. Ramaswamy, 1986, "The Term structure of Interest Rates: Empirical Evidence," mimeo, Stanford University.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica* 50, 1029–1054.

Heath, D., R. Jarrow, and A. Morton, 1988, "Bond Pricing and the Term Structure of Interest Rates: A New Methodology," mimeo, Cornell University.

Heston, S., 1988, "Testing Continuous Time Models of the Term Structure of Interest Rates," mimeo, Carnegie Mellon University.

Ingersoll, J., J. Skelton, and R. Weil, 1978, "Duration Forty Years Later," *Journal of Financial and Quantitative Analysis*, 627–650.

Longstaff, F. A., 1989, "A Nonlinear General Equilibrium Model of the Term Structure of Interest Rates," *Journal of Financial Economics* 23, 195–224.

Luke, Y., 1977, *Alghorithms for the Computation of Mathematical Functions*, Academic Press, Inc., New York, New York.

Mankiw, N. G., 1986, "The Term Structure of Interest Rates Revisited," *Brookings Papers on Economic Activity* 1, 61–96.

Marsh, T., 1980, "Equilibrium Term Structure Models: Test Methodology," *Journal of Finance* 35, 421–435.

Marsh, T. and E. Rosenfeld, 1983, "Stochastic Processes for Interest Rates and Equilibrium Bond Prices," *Journal of Finance* 38, 635–645.

Merton, R. C., 1970, "A Dynamic General Equilibrium Model of the Asset Market and its Application to the Pricing of the Capital Structure of the Firm," Sloan School of Management Working Paper 497-70.

Merton, R. C., 1973, "An Intertemporal Capital Asset Pricing Model," *Econometrica* 41, 867–887.

Oldfield, G. S., and R. J. Rogalski, 1987, "The Stochastic Properties of Term Structure Movements," *Journal of Monetary Economics* 19, 229–245.

Oliver, F. W. J., 1965, "Bessel Functions of Integer Order," *Handbook of Mathematical Functions,* edited by M. A. Abramovitz and I. A. Stegun, New York, Dover.

Richard, S. F., 1978, An Arbitrage Model of the Term Structure of Interest Rates," *Journal of Financial Economics* 6, 33–57.

Roll, R., 1971, "Investment Diversification and Bond Maturity," *Journal of Finance* 26, 51–66.

Roll, R., 1970, *The Behavior of Interest Rates,* New York, Basic Books, Inc.

Schaefer, S. M. and E.S. Schwartz, 1984, "A Two-factor Model of the Term Structure: An Approximate Analytical Solution," *Journal of Financial and Quantitative Analysis* 19, 413–424.

Stambaugh, R. F., 1988, "Information in Forward Rates: Implications for the Term Structure," *Journal of Financial Economics* **21**, 41–70.

Stigum, M., 1983, *The Money Market*, Dow Jones-Irwin, Homewood, Illinois.

Vasicek, O. A., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics* 5, 177–188.

White, H. (ed.), 1983, Non-nested Models, *Journal of Econometrics* 21, 1–160.

White, H., 1984, *Asymptotic Theory for Econometricians*, Academic Press, Orlando.

Table 1
Likelihood Function Values for the Various
Models Estimated Using All 10 Bonds[30]

	1972–1986	1972–1979	1979–1986
One-factor Model			
$\bar{r} = 0$	4.50182×10^3	2.77703×10^3	2.13685×10^3
$\bar{r}$ uncons.	4.50738×10^3	2.77977×10^3	2.14515×10^3
Two-factor Model			
$\bar{r} = 0$	no feasible parameter values		
$\bar{r}$ uncons.	4.93078×10^3	2.93477×10^3	2.41822×10^3

Likelihood Ratio Test Statistics

H_0 vs. H_A	1972–1986	1972–1979	1979–1986
One-factor $\bar{r} = 0$ vs. One-factor $\bar{r}$ uncons.	11.12 (χ_1^2)	6.38 (χ_1^2)	16.60 (χ_1^2)
One-factor $\bar{r} = 0$ vs. Two-factor $\bar{r} = 0$	no feasible parameter values		
One-factor $\bar{r}$ uncons. vs. Two-factor $\bar{r}$ uncons.	846.80 (χ_2^2)	310.00 (χ_2^2)	546.14 (χ_2^2)
Two-factor $\bar{r} = 0$ vs. Two-factor $\bar{r}$ uncons.	no feasible parameter values		
One-factor $\bar{r} = 0$ vs. Two-factor $\bar{r}$ uncons.	857.92 (χ_3^2)	315.48 (χ_3^2)	554.44 (χ_3^2)

[30]The asymptotic distributions of the test statistics for the hypotheses that $\bar{r} = 0$ in the one-factor model are χ_1^2. The asymptotic distributions of the test statistics comparing the one-factor model with $\bar{r}$ unconstrained and the two-factor model with $\bar{r}$ unconstrained ($\bar{r} = -10$) are χ_2^2 if one takes the value of the likelihood function with $\bar{r} = -10$ as equal to the unconstrained value, and the distributions of the test statistics comparing the one-factor model with $\bar{r} = 0$ and the two-factor model with $\bar{r}$ unconstrained are χ_3^2 if one takes the value of the likelihood function with $\bar{r} = -10$ as equal to the unconstrained value. The .05 fractiles of the χ^2 distributions with 1, 2, and 3 degrees of freedom are 3.8415, 5.9915, and 7.8147, and the .01 fractiles are 6.6349, 9.2103, and 11.3449.

Table 2
Parameter Estimates for the One-Factor
Model Estimated Using All 10 Bonds[31]

Subsample and Model	$\ln \mathcal{L}$	κ	σ	θ	λ	$\bar{r}$
1972–1986						
$\bar{r} = 0$	4.50182×10^3	0.16939	0.10232	0.08903	-0.04544	
		(0.0862)	(0.0056)	(0.0450)	(0.0857)	
$\bar{r}$ uncons.	4.50738×10^3	0.17739	0.12622	0.06231	-0.05883	0.02991
		(0.1621)	(0.0079)	(0.0569)	(0.1617)	(0.0028)
$H_0 : \bar{r} = 0$	11.12					
	(χ^2_1)					
1972–1979						
$\bar{r} = 0$	2.77703×10^3	0.47588	0.09269	0.09087	0.07791	
		(0.0594)	(0.0070)	(0.0110)	(0.0588)	
$\bar{r}$ uncons.	2.78022×10^3	0.47791	0.12315	0.05970	0.07543	0.02669
		(0.2445)	(0.0129)	(0.0304)	(0.2411)	(0.0039)
$H_0 : \bar{r} = 0$	6.38					
	(χ^2_1)					
1979–1986						
$\bar{r} = 0$	2.13685×10^3	0.29368	0.11425	0.07936	-0.12165	
		(0.1387)	(0.0086)	(0.0372)	(0.1376)	
$\bar{r}$ uncons.	2.14515×10^3	0.41112	0.16354	0.03310	-0.25236	0.05245
		(0.2620)	(0.0118)	(0.02110)	(0.2615)	(0.0006)
$H_0 : \bar{r} = 0$	16.60					
	(χ^2_1)					

Test Statistics for the Hypotheses that the Parameters
are Constant Throughout the Sample Period

Model	Test Statistic
$\bar{r} = 0$	810.66
	(χ^2_4)
$\bar{r}$ uncons.	835.98
	(χ^2_5)

[31] Standard errors are in parentheses. The asymptotic distributions of the test statistics for the hypotheses that $\bar{r} = 0$ in the first panel are χ^2_1. The asymptotic distributions of the test statistics in the second panel are χ^2_4 and χ^2_5. The .05 fractiles of the χ^2 distributions with 1, 4, and 5 degrees of freedom are 3.8415, 9.4877, and 11.0705, and the .01 fractiles are 6.6349, 13.2767, and 15.0863.

Table 3
Estimated Standard Deviations of the Errors in the Statistical Model for Bond Prices for the One-Factor Model[32]

Subsample and Model	Bond Maturity .5 year	1 year	2 years	3 years	4 years	5 years	7 years	10 years	30 years
1972–1986									
$\bar{r} = 0$	.00157	.00530	.01496	.02725	.03844	.06661	.06107	.07905	.12238
$\bar{r}$ uncons.	.00157	.00531	.01498	.02725	.03842	.06665	.06111	.07941	.12357
1972–1979									
$\bar{r} = 0$	.00124	.00408	.01050	.01779	.02471	.03416	.02912	.03932	.05587
$\bar{r}$ uncons.	.00124	.00408	.01050	.01779	.02471	.03415	.02911	.03927	.05689
1979–1986									
$\bar{r} = 0$	.00179	.00598	.01633	.02782	.03645	.06827	.05957	.07269	.12261
$\bar{r}$ uncons.	.00179	.00600	.01635	.02784	.03645	.06827	.05963	.07311	.12498

[32]The estimated standard deviations are computed under the assumption that the mean of the residuals is zero, and hence are mean-squared errors.

Table 4
Parameter Estimates for the One-Factor
Model Estimated Using 3 Treasury Bills[33]

Sample Period and Model	$\ln \mathcal{L}$	κ	σ	θ	λ	$\bar{r}$
1972–1986						
$\bar{r} = 0$	2.49727×10^3	0.39363	0.10146	0.08695	-0.10678	
		(0.0986)	(0.0054)	(0.0201)	(0.0918)	
$\bar{r}$ uncons.	2.50600×10^3	0.44218	0.12790	0.05772	-0.16246	0.02880
		(0.1462)	(0.0079)	(0.0185)	(0.14199)	(0.0020)
$H_0 : \bar{r} = 0$	17.46					
	(χ_1^2)					
1972–1979						
$\bar{r} = 0$	1.37141×10^3	0.58862	0.09154	0.08148	-.08166	
		(0.1524)	(0.0068)	(0.0185)	(0.1344)	
$\bar{r}$ uncons.	1.37417×10^3	0.62868	0.11973	0.05517	-.13105	0.02530
		(0.2438)	(0.0124)	(0.0207)	(0.2323)	(0.0041)
$H_0 : \bar{r} = 0$	5.52					
	(χ_1^2)					
1979–1986						
$\bar{r} = 0$	1.15540×10^3	0.41064	0.11373	0.08547	-0.13383	
		(0.1465)	(0.0087)	(0.0276)	(0.1332)	
$\bar{r}$ uncons.	1.16552×10^3	0.51827	0.16727	0.03723	-0.26577	0.05175
		(0.2989)	(0.0130)	(0.0210)	(0.2933)	(0.0008)
$H_0 : \bar{r} = 0$	20.24					
	(χ_1^2)					

Test Statistics for the Hypotheses that the Parameters
are Constant Throughout the Sample Period

Model	Test Statistic
$\bar{r} = 0$	59.08
	(χ_4^2)
$\bar{r}$ uncons.	66.78
	(χ_5^2)

[33]Standard errors are in parentheses. The asymptotic distributions of the test statistics for the hypotheses that $\bar{r} = 0$ in the first panel are χ_1^2. The asymptotic distributions of the test statistics in the second panel are χ_4^2 and χ_5^2. The .05 fractiles of the χ^2 distributions with 1, 4, and 5 degrees of freedom are 3.8415, 9.4877, and 11.0705, and the .01 fractiles are 6.6349, 13.2767, and 15.0863.

Table 5
Parameter Estimates for the Two-Factor Model Estimated Using All 10 Bonds[34]

Subsample and Model	$\ln \mathcal{L}$	κ	σ	$\theta + \bar{r}$	λ	κ_2	σ_2	θ_2
1972–1986								
$\bar{r} = 0$	no feasible parameter values							
$\bar{r}$ uncons.	4.93110×10^3	7.45252	0.01969	0.02641	-0.00479	0.07974	0.11703	0.09296
		(0.6269)	(0.0014)	(0.0053)	(0.0017)	(0.0089)	(0.0080)	(0.0064)
$H_0 : \bar{r} = 0$	∞ (χ_1^2)							
1972–1979								
$\bar{r} = 0$	no feasible parameter values							
$\bar{r}$ uncons.	2.93718×10^3	14.50615	0.03209	-0.03148	-0.02211	0.45896	0.07482	0.09267
		(3.7655)	(0.0108)	(0.0628)	(0.0113)	(0.0281)	(0.0287)	(0.0627)
$H_0 : \bar{r} = 0$	∞ (χ_1^2)							
1979–1986								
$\bar{r} = 0$	no feasible parameter values							
$\bar{r}$ uncons.	2.41844×10^3	5.11342	0.01868	0.04070	-0.00093	0.13172	0.15634	0.10138
		(0.4660)	(0.0015)	(0.0057)	(0.0022)	(0.0150)	(0.0118)	(0.0063)
$H_0 : \bar{r} = 0$	∞ (χ_1^2)							

Test Statistics for the Hypotheses that the Parameters are Constant Throughout the Sample Period

Model	Test Statistic
$\bar{r} = 0$	no feasible parameter values (χ_7^2)
$\bar{r}$ uncons.	849.04 (χ_8^2)

[34]Standard errors are in parentheses. The parameter $\bar{r} = -10$ in the results labelled "$\bar{r}$ uncons." The asymptotic distributions of the test statistics for the hypotheses that $\bar{r} = 0$ in the first panel are χ_1^2 if one takes the value of the likelihood function with $\bar{r} = -10$ as equal to the unconstrained value. The asymptotic distributions of the test statistics in the second panel are χ_7^2 and χ_8^2, again if one takes the value of the likelihood function with $\bar{r} = -10$ as equal to the unconstrained value. The .05 fractile of the χ^2 distribution with 1 degree of freedom is 3.8415, and the .01 fractile is 6.6349. The .05 fractiles of the χ^2 distributions with 7 and 8 degrees of freedom are 14.0671 and 15.5073, and the .01 fractiles are 18.4753 and 20.0902.

Table 6
Estimated Standard Deviations of the Errors in the Statistical Model for Bond Prices for the Two-Factor Model[35]

Subsample and Model	Bond Maturity 1 year	2 years	3 years	4 years	5 years	7 years	10 years	30 years
1972–1986								
$\bar{r} = 0$	no feasible parameter values							
$\bar{r}$ uncons.	.00206	.01052	.02182	.03286	.06069	.05346	.07021	.11121
1972–1979								
$\bar{r} = 0$	no feasible parameter values							
$\bar{r}$ uncons.	.00175	.00863	.01595	.02392	.03291	.02678	.03584	.05449
1979–1986								
$\bar{r} = 0$	no feasible parameter values							
$\bar{r}$ uncons.	.00226	.00946	.01973	.02555	.05892	.04519	.05660	.10760

[35] The estimated standard deviations are computed under the assumption that the mean of the residuals was zero, and hence are mean-squared errors.

Table 7
Likelihood Function Values for the Various Models Estimated Using 3 Treasury Bills[36]

	1972–1986	1972–1979	1979–1986
One-factor Model			
$\bar{r} = 0$	2.49727×10^3	1.37141×10^3	1.15540×10^3
$\bar{r}$ uncons.	2.50600×10^3	1.37417×10^3	1.16552×10^3
Two-factor Model			
$\bar{r} = 0$	no feasible parameter values		
$\bar{r}$ uncons.	2.69932×10^3	1.45573×10^3	1.26995×10^3

Likelihood Ratio Test Statistics

H_0 vs. H_A	1972–1986	1972–1979	1979–1986
One-factor $\bar{r} = 0$ vs. One-factor $\bar{r}$ uncons.	17.46 (χ^2_1)	5.52 (χ^2_1)	20.24 (χ^2_1)
One-factor $\bar{r} = 0$ Two-factor $\bar{r} = 0$	no feasible parameter values		
One-factor $\bar{r}$ uncons. vs. Two-factor $\bar{r}$ uncons.	386.64 (χ^2_2)	163.12 (χ^2_2)	208.86 (χ^2_2)
Two-factor $\bar{r} = 0$ Two-factor $\bar{r}$ uncons.	no feasible parameter values		
One-factor $\bar{r} = 0$ Two-factor $\bar{r}$ uncons.	202.05 (χ^2_3)	168.64 (χ^2_3)	229.10 (χ^2_3)

[36] The asymptotic distributions of the test statistics for the hypotheses that $\bar{r} = 0$ in the first panel are χ^2_1. The asymptotic distributions of the test statistics comparing the one-factor model with $\bar{r}$ unconstrained and the two-factor model with $\bar{r}$ unconstrained ($\bar{r} = -10$) are χ^2_2 if the value of the likelihood function with $\bar{r} = -10$ is taken to be equal to the unconstrained value, and the distributions of the test statistics comparing the one-factor model with $\bar{r} = 0$ and the two-factor model with $\bar{r}$ unconstrained are χ^2_3 if one takes the value of the likelihood function with $\bar{r} = -10$ as equal to the unconstrained value. The .05 fractiles of the χ^2 distributions with 1, 2, and 3 degrees of freedom are 3.8415, 5.9915, and 7.8147, and the .01 fractiles are 6.6349, 9.2103, and 11.3449.

Table 8
Parameter Estimates for the Two-Factor Model Estimated Using 3 Treasury Bills[37]

Sample Period and Model	$\ln \mathcal{L}$	κ	σ	$\theta+\bar{r}$	λ	κ_2	σ_2	θ_2
1972–1986								
$\bar{r}=0$	no feasible parameter values							
$\bar{r}$ uncons.	2.69932×10^3	8.80420	0.02244	0.02741	-0.00678	0.18168	0.13339	0.06337
		(0.9418)	(0.0023)	(0.0037)	(0.0022)	(0.0327)	(0.0093)	(0.0066)
$H_0: \bar{r}=0$	∞							
	(χ_1^2)							
1972–1979								
$\bar{r}=0$	no feasible parameter values							
$\bar{r}=-10$	1.45573×10^3	16.40774	0.03734	-0.03182	-0.02754	0.31688	0.07111	0.09426
		(4.3838)	(0.0141)	(0.0642)	(0.0150)	(0.0633)	(0.0284)	(0.0643)
$H_0: \bar{r}=0$	∞							
	(χ_1^2)							
1979–1986								
$\bar{r}=0$	no feasible parameter values							
$\bar{r}$ uncons.	1.26995×10^3	7.35352	0.02255	0.03942	-0.00276	0.26459	0.17031	0.07070
		(0.9283)	(0.0025)	(0.0047)	(0.0028)	(0.0508)	(0.0148)	(0.0067)
$H_0: \bar{r}=0$	∞							
	(χ_1^2)							

Test Statistics for the Hypotheses that the Parameters are Constant Throughout the Sample Period

Model	Test Statistic
$\bar{r}=0$	no feasible parameter values
	(χ_7^2)
$\bar{r}$ uncons.	52.72
	(χ_8^2)

[37] Standard errors are in parentheses. The parameter $\bar{r} = -10$ in the results labelled "$\bar{r}$ uncons." The asymptotic distributions of the test statistics for the hypotheses that $\bar{r} = 0$ in the first panel are χ_1^2 if one takes the value of the likelihood function with $\bar{r} = -10$ as equal to the unconstrained value. The asymptotic distributions of the test statistics in the second panel are χ_7^2 and χ_8^2, again if one takes the value of the likelihood function with $\bar{r} = -10$ as equal to the unconstrained value. The .05 fractile of the χ^2 distribution with 1 degree of freedom is 3.8415, and the .01 fractile is 6.6349. The .05 fractiles of the χ^2 distributions with 7 and 8 degrees of freedom are 14.0671 and 15.5073, and the .01 fractiles are 18.4753 and 20.0902.

Table 9

Constrained and Unconstrained Estimates of the Parameters β_{ji}[38]

Subsample and Model	$\ln \mathcal{L}$	β_{10}	β_{11}	β_{12}	β_{20}	β_{21}	β_{22}	β_{30}	β_{31}	β_{32}
1972–1986										
CIR model	2699.32	-0.0024	-0.1133	1.1111	-0.0008	-0.6364	1.6290	0.0030	-0.9448	1.9026
general model	2729.09	-0.0000	0.5879	0.4092	0.0010	-0.0374	1.0343	0.0051	-0.4340	1.3929
H_0 : CIR model	62.68 (χ^2_5)									
1972–1979										
CIR model	1455.73	-0.0039	-0.2419	1.2488	-0.0013	-0.6648	1.6574	0.0039	-0.8580	1.7930
general model	1474.46	-0.0032	0.3422	0.6797	0.0006	-0.0731	1.0725	0.0063	-0.3508	1.2800
H_0 : CIR model	43.44 (χ^2_5)									
1979–1986										
CIR model	1269.95	-0.0023	-0.0989	1.1013	-0.0003	-0.6314	1.6238	0.0057	-0.9350	1.8760
general model	1284.36	0.0031	0.1954	0.7625	0.0023	-0.3713	1.3448	0.0079	-0.7107	1.6372
H_0 : CIR model	29.56 (χ^2_5)									

Test Statistics for the Hypotheses that the Parameters are Constant Throughout the Sample Period

Model	Test Statistic
CIR model	52.72 (χ^2_8)
general model	59.46 (χ^2_{13})

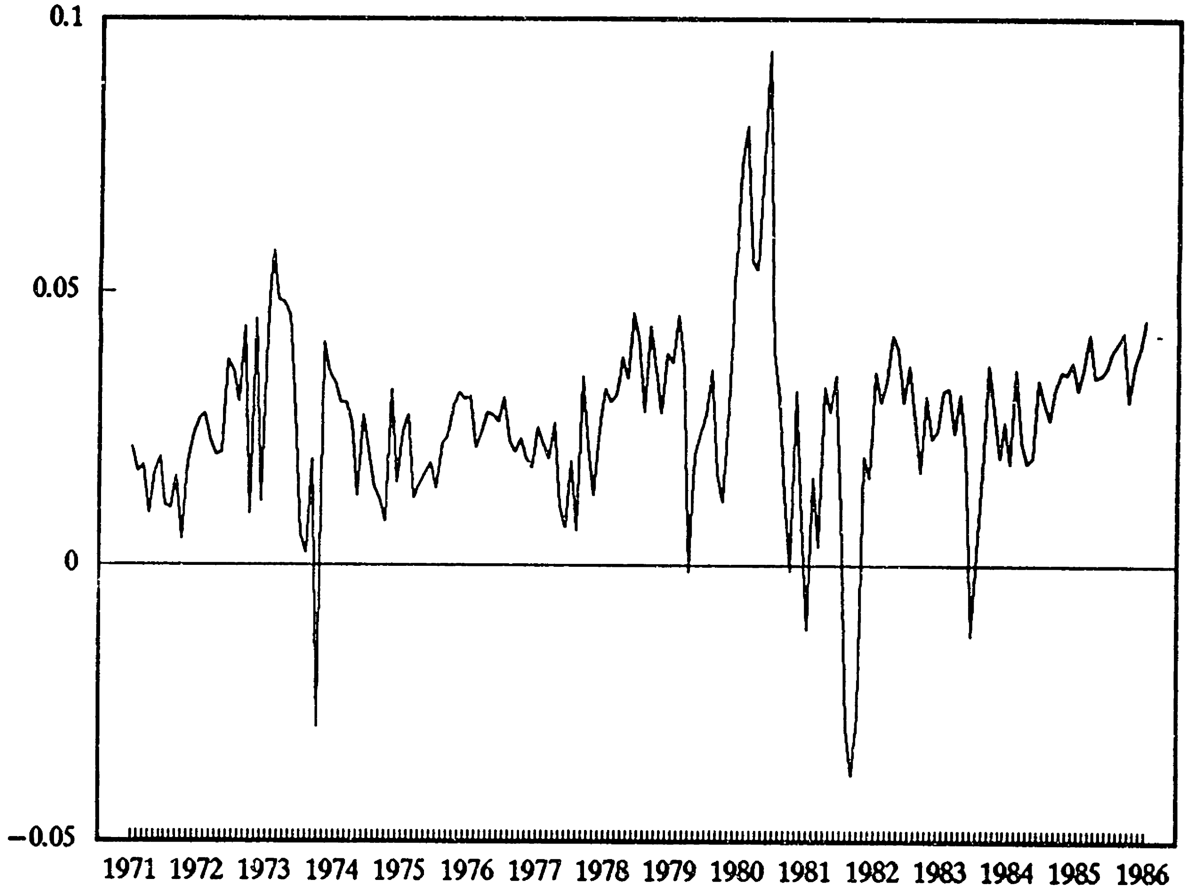
[38]The asymptotic distributions of the test statistics in the first panel for the hypotheses that the constraints are satisfied are χ^2_5. The asymptotic distributions of the test statistics in the second panel are χ^2_8 and χ^2_{13}. The .05 fractile of the χ^2 distribution with 5 degrees of freedom is 11.0705, and the .01 fractile is 15.0863. The .05 fractile of the χ^2 distributions with 8 and 13 degrees of freedom are 15.5073 and 22.3621, and the .01 fractiles are 20.0902 and 27.6883.

Table 10
Predictive Performance of Various Models:
Mean Squared Prediction Errors[39]

Forecast Error	Mean Squared Prediction Errors ($\times 10^4$) One-Factor Model	Two-Factor Model	General Markov Model	Martingale Model
1972–1986				
$Y^*(\tau_1) - \overline{Y}^*(\tau_1)$	0.7124	0.8311	0.7303	0.6905
$Y^*(\tau_2) - \overline{Y}^*(\tau_2)$	0.6465	0.7325	0.6647	0.6495
$Y^*(\tau_3) - \overline{Y}^*(\tau_3)$	0.6939	0.6838	0.6394	0.6047
1972–1979				
$Y^*(\tau_1) - \overline{Y}^*(\tau_1)$	0.3890	0.3681	0.3186	0.3455
$Y^*(\tau_2) - \overline{Y}^*(\tau_2)$	0.3140	0.2817	0.2495	0.2588
$Y^*(\tau_3) - \overline{Y}^*(\tau_3)$	0.3510	0.2848	0.2675	0.2456
1979–1986				
$Y^*(\tau_1) - \overline{Y}^*(\tau_1)$	1.0557	1.3384	1.2075	1.0593
$Y^*(\tau_2) - \overline{Y}^*(\tau_2)$	0.9998	1.2211	1.1413	1.0672
$Y^*(\tau_3) - \overline{Y}^*(\tau_3)$	1.0430	1.0946	1.0464	0.9886

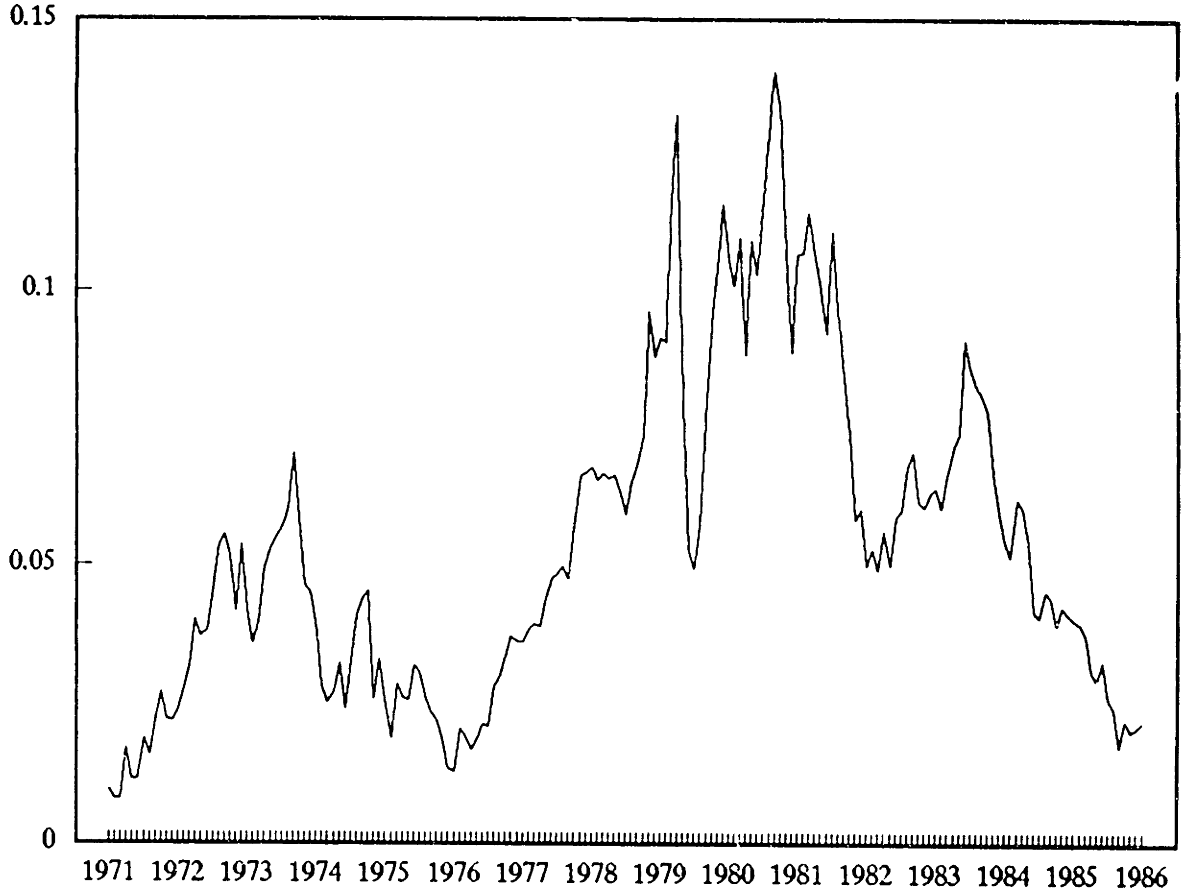
[39] The forecast errors are the differences between the actual and predicted yields. The predicted yields were calculated using the parameter estimates of the translated one and two-factor models and the general markov model for yields estimated using 3 Treasury Bills and the naive martingale model.

Figure 1
Time Series of the Interest Rate r for
the Translated Two-Factor Model, 1972–1986[40]



[40]The estimates of the values of the state variable were calculated using the parameter estimates of the translated two-factor model estimated using all 10 bonds.

Figure 2
Time Series of the State Variable y for
the Translated Two-Factor Model, 1972–1986[41]



[41] The estimates of the values of state variable were calculated using the parameter estimates of the translated two-factor model estimated using all 10 bonds.

CHAPTER 3
AN EXTENDED SIMULATED MOMENTS ESTIMATOR OF CONTINUOUS TIME ASSET PRICING MODELS

Abstract This essay presents an extended simulated moments estimator (ESME) of the parameters of continuous time asset pricing models in which the underlying state variables follow a (vector) diffusion process and asset prices can be expressed as conditional expectations of known functions of the sample path of the state vector. The ESME allows the estimation of models in which the asset pricing function is neither known in closed form nor easily evaluated numerically. The idea underlying the estimation strategy is that, when asset prices can be expressed as conditional expectations, it is possible to simulate the moments of certain functions of asset prices even when it is not feasible to compute the asset prices as functions of the underlying state variables. Specifically, moments of certain functions of the asset prices are obtained by constructing a discrete approximation of an expanded system of stochastic differential equations related to the stochastic differential equation describing the motion of the underlying state variables and using the approximation to simulate the moments of certain functions of the sample path of the expanded system. If these functions are chosen appropriately, probability limits of normalized sums of these functions are equal to the probability limits of normalized sums of certain functions of the data when the simulation is performed using the true parameter vector, and not otherwise. I show the consistency and asymptotic normality of the estimator and demonstrate its computational feasibility by using it to estimate a one-factor term structure model due to Cox, Ingersoll, and Ross.

1 Introduction

The last twenty years has seen the development of a number of continuous time asset pricing models in which the prices of certain assets can be expressed as conditional expectations of functions of the sample paths of underlying state variables. These state variables may be observed or unobserved, and may include the prices of other assets. For example, in the option pricing model of Black and Scholes (1973) there is a single underlying state variable, the price of the common stock, and the price of a European option is equal to the conditional expectation (under the appropriate probability measure) of a function of the stock price on the expiration date of the option. In the term structure model of Cox, Ingersoll, and Ross (1985b) the bond price can be written as the conditional expectation of a function of the sample path of the unobservable instantaneous riskless rate of interest, while in the stochastic volatility option pricing models of Wiggins (1987), Scott (1987), and Melino and Turnbull (1988) the prices of options depend upon the price and (unobservable) instantaneous volatility of the underlying asset. Empirical implementation of such models is hindered by the fact that maximum likelihood and method of moments estimators require that the econometrician know either the conditional density or the moments of the data generating process, and analytic expressions for these are known only for very special cases of diffusion processes. A second difficulty is that for many models an analytic expression for the asset pricing function (i.e., the appropriate conditional expectation) is known only for special cases of the underlying state variable process, or not at all. This makes estimation using existing techniques extraordinarily difficult, if not impossible.

The first difficulty, that of adducing either the conditional density or the moments of the data generating process, can be circumvented by using the estimation strategy of Duffie and Singleton (1989) (DS). They advance a simulated moments estimator (SME) of the parameters of asset pricing models in which the underlying state variables follow general markov processes and the asset pricing function is either known in closed form or easily evaluated numerically.[1] The basic idea of their approach is to simulate the moments of the data generating process and choose the estimate of the parameter vector to minimize the (appropriately measured) distance between the simulated and sample moments. In essence, they use simulated moments to replace the population moments in the generalized method of moments (GMM) of Hansen and Singleton (1982), in which the parameter vector is chosen to minimize the distance between the population and sample moments. The simulated moments of continuous time processes are obtained by simulating a finely spaced discrete approximation of the state process, computing the asset prices as functions of the state vector, and then sampling the approximating processes at an interval corresponding to the observation interval of the econometrician.

This essay addresses the second problem hindering the empirical implementation of continuous time asset pricing models, namely that in many models the asset pricing function is known in closed form only for very special cases of the underlying state variable process, or not at all. I present an extended simulated moments estimator (ESME) of the parameters of asset pricing models in which the underlying state variables follow continuous time markov processes and the endogenously determined asset prices can be expressed as conditional expectations of known functions of the sample paths of the state variables.[2] The idea underlying the estimation strategy in this paper is that, when asset prices can be written as conditional expectations, it is possible to simulate the moments of functions of asset prices even when it is not possible to compute the asset prices as functions of the underlying state variables.

The estimation strategy is as follows. The point of departure is the SME of DS. The first step is to simulate the underlying state variable process exactly as in DS. At this point the approach of DS would involve computing the asset prices as functions of the underlying state variables

[1] Lee and Ingram (1989) have also proposed a simulated moments estimator of time series models MacFadden (1989) and Pakes and Pollard (1989) have studied simulation-based estimators in cross-sectional settings.

[2] The requirement that the endogenously determined asset prices be expectations of known functions of the underlying state variables precludes using the ESME with assets that have an "American" feature unless it can be shown that this feature has no value, as, for example, is the case with an American call option on a non-dividend paying stock.

and forming a moment condition using moments of functions of the simulated state variables and asset prices and moments of the same functions of the data. However, in this paper it is assumed that the asset prices, i.e. the conditional expectations of functions of the sample paths of the state variables, cannot be computed. Therefore functions of the simulated asset prices cannot be computed and the moment condition cannot be formed. The estimation strategy of the ESME is to replace the functions of the simulated asset prices with estimates consisting of certain functions of simulated sample paths of an expanded set of stochastic processes related to the state variable processes. Under the appropriate probability measure, the conditional expectations, and therefore expectations, of the differences between the functions of asset prices and the estimates have mean zero. Under regularity conditions these differences satisfy a law of large numbers, and moment conditions constructed using the simulated asset prices can be replaced with moment conditions constructed using the estimates of the simulated asset prices.

To be more precise, let $f : \Re^A \to \Re^B$ be a measurable function, where A is the dimension of the data vector to be used in estimation and B is the dimension of the moment condition. In the simulated moments estimator of DS, estimates are obtained by choosing the estimates to minimize the (appropriately measured) distance between certain sample moments calculated from the data and the sample moments of a simulated stochastic process generated using the equation of motion which describes the data. The DS moment condition is of the form $\frac{1}{T}\sum_{t=0}^{T}\bar{f}_t - \frac{1}{T(T)}\sum_{s=0}^{T(T)} f_s^\beta$, where $\bar{f}_t$ is a function of the data observed at time t and f_s^β denotes the same function of a sequence simulated using the parameter vector β. The simulated moments are used to replace the expectation in the GMM moment condition of Hansen and Singleton (1982) $\frac{1}{T}\sum_{t=0}^{T}\bar{f}_t - E(f_t^\beta)$, where the expectation is taken with respect to the stationary distribution of the data generating process.

I perform additional simulations to construct estimates $\hat{f}_s^\beta$ of the functions f_s^β used in the DS moment conditions. Let $\mathcal{P}$ be the probability measure on the space on which the simulations are defined,[3] and let $\mathcal{G}_s$ denote the sigma-field of events distinguishable at time s. I use a moment condition $m_T^{T(T)}(\beta) = \frac{1}{T}\sum_{t=0}^{T}\bar{f}_t - \frac{1}{T(T)}\sum_{s=0}^{T(T)}\hat{f}_s^\beta$, where $E^{\mathcal{P}}(\hat{f}_s^\beta - f_s^\beta|\mathcal{G}_s) = 0$ and the sequence of differences $\hat{f}_s^\beta - f_s^\beta$ admits a law of large numbers. This estimation strategy is useful when the asset price is given by the conditional expectation of a function of the sample path of a stochastic process and this conditional expectation cannot be computed in closed form. In this case the vector f_s^β also cannot be computed, and the DS moment condition cannot be formed. In this paper I describe how to construct certain functions $\hat{f}_s^\beta$ that have the property that $E^{\mathcal{P}}(\hat{f}_s^\beta - f_s^\beta|\mathcal{G}_s) = 0$. Estimation can then be based upon simulated moments of these functions.

[3]The simulated process is independent of the data generating process. The probability space underlying the estimator is the product space formed from the probability space upon which the process assumed to generate the data is defined and the probability space upon which the simulation is defined.

While the estimator involves a great deal of simulation, it is computationally feasible even when the state space is of relatively high dimension. In contrast, maximum likelihood, GMM, and SME estimators are not computationally feasible when the asset pricing function is unknown and cannot be computed at reasonable cost. The simulated moments estimator of DS overcomes the problem of adducing the conditional density or moments of the underlying state variables, but still requires that one be able to compute the asset prices. While asset prices in the continuous time models of interest satisfy partial differential equations, solution of these equations by finite difference methods is extraordinarily difficult and costly when the state space is of even moderately high dimension. For this reason, estimation of models for which an analytic expression for the asset pricing function is not known and the state space is of even moderately high dimension is essentially impossible using existing techniques.

The key to the approach in this paper is to perform the simulation on the appropriate probability space. The simulated state variable process is defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, P)$ which is a "copy" of the probability space upon which the process assumed to generate the data is defined. Ergodicity assumptions that give a law of large numbers and a central limit theorem are made about processes defined on this probability space. However, the asset prices are given by conditional expectations with respect to probability measures Q^t equivalent to the conditional probabilities of P. No such property holds under the conditional probabilities of P. I need to construct a probability space, set of stochastic processes defined on this space, and estimates of asset prices that mimic the property of Q^t that asset prices are given by expectations and also allow a law of large numbers and a central limit theorem. I must also chose the functions $\hat{f}_s^\beta$ so that $E^{\mathcal{P}}(\hat{f}_s^\beta - f_s^\beta | \mathcal{G}_s) = 0$.

In order to accomplish these goals, the simulated state variable process is, as above, defined on a "copy" of the probability space upon which the data generating process is defined. Additional simulated processes related to the state variable process are used to construct the functions $\hat{f}_s^\beta$, for $s = 0, 1, \ldots, \mathcal{T}$. These are defined on probability spaces $(\Omega_i^s, \mathcal{F}_i^s, \mathbf{F}_i^s, P_i^s)$, for $s = 0, 1, \ldots, \mathcal{T}$ and $i = 1, \ldots, M$, where M is the order of the highest order moment used in estimation. The entire set of simulations used to generate both the state variable process and the functions $\hat{f}_s^\beta$ is defined on a rather complicated product space constructed from $(\Omega, \mathcal{F}, \mathbf{F}, P)$ and $(\Omega_i^s, \mathcal{F}_i^s, \mathbf{F}_i^s, P_i^s)$, for $s = 0, 1, \ldots, \mathcal{T}$, and $i = 0, 1, \ldots, M$. The probability measure $\mathcal{P}$ on this space is constructed from both P and the P_i^s.[4] This probability space, the processes defined

[4]One need not be concerned about the fact that the probability space upon which the simulated processes are defined has a different structure than the probability space upon which the processes assumed to generate the data are defined. The estimator requires that the moments of the simulated processes and the moments of the data have the same probability limits when the simulation is performed using the true parameter vector, and have different probability limits otherwise, but no other features of the simulated processes, or the probability space upon which they are defined, are important. In particular, it is not necessary that the probability space upon which the simulated processes are defined have the same structure as the probability space upon which the processes assumed to generate the data are defined.

on it, and the functions $\hat{f}_s^\beta$ have the property that $E^{\mathcal{P}}(\hat{f}_s^\beta - f_s^\beta | \mathcal{G}_s) = 0$.

Bossaerts and Hillion (1989) have advanced a simulation-based estimator of models of the pricing of "European" type contingent claims which can be used when an analytic expression for the asset pricing function is not known and the state vector is observable.[5] Their procedure involves choosing the model parameters to minimize the distance between a time-series cross-section of observed contingent claims prices and estimates of theoretical prices obtained by simulation. At each point in time, a cross-sectional simulation using a moderately large number of draws is used to estimate each asset price.[6] The limitation that the state vector must be observed arises because the values of the state variables at each point in time are needed to simulate the contingent claims prices. Another limitation of the procedure is that it does not exploit any knowledge of the dynamics of the underlying state vector that the econometrician might possess. That is, Bossaerts and Hillion do not make any use of the implications of the underlying model about the time-series properties of asset prices. In particular, they do not use second or higher order moments of the contingent claims prices, or of lagged moments.

The estimation strategy in this paper, including the construction of the probability space upon which the simulated processes are defined, the processes to be simulated, and the construction of the moment conditions, is somewhat complicated. In order to make clear the basic idea, in the next section of the paper (Section 2) I present the approach in the context of a simple one-factor bond pricing model. I begin by discussing the SME of DS, and then show how the ESME can be constructed by expanding the probability space on which the simulated process is defined and performing additional simulations.

Section 3 presents the setting for which the estimator is intended, the simulated processes, and the estimator.

In the models I consider, the underlying state variables follow diffusion processes. Moments of certain functions of the asset prices are obtained by constructing discrete approximations of the stochastic differential equation describing the motion of the underlying state variables and the additional processes used to construct the estimates of the asset prices and using the approximating processes to simulate the moments of the functions $\hat{f}_s^\beta$. Existing literature on approximating stochastic differential equations in mean square provides bounds on the mean square error only over a fixed time interval $[0, \mathcal{T}]$ (see, for example, Pardoux and Talay (1985) or Gard (1988)). However, in the asymptotic theory one must let $\mathcal{T} \to \infty$, and the existing results are not adequate because the bound on the mean square error is a function of $\mathcal{T}$.

One strategy is to take the approximating discrete-time process as the true probability model, and assume that asset prices can be expressed as conditional expectations of sample

[5] Bossaerts (1989) studies the extension of the ideas of Bossaerts and Hillion (1989) to the case when the option may be exercised at any of a finite number of times prior to expiration.

[6] Bossaerts and Hillion (1989) use between 20 and 100 draws to estimate each option price.

paths of the approximating process. This approach is taken in Section 4, where the consistency and asymptotic normality of the estimator are shown under the assumption that the approximating process is the true probability model.

An alternative strategy is to show that the approximating process for the state variables converges in mean square to the underlying diffusion process for all times $s \in [0, \infty)$. In Section 5 I pursue this strategy for certain bond pricing models. I show that the approximating process converges in mean square to the underlying diffusion for all $s \in [0, \infty)$, and that the same convergence holds for the moments to be used in estimation. This allows me to obtain the stronger result that the ESME is consistent and asymptotically normal for the parameters of the continuous time bond pricing models.

In Section 6, I demonstrate the computational feasibility of the ESME by using it to estimate the parameters of a term structure model. The model I estimate is the basic one-factor model of the price of a real bond due to CIR (1985b) in which the instantaneous riskless rate of interest follows a mean-reverting process with linear drift and diffusion coefficient proportional to the square root of the interest rate. For the purpose of demonstrating the feasibility of the ESME I reinterpret this model as a model of the prices of nominal bonds. This particular model was chosen because analytic expressions for the moments of the bond prices can be obtained and the model can be estimated by the GMM. This allows me to compare the estimates from the ESME to those obtained using GMM.

Section 7 concludes the paper.

2 An Example

The ideas can be illustrated in the context of a one state variable model for bond prices due to CIR (1985b). The relevant features of this model are as follows.

2.1 The Model

Fix a filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{F}}, \tilde{P})$ for the time set $[0, T]$ where $\tilde{\mathbf{F}} = \{\tilde{\mathcal{F}}_t; t \geq 0\}$ is the filtration generated by a standard Brownian Motion in $\Re$. On this probability space define a stochastic process

$$d\tilde{r}(t) = \kappa_0(\theta_0 - \tilde{r}(t))dt + \sigma_0\sqrt{\tilde{r}(t)}d\widetilde{W}(t), \tag{1}$$

with initial condition $\tilde{r}(0) = \tilde{r}_0$. The partial differential equation satisfied by the price at time t of a bond with maturity τ, $\widetilde{H}(\tilde{r}(t), \beta_0)$, is (CIR (1985b))

$$1/2\sigma_0^2\tilde{r}\widetilde{H}_{\tilde{r}\tilde{r}} + (\kappa_0\theta_0 - (\kappa_0 + \lambda_0)\tilde{r})\widetilde{H}_{\tilde{r}} - \widetilde{H}_\tau - \tilde{r}\widetilde{H} = 0, \tag{2}$$

with boundary condition $\widetilde{H}(\tilde{r}(t+\tau),0,\beta_0)=1$. Here λ_0 denotes the market price of interest rate risk, and $\beta_0 \equiv (\kappa_0,\sigma_0,\theta_0,\lambda_0)'$.

Alternatively, the bond price can be expressed as an expectation.

Let $\xi^t(t')$, for $t' \geq t$ and $t=0,1,\dots,T$, be the processes defined by[7]

$$\xi^t(t') = \exp\left[-\int_t^{t'} \frac{\lambda_0\sqrt{\tilde{r}(u)}}{\sigma_0} d\widetilde{W}(u) - \frac{1}{2}\int_t^{t'} \frac{\lambda_0^2 \tilde{r}(u)}{\sigma_0^2} du\right]$$

with initial condition $\xi^t(t)=1$. Define probabilities Q^t, $t=0,1,\dots,T$ by

$$Q^t(A) = E^{\widetilde{P}}[1_A \xi^t(t+\tau) \mid \widetilde{\mathcal{F}}_t], \quad \text{for } A \in \widetilde{\mathcal{F}}_{t+\tau}. \tag{3}$$

The probability $\widetilde{P}$ and its conditional probabilities are to be interpreted as the beliefs of the agents in the economy, while Q^t is a "risk-adjusted" probability. For $t' \in [t,t+\tau]$, the interest rate satisfies

$$d\tilde{r}(t') = (\kappa_0\theta_0 - (\kappa_0+\lambda_0)\tilde{r}(t'))dt' + \sigma_0\sqrt{\tilde{r}(t')}d\widehat{W}^t(t'),$$

where by Girsanov's theorem $\widehat{W}^t(t') = W(t) + \int_t^{t'} \frac{\lambda_0\sqrt{\tilde{r}(u)}}{\sigma_0} du$ is a Brownian Motion under Q^t. From CIR (1985b), the price at time t of a bond with maturity τ is given by the expectation

$$\widetilde{H}(\tilde{r}(t),\beta_0) = E^{Q^t}\left(e^{-\int_t^{t+\tau}\tilde{r}(u)du} \mid \widetilde{\mathcal{F}}_t\right). \tag{4}$$

Now consider another probability space $(\Omega^t,\mathcal{F}^t,\mathbf{F}^t,P^t)$ for the time set $[t,t+\tau]$, and define

$$dr^t(t') = (\kappa_0\theta_0 - (\kappa_0+\lambda_0)r^t(t'))dt' + \sigma_0\sqrt{r^t(t')}dW^t(t'),$$

where $r^t(t) = \tilde{r}(t)$ and W^t is a Brownian Motion under P^t. Conditional on the information available at time t, the random variable $e^{-\int_t^{t+\tau} r^t(u)du}$ has the same distribution under P^t as $e^{-\int_t^{t+\tau}\tilde{r}(u)du}$ has under Q^t, and, similar to (4),

$$\widetilde{H}(\tilde{r}(t),\beta_0) = E^{P^t}\left(e^{-\int_t^{t+\tau} r^t(u)du} \mid \widetilde{\mathcal{F}}_t\right). \tag{5}$$

This expression for the bond price is the basis of the estimation procedure.[8]

[7] Throughout, time superscripts on processes indicate that the processes start at that time. For example, the process $\xi^t(t')$ is defined for $t' \geq t$.

[8] From CIR (1985b),

$$\widetilde{H}(\tilde{r}(t),\beta_0) = A(\tau,\beta_0)e^{-B(\tau,\beta_0)\tilde{r}(t)},$$

where

$$A(\tau,\beta_0) \equiv \left[\frac{2\gamma e^{(\gamma+\lambda+\kappa)\tau/2}}{(\gamma+\lambda+\kappa)(e^{\gamma\tau}-1)+2\gamma}\right]^{2\kappa\theta/\sigma^2},$$

While the GMM or the SME of DS could be used to estimate this particular bond pricing model, in general the bond pricing function is known only for very special cases of the underlying state variable process. Since DS assume that the asset price is a known function of the state variables and it is only the moments of the asset price that are unknown, in general their approach cannot be used.[9] The problem in estimating bond pricing models is precisely that the bond pricing function is known only for very restrictive assumptions about the dynamics of the underlying state variables. The object is to see how this difficulty can be overcome using an extension of the simulated moments estimator.

2.2 The Simulated Moments Estimator

As a point of departure, I describe the simulated moments estimator of DS might be applied to this bond pricing model.[10] It is assumed that the data are generated by (1) and (4), and therefore that $\beta_0 = (\kappa_0, \sigma_0, \theta_0, \lambda_0)'$ is the true value of the parameter vector. The econometrician observes the values of the process (4) at times $t = 0, 1, \ldots, T$.

Consider a second probability space $(\Omega, \mathcal{F}, \mathbf{F}, P)$ on the time set $[0, \mathcal{T}]$ where $\mathbf{F} = \{\mathcal{F}_s; s \geq 0\}$ is the filtration generated by a standard Brownian Motion in $\Re$. On this probability space define a stochastic process

$$dr(s) = \kappa(\theta - r(s))ds + \sigma\sqrt{r(s)}dW(s), \tag{6}$$

with initial condition $r(0) = r_0$. Define another stochastic process

$$H(r(s), \beta) = A(\tau, \beta)e^{-B(\tau,\beta)r(s)}, \tag{7}$$

where $\beta \equiv (\kappa, \sigma, \theta, \lambda)'$ and $A(\tau, \beta)$ and $B(\tau, \beta)$ are defined in footnote 8. It is assumed that the econometrician simulates the processes (6) and (7). In estimation, the econometrician will use values of the process (7) observed at times $s = 0, 1, \ldots, \mathcal{T}$.

The probability space underlying the SME is the product space $(\tilde{\Omega} \times \Omega, \tilde{\mathcal{F}} \times \mathcal{F}, \tilde{P} \times P)$ where, for any $A \in \tilde{\mathcal{F}}_t$, $B \in \mathcal{F}_s$, $(\tilde{P} \times P)(A \times B) = \tilde{P}(A)P(B)$.

$$\begin{aligned} B(\tau, \beta_0) &\equiv \frac{2(e^{\gamma\tau} - 1)}{(\gamma + \lambda + \kappa)(e^{\gamma\tau} - 1) + 2\gamma}, \\ \gamma &\equiv ((\kappa + \lambda)^2 + 2\sigma^2)^{1/2}. \end{aligned}$$

[9] While finite difference methods could be used to compute the asset prices in models with only one state variable, they are not computationally feasible when the state space is of higher dimension because most estimation problems require that the asset pricing function be evaluated hundreds or thousands of times.

[10] Of course, the generalized method of moments of Hansen and Singleton (1982) could be applied to this model because the joint distribution functions of the interest rate are known and analytic expressions for the moments of both yields and prices can be found.

Let $\tilde{Z}_t = [\tilde{H}(r(t),\beta_0), \tilde{H}(r(t-1),\beta_0), \ldots, \tilde{H}(r(t-\ell),\beta_0)]$ and $Z_s = [H(r(s),\beta), H(r(s-1),\beta), \ldots, H(r(s-\ell),\beta)]$ for some positive integer $\ell < \infty$, let $f : \Re^{\ell} \to \Re^B$, where $B \geq 4$, be a measurable function, and consider the sequences $\{\tilde{f}_t = f(\tilde{Z}_t)\}_{t=0}^{T}$ (the data) and $\{f_s^{\beta} = f(Z_s)\}_{s=0}^{T}$ (the simulated observations). In the simulated moments estimator, estimation is based upon sample moments of these two sequences. Specifically, the econometrician forms a moment condition

$$m_T^{\mathcal{T}(T)}(\beta) \equiv \frac{1}{T}\sum_{t=0}^{T}\tilde{f}_t - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)} f_s^{\beta}, \tag{8}$$

and chooses the estimator b_T to make this moment condition close to zero.

I emphasize that the probability space upon which the simulation is performed need not be a "copy" of the probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{F}}, \tilde{P})$ upon which the processes assumed to generate the data are defined, and the simulated process need not be a "copy" of the process used to generate the data.[11] It is necessary only that those moments of the simulated process used in estimation agree with the moments of the data generating process when $\beta = \beta_0$ and disagree otherwise. While the simulated moments estimator of DS does not require that the econometrician take advantage of this freedom, I exploit it below with the extended simulated moments estimator. I find it necessary to introduce a complicated probability space upon which to define the simulated processes. The simulated processes defined on that space are of much higher dimension that the processes assumed to generate the data.

[11] For example, consider a filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, P)$ on the time set $[0, \mathcal{T}]$ where $\mathbf{F} = \{\mathcal{F}_s; s \geq 0\}$ is the filtration generated by a standard Brownian Motion in $\Re^2$. On this probability space define stochastic processes $r(s)$ and $H(r(s),\beta)$ as follows:

$$\begin{aligned} r(s) &= 2\theta \quad \text{for } 0 \leq s \leq 1, \\ r(s) &= r(1) + \int_1^s \kappa(\theta - r(u))du + \int_1^s \begin{bmatrix} \frac{\sigma}{\sqrt{2}}\sqrt{r(u)} & \frac{\sigma}{\sqrt{2}}\sqrt{r(u)} \end{bmatrix} \begin{bmatrix} dW_1(u) \\ dW_2(u) \end{bmatrix} \quad \text{for } s \geq 1, \\ H(r(s),\beta) &= A(\tau,\beta)e^{-B(\tau,\beta)r(s)}. \end{aligned}$$

Simulated moments estimation could be based upon this process because the ergodic distribution of $H(r(s),\beta)$ is identical to that of $\tilde{H}(\tilde{r}(t),\beta_0)$ when $\beta = \beta_0$.

While this is a silly example, it makes the points that the probability space upon which the simulated process is defined need not be identical to the probability space upon which the process assumed to generate the data is defined, and the simulated process need not be identical to the process assumed to generate the data. All that is required is that the econometrician define a simulated proce and construct sequences of functions $\{\tilde{f}_t\}_{t=0}^{T}$ and $\{\hat{f}_s^{\beta}\}_{s=0}^{\mathcal{T}(T)}$ with the property that $m_T(\beta) = \frac{1}{T}\sum_{t=0}^{T}\tilde{f}_t - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)} f_s^{\beta} \to 0$ in probability when $\beta = \beta_0$, and not otherwise.

In practice, noone would ever define the simulated processes on a probability space more complicated than the probability space on which the data generating process is defined, and noone would ever define a simulated process that differed from the data generating process unless inability to observe the initial condition of the data generating process forced one to do so. In this case the econometrician presumably would construct a simulated process that differed from the data generating process only in the initial condition. It is nonetheless true that the econometrician has some f dom in constructing the simulation.

2.3 The Extended Simulated Moments Estimator

The idea of the extension is to replace the simulated functions f_s^β with estimates $\hat{f}_s^\beta$, where $E^{\mathcal{P}}(\hat{f}_s^\beta - f_s^\beta | \mathcal{G}_s) = 0$ and the sequence of differences $\{\hat{f}_s^\beta - f_s^\beta\}$ admits a law of large numbers. (Here $\mathcal{P}$ denotes the probability measure on the probability space upon which the simulation is defined, while $\mathcal{G}_s$ is the sub-sigma-field of events distinguishable at time s.) It turns out that it is not necessary that the functions $\hat{f}_s^\beta$ be precise estimates of the functions f_s^β. Moment conditions to replace the condition (8) of DS can be formed using only a few simulations to obtain the estimates $\hat{f}_s^\beta$ used to replace the functions f_s^β used in (8). In particular, it is not necessary to perform enough additional simulation or other computation to obtain a precise estimate of the function for each $s \in \{0, 1, \ldots, \mathcal{T}\}$.

The data is assumed to be generated by the model described in subsection 2.1 above. There is an underlying filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{F}}, \tilde{P})$ for the time set $[0, T]$, the interest rate follows (1), and the observed bond prices are given by the conditional expectation (4). I proceed by first simulating the state variable process on a "copy" of the probability space underlying the data generating process. I then construct a product space and a set of processes defined on it and use these to construct the functions $\hat{f}_s^\beta$.

2.3.1 The simulated processes

The state variable process

Fix a filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, P)$ for the time set $[0, \mathcal{T}]$ where $\mathbf{F} = \{\mathcal{F}_s; s \geq 0\}$ is the filtration generated by a standard Brownian Motion in $\Re$. On this probability space define a stochastic process for the underlying state variable

$$dr(s) = \kappa(\theta - r(s))ds + \sigma\sqrt{r(s)}dW(s), \tag{9}$$

with initial condition $r(0) = r_0$. This is the same probability space used in the SME in Subsection 2.2 above, and the state variable process (9) is identical to (6).

Let $\xi^s(s')$, for $s' \geq s$ and $s = 0, 1, \ldots, \mathcal{T}$, be the processes defined by

$$\xi^s(s') = \exp\left[-\int_s^{s'} \frac{\lambda\sqrt{r(u)}}{\sigma} dW(u) - \frac{1}{2}\int_s^{s'} \frac{\lambda^2 r(u)}{\sigma^2} du\right].$$

Define probabilities Q^s, for $s = 0, 1, \ldots, \mathcal{T}$, by $Q^s(A) = E^P[1_A \xi^s(s + \tau) | \mathcal{F}_s]$, for $A \in \mathcal{F}_{s+\tau}$. The process on $[s, s + \tau]$ defined by $\widehat{W}^s(s) = 0$, $d\widehat{W}^s(s') = dW(s') + \frac{\lambda}{\sigma}\sqrt{r(s')}ds'$ is a Brownian Motion under Q^s, and, for $s' \geq s$, the simulated state variable process (9) can be written

$$dr(s') = (\kappa\theta - (\kappa + \lambda)r(s'))ds' + \sigma\sqrt{r(s')}d\widehat{W}(s'), \tag{10}$$

The simulated bond price is given by the conditional expectation

$$H(r(s),\beta) = E^{Q^s}(e^{-\int_s^{s+\tau} r(u)du} \mid \mathcal{F}_s).$$

The strategy below is to construct random variables that have the same conditional distribution as integer powers of $e^{-\int_s^{s+\tau} r(u)du}$ and use these random variables in the moment conditions.

The product space and the processes defined on it

For each integer s, $s = 0, 1, \ldots, \mathcal{T}$, construct M probability spaces $(\Omega_i^s, \mathcal{F}_i^s, \mathbf{F}_i^s, P_i^s)$, for $i = 1, \ldots, M$, on the time set $[s, s+\tau]$ where $\mathbf{F}_i^s = \{\mathcal{F}_{i,u}^s; s \le u \le s+\tau\}$ is the filtration generated by a standard Brownian Motion in $\Re$. Construct the product space $\Omega \times \left(\times_{s=0}^{\mathcal{T}} \left(\times_{i=1}^{M} \Omega_i^s\right)\right)$, with product sub-sigma-field at time u given by $\mathcal{G}_u = \mathcal{F}_u \times \left(\times_{s=0}^{j} \left(\times_{i=1}^{M} \mathcal{F}_{i,u}^s\right)\right)$, where j is the largest integer less than or equal to u. The filtration is $\mathbf{G} = \{\mathcal{G}_s; s \ge 0\}$ and the product measure $\mathcal{P}$ is $\mathcal{P}(B) = \mathcal{P}(A \times \left(\times_{s=0}^{j} \left(\times_{i=1}^{M} A_i^s\right)\right) = P(A) \prod_{s=0}^{j} \prod_{i=1}^{M} P_i^s(A_i^s)$ where $B \in \mathcal{G}_u$, $A \in \mathcal{F}_u$, and $A_i^s \in \mathcal{F}_{i,u}^s$.

On each of the probability spaces $(\Omega_i^s, \mathcal{F}_i^s, \mathbf{F}_i^s, P_i^s)$, define a process

$$dr_i^s(s') = (\kappa\theta - (\kappa + \lambda) r_i^s(s'))ds' + \sigma\sqrt{r_i^s(s')}dW_i^s(s'), \tag{11}$$

Define[12] $\widehat{H}_i(r(s),\beta) = e^{-\int_s^{s+\tau} r_i^s(u)du}$, for $i = 1, \ldots, M$. The processes (11) are identical in distribution to the simulated state variable process under Q^s given by (10), and the distribution of $\widehat{H}_i(r(s),\beta)$ under P_i^s is identical to the distribution of $e^{-\int_s^{s+\tau} r(u)du}$ under Q^s. Below I repeatedly use the following fact:

$$\begin{aligned} E^{\mathcal{P}}(\widehat{H}_i(s,\beta) \mid \mathcal{G}_s) &= E^{P_i^s}(\widehat{H}_i(s,\beta) \mid \mathcal{F}_s) \\ &= E^{Q^s}(e^{-\int_s^{s+\tau} r(u)du} \mid \mathcal{F}_s) \\ &= H(r(s),\beta). \end{aligned}$$

The first equality follows from the construction of the probability $\mathcal{P}$, the second follows from the fact that the conditional (on $\mathcal{F}_s$) distribution of $\widehat{H}_i(r(s),\beta)$ under P_i^s is identical to the conditional distribution of $e^{-\int_s^{s+\tau} r(u)du}$ under Q^s, and the third equality is a property of the bond pricing model.

2.3.2 A law of large numbers for the first moment

It is now possible to obtain a law of large numbers under $\mathcal{P}$, i.e., to show that certain functions of the simulated processes converge in probability to certain functions of the data when $\beta = \beta_0$. That this is possible for the first moment is suggested by the fact that

[12] Throughout this section, the exposition is as though one could actually compute this function.

$$
\begin{aligned}
&E^{\mathcal{P}}(\widehat{H}_i(r(s),\beta) - H(r(s),\beta) \mid \mathcal{G}_0) \\
&\quad = E^{\mathcal{P}}\left(E^{\mathcal{P}}(\widehat{H}_i(r(s),\beta) - H(r(s),\beta) \mid \mathcal{G}_s) \mid \mathcal{G}_0\right) \\
&\quad = E^{P}\left(E^{P_i^s}(\widehat{H}_i(r(s),\beta) - H(r(s),\beta) \mid \mathcal{F}_{i,s}^s) \mid \mathcal{F}_0\right) \\
&\quad = E^{P}(0 \mid \mathcal{F}_0) \\
&\quad = 0,
\end{aligned}
$$

for $i = 1, \ldots, M$, and the differences $\widehat{H}_i(r(s),\beta) - H(r(s),\beta)$ are bounded. What I actually use is the fact that after appropriate rescaling of the time variables in (11) each of the sequences

$$
\{\widehat{H}_i(r(s),\beta) - H(r(s),\beta), \mathcal{G}_{s+\tau}\}_{s=0}^{\mathcal{T}(T)} \tag{12}
$$

is a martingale difference sequence.

Suppose $\tau \leq 1$, and consider the sequence (12). The sequence $\{\widehat{H}_i(r(s),\beta) - H(r(s),\beta)\}_{s=0}^{\mathcal{T}(T)}$ is adapted to the sequence $\{\mathcal{G}_{s+\tau}\}_{s=0}^{\mathcal{T}(T)}$, and

$$
\begin{aligned}
E^{\mathcal{P}}(\widehat{H}_i(r(s),\beta) - H(r(s),\beta) \mid \mathcal{G}_s) &= E^{P_i^s}(\widehat{H}_i(r(s),\beta) - H(r(s),\beta) \mid \mathcal{F}_{i,s}^s) \\
&= 0,
\end{aligned}
$$

so (12) is a martingale difference sequence. The sequence (12) is not a martingale difference sequence when $\tau > 1$. However, one can make $\tau \leq 1$ by appropriate rescaling of the time variable in the stochastic processes (11). Below, without loss of generality I assume $\tau = 1$ whenever it is useful for (12) or a similar sequence to be a martingale difference sequence.

That (12) is a martingale difference sequence, combined with the geometric ergodicity of the state variable process, allows one to prove a law of large numbers for a moment condition involving the first moment. Specifically, $\frac{1}{T}\sum_{t=0}^{T} \widetilde{H}(\tilde{r}(t),\beta_0) - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)} \widehat{H}_i(r(s),\beta) \to 0$ in probability when $\beta = \beta_0$.

From the triangle inequality,

$$
\begin{aligned}
&\left|\frac{1}{T}\sum_{t=0}^{T} \widetilde{H}(\tilde{r}(t),\beta_0) - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)} \widehat{H}_i(r(s),\beta)\right| \\
&= \left|\frac{1}{T}\sum_{t=0}^{T} \widetilde{H}(\tilde{r}(t),\beta_0) - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)} H(r(s),\beta) + \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)} (H(r(s),\beta) - \widehat{H}_i(r(s),\beta))\right| \\
&\leq \left|\frac{1}{T}\sum_{t=0}^{T} \widetilde{H}(\tilde{r}(t),\beta_0) - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)} H(r(s),\beta)\right| + \left|\frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)} (H(r(s),\beta) - \widehat{H}_i(r(s),\beta))\right|
\end{aligned}
$$

The first term to the right of the inequality converges to zero in probability exactly as in DS.

For the second term, it was shown above that $\{H(r(s),\beta) - \widehat{H}_i(r(s),\beta), \mathcal{G}_{s+\tau}\}_{s=0}^{\mathcal{T}}$ is a martingale difference sequence under $\mathcal{P}$. Also, from the properties of the exponential function, $E^{\mathcal{P}}(|H(r(s),\beta) - \widehat{H}_i(r(s),\beta)|^2|\mathcal{G}_0) < 1$ for all s. Therefore the second term converges to zero almost surely, and hence in probability (White (1984), Theorem 3.77, Exercise 3.78).

2.3.3 A law of large numbers for the second moment

The moment conditions involving the higher moments are slightly more complicated. Construct $\widehat{H}_i(r(s),\beta)$ as before and then construct $\widehat{H}_j(r(s),\beta)$ as $\widehat{H}_j(r(s),\beta) \equiv e^{-\int_s^{s+\tau} r_j^s(u)du}$ where $r_j^s(u)$ $(j \neq i)$ is given by (11). Then

$$\begin{aligned} E^{\mathcal{P}}(\widehat{H}_i(r(s),\beta)\widehat{H}_j(r(s),\beta) \mid \mathcal{G}_s) &= E^{P_i^s \times P_j^s}(\widehat{H}_i(r(s),\beta)\widehat{H}_j(r(s),\beta) \mid \mathcal{F}_s) \\ &= E^{P_i^s}(\widehat{H}_i(r(s),\beta) \mid \mathcal{F}_s)E^{P_j^s}(\widehat{H}_j(r(s),\beta) \mid \mathcal{F}_s) \\ &= (E^{Q^s}(e^{-\int_s^{s+\tau} r(u)du} \mid \mathcal{F}_s))^2 \\ &= H(r(s),\beta)^2. \end{aligned}$$

Here $P_i^s \times P_j^s$ is the product measure on the product space constructed from $(\Omega_i^s, \mathcal{F}_i^s, \mathbf{F}_i^s, P_i^s)$ and $(\Omega_j^s, \mathcal{F}_j^s, \mathbf{F}_j^s, P_j^s)$. The first equality follows from the construction of the probability $\mathcal{P}$, the second follows from the independence of $\widehat{H}_i(r(s),\beta)$ and $\widehat{H}_j(r(s),\beta)$ given $\mathcal{F}_s$, the third follows from the fact that the conditional (on $\mathcal{F}_s$) distribution of $\widehat{H}_i(r(s),\beta)$ under P_i^s is identical to the conditional distribution of $e^{-\int_s^{s+\tau} r(u)du}$ under Q^s, and the fourth equality is a property of the bond pricing model.

Also, after rescaling the time variables in (11), the sequence $\{\widehat{H}_i(r(s),\beta)\widehat{H}_j(r(s),\beta) - H(r(s),\beta)^2, \mathcal{G}_{s+\tau}\}_{s=0}^{\mathcal{T}(T)}$ is a martingale difference sequence. This, together with the geometric ergodicity of the state variable process, allows one to conclude that

$$\frac{1}{T}\sum_{t=0}^{T} \widetilde{H}(\tilde{r}(t),\beta_0)^2 - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)} \widehat{H}_i(r(s),\beta)\widehat{H}_j(r(s),\beta) \to 0 \qquad (13)$$

in probability when $\beta = \beta_0$.

Conditions involving the higher moments can be constructed using additional interest rate processes r_k^s, $k \neq i,j$.

As indicated above, the estimates of bond prices $\widehat{H}_i(r(s),\beta)$ cannot be computed exactly. Also, in general, the econometrician will not be able to simulate exactly the state variable process. The strategy is to use the discrete time approximation described in subsection 3.3 below.

3 The Extended Simulated Moments Estimator

In this section I describe the setting for which the estimator is intended and then turn to the simulated processes and the construction of the moment conditions.

3.1 The data generating process

Fix a filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{F}}, \tilde{P})$ for the time set $[0, T]$ where $\tilde{\mathbf{F}} = \{\tilde{\mathcal{F}}_t; t \geq 0\}$ is the filtration generated by a standard Brownian Motion in $\Re^p$. On this probability space define an L-dimensional stochastic process

$$d\tilde{X}(t) = \mu(\tilde{X}(t), \beta_0)dt + \sigma(\tilde{X}(t), \beta_0)d\widetilde{W}(t), \tag{14}$$

with initial condition $\tilde{X}(0) = \tilde{X}_0$, where $\widetilde{W}(t) = [\widetilde{W}_1(t), \ldots, \widetilde{W}_p(t)]'$ is a standard Brownian Motion in $\Re^p$. Here $\beta_0 \in \Theta \subset \Re^q$ denotes the vector of unknown parameters to be estimated. This process describes the motion of a vector of state variables which may include the prices of certain assets. It is not necessary that the econometrician observe the state variable process $\tilde{X}$.

The econometrician observes a K-dimensional vector of asset prices $\tilde{H}(t)$ which are functions of the state vector $\tilde{X}(t)$, the time to maturity or expiration of the asset τ, and the vector of unknown parameters β_0,

$$\tilde{H}(t) = \tilde{H}(\tilde{X}(t), \beta_0). \tag{15}$$

The econometrician knows neither an analytic expression for the function $\tilde{H}$ nor the dynamics of the asset prices (i.e., the functional form of the drift and diffusion coefficients).[13] However, there exists a measurable function $g : \tilde{\mathbf{X}}^t \times \Theta \to \Re^K$, known to the econometrician, and a probability Q^t such that, for every time t,

$$\tilde{H}(\tilde{X}(t), \beta_0) = E^{Q^t}(g(\tilde{x}^t, \beta_0) \mid \mathcal{F}_t), \tag{16}$$

where $\tilde{\mathbf{X}}^t$ denotes the set of sample paths of the process $\tilde{X}$ on $[t, t+\tau]$ and $\tilde{x}^t$ denotes one element of $\tilde{\mathbf{X}}^t$. The process $\tilde{X}$ satisfies a stochastic differential equation

$$d\tilde{X}(u) = \hat{\mu}(\tilde{X}(u), \beta_0)du + \sigma(\tilde{X}(u), \beta_0)d\widehat{W}^t(u), \tag{17}$$

on the time set $[t, t+\tau]$, where $\widehat{W}^t(u) = [\widehat{W}_1^t(u), \ldots, \widehat{W}_p^t(u)]'$ is a Brownian Motion in $\Re^p$ under Q^t.

[13] The case when the dynamics of the asset prices are known has been treated by DS. If the dynamics of some elements of the vector $\tilde{H}(t)$ are known or some elements of $\tilde{X}(t)$ are observable the approach here may be combined with that of DS.

For example, in the bond pricing model of CIR given by (1) and (4) above the state variable process is $d\tilde{X}(t) = d\tilde{r}(t) = \kappa_0(\theta_0 - \tilde{r}(t))dt + \sigma_0\sqrt{\tilde{r}(t)}d\widetilde{W}(t)$, and the bond price at time t, $\tilde{H}(\tilde{r}(t),\beta_0)$, is given by $\tilde{H}(\tilde{r}(t),\beta_0) = E^{Q^t}(e^{-\int_t^{t+\tau}\tilde{r}(u)du}|\mathcal{F}_t)$, where Q^t is defined by (3). For $u \geq t$, $d\tilde{X}(u) = d\tilde{r}(u) = (\kappa_0\theta_0 - (\kappa_0 + \lambda_0\tilde{r}(u)))du + \sigma_0\sqrt{\tilde{r}(u)}d\widehat{W}^t(u)$.

The functions μ, $\hat{\mu}$, and σ satisfy growth and uniform Lipshitz conditions for all $\beta \in \Theta$. That is, there exists a constant $K > 0$ such that

$$\begin{aligned} |\mu(x,\beta) - \mu(y,\beta)| + |\hat{\mu}(x,\beta) - \hat{\mu}(y,\beta)| + |\sigma(x,\beta) - \sigma(y,\beta)| &\leq K|x-y|, \\ |\mu(x,\beta)|^2 + |\hat{\mu}(x,\beta)|^2 + |\sigma(x,\beta)|^2 &\leq K^2(1+|x|)^2. \end{aligned}$$

This assumption is made for all $\beta \in \Theta$ in order to include the simulated processes below.

3.2 The simulated processes

Ideally, the econometrician would simulate a continuous-time state variable process along with the set of "risk-adjusted" processes used to generate estimates of the functions $\hat{f}_s^\beta$. In general, it is impossible to simulate exactly the continuous time processes, and a discrete time approximation must be used. I begin by indicating the underlying continuous-time process that the econometrician will approximate, and then indicate the discrete approximation.

The state variable process

Fix a filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, P)$ for the time set $[0,\mathcal{T}]$ where $\mathbf{F} = \{\mathcal{F}_s; s \geq 0\}$ is the filtration generated by a standard Brownian Motion in $\Re^p$. On this probability space define an L-dimensional stochastic process for the simulated state variable

$$dX(s) = \mu(X(s),\beta)ds + \sigma(X(s),\beta)dW(s), \tag{18}$$

with initial condition $X(0) = X_0$, where $W(s) = [W_1(s),\ldots,W_p(s)]'$ is a Brownian Motion in $\Re^p$.

The simulated processes are constructed so as to inherit certain properties of the state variable process. In particular, for all $\beta \in \Theta$, there exists a measurable function $g : \mathbf{X}^s \times \Theta \to \Re^K$ and a probability Q^s such that, for every time s, the simulated asset prices

$$H(X(s),\beta) \equiv E^{Q^s}(g(x^s,\beta) \mid \mathcal{F}_s) \tag{19}$$

are well defined, where $\mathbf{X}^s$ denotes the set of sample paths of the process X on $[s, s+\tau]$ and x^s denotes one element of $\mathbf{X}^s$. The process X satisfies the stochastic differential equation

$$dX(u) = \hat{\mu}(X(u),\beta)du + \sigma(X(u),\beta)d\widehat{W}^s(u), \tag{20}$$

on the interval $[s, s+\tau]$, where $\widehat{W}^s(u) = [\widehat{W}_1^s(u), \ldots, \widehat{W}_p^s(u)]'$ is a Brownian Motion under Q^s defined on $[s, s+\tau]$.

The product space and the "risk-adjusted" processes defined on it

Let M be the order of the highest order moment used in estimation. For each integer s, $s = 0, 1, \ldots, \mathcal{T}$, construct M probability spaces $(\Omega_i^s, \mathcal{F}_i^s, \mathbf{F}_i^s, P_i^s)$, for $i = 1, \ldots, M$, on the time set $[s, s+\tau]$ where $\mathbf{F}_i^s = \{\mathcal{F}_{i,u}^s; s \le u \le s+\tau\}$ is the filtration generated by a standard Brownian Motion in $\Re^p$. The simulation is performed on the product space $\Omega \times \left(\times_{s=0}^{\mathcal{T}} \left(\times_{i=1}^{M} \Omega_i^s\right)\right)$, with sub-sigma-field at time u given by $\mathcal{G}_u = \mathcal{F}_u \times \left(\times_{s=0}^{j} \left(\times_{i=1}^{M} \mathcal{F}_{i,u}^s\right)\right)$, where j is the largest integer less than or equal to u. The filtration is $\mathbf{G} = \{\mathcal{G}_s; s \ge 0\}$, and the product measure $\mathcal{P}$ is $\mathcal{P}(B) = \mathcal{P}(A \times \left(\times_{s=0}^{j} \left(\times_{i=1}^{M} A_i^s\right)\right) = P(A) \prod_{s=0}^{j} \prod_{i=1}^{M} P_i^s(A_i^s)$ where $B \in \mathcal{G}_u$, $A \in \mathcal{F}_u$, and $A_i^s \in \mathcal{F}_{i,u}^s$. For each s, $s = 0, 1, \ldots, \mathcal{T}$, define M stochastic processes on the time set $[s, s+\tau]$

$$dX_i^s(s) = \hat{\mu}(X_i^s(u), \beta)du + \sigma(X_i(u), \beta)dW_i^s(u), \tag{21}$$

with initial condition $X_i^s(s) = X(s)$, $i = 1, \ldots, M$, where $W_i^s(u)$ defined on $[s, s+\tau]$ is a standard Brownian Motion under P_i^s in $\Re^p$.

Let x_i^s denote a sample path of the process $X_i^s(u)$. The estimation strategy is to replace the simulated asset prices (19) with functions $g(x_i^s, \beta)$ where the differences $g(x_i^s, \beta) - E^{\mathcal{P}}(g(x_i^s, \beta)|\mathcal{G}_s)$ form a martingale difference sequence. The law of large numbers below requires that one control certain moments of the martingale difference sequence. A law of large numbers holds if, for all $\beta \in \Theta$ and all $s \in \{0, \ldots, \mathcal{T}\}$, $E^{\mathcal{P}}[(g(x_i^s, \beta) - E^{\mathcal{P}}(g(x_i^s, \beta)|\mathcal{G}_s))^{2q}|\mathcal{G}_0] < \Delta < \infty$ for some $q \ge 1$.

3.3 Approximation of the simulated processes

The state variable process

Divide each unit of time in n equal subdivisions, and consider the equally spaced partition $0 = s_0 < s_1 < \ldots < s_k < \ldots < s_{n\mathcal{T}} = \mathcal{T}$ of the interval $[0, \mathcal{T}]$. Draw a sequence of normal random variables $\{\epsilon(k)\}_{k=1}^{n\mathcal{T}}$. Define

$$\begin{aligned} X^n(s_0) &= X(0), \\ X^n(s_{k+1}) &= X^n(s_k) + \Delta X^n(s_k), \\ \Delta X^n(s_k) &= h\mu(X^n(s_k), \beta) + \sqrt{h}\sigma(X^n(s), \beta)\epsilon(k+1), \end{aligned} \tag{22}$$

where $h = 1/n$. The sequence $\{X^n(s_k)\}_{k=0}^{n\mathcal{T}}$ is the approximate solution to (18). Construct a sequence $\{X^{\beta n}(s_{nj})\}_{j=0}^{\mathcal{T}} = \{X^n(s)\}_{s=0}^{\mathcal{T}}$ from every n-th term of $\{X^n(s_k)\}_{k=0}^{n\mathcal{T}}$. The time increments between elements of this sequence are equal to the observation interval of the econometrician.

The additional simulated processes

Let M denote the order of the highest order moment used in estimation. For each element of $\{X^{\beta n}(s)\}_{s=0}^{\mathcal{T}}$, generate M sets of innovations $\{\epsilon_i^s(k)\}_{k=1}^{n\tau}$, $i = 1, \ldots, M$, and construct M simulated processes $X_i^{s,n}$ on the interval $[s, s+\tau]$ as follows:

$$\begin{aligned} X_i^{s,n}(s) &= X^n(s), \\ X_i^{s,n}(s+h(k+1)) &= X_i^{s,n}(s+hk) + \Delta X_i^{s,n}(s+hk), \\ \Delta X_i^{s,n}(s+hk) &= h\hat{\mu}(X_i^{s,n}(s+hk), \beta) + \sqrt{h}\sigma(X_i^{s,n}, \beta)\epsilon_i(k+1), \\ X_i^{s,n}(u) &= X_i^{s,n}(s_k) \quad \text{for } s+hk \leq u < s+h(k+1). \end{aligned} \tag{23}$$

Then, for each element of $\{X^n(s)\}_{s=0}^{\mathcal{T}}$, construct the functions $g(x_i^{s,n}, \beta)$, $i = 1, \ldots, M$. For fixed $\mathcal{T}$, the mean square convergence of the approximation is a standard result (see, e.g., Gard (1988)) when $X_i^{s,n}(s) = X^n(s)$: $E^{\mathcal{P}}[X_i^{s,n}(u) - (X_i^s(u))^2|\mathcal{G}_0] \leq Kh$ and $E^{\mathcal{P}}[(X_i^{s,n}(u) - X_i^s(u))^2|\mathcal{G}_0] \leq Kh$ for $s \leq u \leq s+\tau$ as $h \to 0$.

3.4 Construction of the estimator

Estimation of the vector β_0 is based upon functions of the data $\tilde{f}_t \equiv \tilde{f}(\widetilde{H}(\widetilde{X}(t), \beta_0))$ and simulated state process $\hat{f}_s^{\beta n} \equiv \hat{f}(g(x_i^{s,n}, \beta)$. The j-th components of these functions are given by

$$\tilde{f}_{t,j} = \prod_{i=1}^{m(j)} \widetilde{H}(\widetilde{X}(t-\ell(i,j)), \beta_0), \tag{24}$$

$$\hat{f}_{s,j}^{\beta n} = \prod_{i=1}^{m(j)} g(x_i^{s-\ell(i,j),n}, \beta), \tag{25}$$

where $1 \leq m(j) \leq M$ and $\ell(i,j)$, for $i = 1, \ldots, m$ is a nonnegative integer less than ∞.

Suppose that a sample of T observations $\{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_T\}$ defined by (24) is available for estimation. For any particular $\beta \in \Theta$ and simulation sample size $\mathcal{T}(T)$, one can generate $\mathcal{T}(T)$ simulated observations $\{\hat{f}_1^{\beta n}, \hat{f}_2^{\beta n}, \ldots, \hat{f}_{\mathcal{T}(T)}^{\beta n}\}$ defined by (25) from representation (22) of the state variable process $\{X^n\}$ and representation (23) of the processes $\{X_i^{s,n}\}$. Let $m_{nT}^{\mathcal{T}(T)} : \Theta \to \Re^B$ be defined by

$$m_T^{n\mathcal{T}(T)} = \frac{1}{T}\sum_{t=0}^{T} \tilde{f}_t - \frac{1}{\mathcal{T}(T)} \sum_{s=0}^{\mathcal{T}(T)} \hat{f}_s^{\beta n}. \tag{26}$$

Let $\{W_T\}$ be a sequence of $M \times M$ positive semi-definite matrices (possibly depending on the sample information) of rank greater than or equal to Q. The (extended) simulated moments estimator for β_0 is then the sequence defined by

$$b_T = \arg\min_{\beta \in \Theta} m_T^{n\mathcal{T}(T)}(\beta)' W_T m_T^{n\mathcal{T}(T)}(\beta).$$

4 Large sample properties

Proving the consistency of the estimator described above requires that one control the approximation error in (22) as $\mathcal{T} \to \infty$. Unfortunately, it is not sufficient simply to let $n \to \infty$. Existing literature on approximating stochastic differential equations in mean square provides bounds on the mean square error only over a fixed time interval $[0, \mathcal{T}]$. However, in the asymptotic theory one must let $\mathcal{T} \to \infty$, and the existing results are not adequate because the bound on the mean square error is a function of $\mathcal{T}$. In fact, the bound is typically exponential in $\mathcal{T}$, and in general the consistency of the estimator will require that n grow more than exponentially with $\mathcal{T}$. Given that the total number of psuedo-random variables required to compute the moment condition is of the order of $n\mathcal{T}$, such a result is of limited value.

One strategy is to take a discrete-time process such as (22) as the true probability model for the data, and assume that asset prices can be expressed as conditional expectations of sample paths of the discrete-time process. In this section the consistency and asymptotic normality of the ESME are shown under this assumption.

4.1 The data generating process

The process assumed to be followed by the data is as follows. Divide each unit of time into n equal subdivisions, and consider the equally spaced partition $0 = t_0 < t_1 < \ldots < t_k < \ldots < t_{nT} = T$ of the interval $[0, T]$. Draw a sequence of normal random vectors $\{\epsilon(k)\}_{k=1}^{nT}$. Define

$$\begin{aligned} \widetilde{X}^n(t_0) &= \widetilde{X}(0), \\ \widetilde{X}^n(t_{k+1}) &= \widetilde{X}^n(t_k) + \Delta\widetilde{X}^n(t_k), \\ \Delta\widetilde{X}^n(t_k) &= h\mu(\widetilde{X}^n(t_k), \beta_0) + \sqrt{h}\sigma(\widetilde{X}^n(t_k), \beta_0)\epsilon(k+1). \end{aligned} \tag{27}$$

For the balance of this section, let $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbf{F}}, \widetilde{P})$ denote the probability space in which the elements of $\widetilde{\Omega}$ consist of sequences $\{\epsilon(k)\}_{k=1}^{nT}$, $\widetilde{\mathcal{F}}$ is the sigma-field generated by the sequences $\{\epsilon(k)\}_{k=1}^{nT}$, the filtration $\widetilde{\mathbf{F}} = \{\widetilde{\mathcal{F}}_t; t \geq 0\}$, and $\epsilon(k) \sim N(0, I_p)$ under $\widetilde{P}$, where I_p denotes the p-dimensional identity matrix.

Assumption 1 *For every time t there exists a stochastic process $\widetilde{X}^{t,n}$ defined on $[t, t+\tau]$, a measurable function $g : \widetilde{\mathbf{X}}^t \times \Re_+ \times \Theta \to \Re^K$, and a probability $Q^{t,n}$ such that*

$$\widetilde{H}^n(\widetilde{X}^n(t), \beta_0) = E^{Q^{t,n}}(g(\tilde{x}^{t,n}, \beta) \mid \mathcal{F}_t) \tag{28}$$

where $\widetilde{\mathbf{X}}^t$ denotes the set of sample paths of the process $\widetilde{X}^{t,n}$ and $\tilde{x}^{t,n}$ denotes one element of $\widetilde{\mathbf{X}}^t$. The process $\widetilde{X}^{t,n}(u)$ is defined by

$$\begin{aligned} \widetilde{X}^{t,n}(t) &= X(t), \\ \widetilde{X}^{t,n}(t+h(k+1)) &= \widetilde{X}^{t,n}(t+hk) + \Delta\widetilde{X}^{t,n}(t+hk), \\ \Delta\widetilde{X}^{t,n}(t+hk) &= h\hat{\mu}(\widetilde{X}^{t,n}(t+hk), \beta_0) + \sqrt{h}\sigma(\widetilde{X}_i^{\beta_n,t}, \beta_0)\epsilon(k+1), \\ \widetilde{X}^{t,n}(u) &= \widetilde{X}^{t,n}(t+hk) \quad \text{for } t+hk \leq u < t+h(k+1). \end{aligned} \tag{29}$$

where $\epsilon(k) \sim N(0, I_p)$ under $Q^{t,n}$.

The processes (27), (28), and (29) are approximations of (14), (16), and (17).

Assumption 2 *The data are generated by (27) and (28).*

The econometrician observes a sequence of K-dimensional vectors of asset prices $\{\widetilde{H}^n(\widetilde{X}^n(t), \beta_0)\}_{t=0}^T$ given by (28) and makes use of this sequence in estimation.

4.2 The simulated processes

Given that the data are assumed to follow a discrete time process, I proceed by making assumptions about the discrete time data generating processes (27) and (28) and the discrete time simulated processes (22) and (23). For the balance of this section, let $(\Omega, \mathcal{F}, \mathbf{F}, P)$ denote the probability space in which the elements of Ω consist of sequences $\{\epsilon(k)\}_{k=1}^{nT}$, let $\mathcal{F}$ denote the sigma-field generated by the sequences $\{\epsilon(k)\}_{k=1}^{nT}$, let the filtration $\mathbf{F} = \{\mathcal{F}_s; s \geq 0\}$, and let $\epsilon(k) \sim N(0, I_p)$ under P. Also, use $(\Omega_i^s, \mathcal{F}_i^s, \mathbf{F}_i^s, P_i^s)$ to denote the probability spaces in which the elements of Ω_i^s consist of sequences $\{\epsilon_i^s(k)\}_{k=1}^{n\tau}$, $\mathcal{F}_i^s$ is the sigma-field generated by the sequences $\{\epsilon_i^s(k)\}_{k=1}^{n\tau}$, the filtration $\mathbf{F}_i^s = \{\mathcal{F}_{i,u}^s; s \leq u \leq s+\tau\}$, and $\epsilon_i^s(k) \sim N(0, I_p)$ under P_i^s.

Assumption 3 *For all $\beta \in \Theta$ and for every time s there exist M stochastic processes $X_i^{s,n}$, $i = 1, 2, \ldots, M$, defined on $[s, s+\tau]$ by (23), a measurable function $g : \mathbf{X}^s \times \Theta \to \Re^K$, and M conditional probabilities Q^s such that the simulated asset prices*

$$H(X^n(s), \beta) = E^{P_i^s}(g(x_i^{s,n}, \beta) \mid \mathcal{F}_s) \quad (i = 1, 2, \ldots, M)$$

are well defined, where $\mathbf{X}_i^s$ *denotes the set of sample paths of the process* $X_i^{s,n}$ *and* $x_i^{s,n}$ *denotes one element of* $\mathbf{X}^s$.

Let the j-th component of $\tilde{f}_t$, denoted $\tilde{f}_{t,j}$, be defined by

$$\tilde{f}_{t,j} = \prod_{i=1}^{m(j)} \tilde{H}(\tilde{X}^n(t-\ell(i,j)),\beta_0), \tag{30}$$

and let $\hat{f}_{s,j}^{\beta n}$ be defined by (25).

The following assumptions and preliminaries (which, except for assumption (4), are identical to or variants of assumptions in DS) are needed to obtain the asymptotic properties of the estimator.

Assumption 4 *For all* $\beta \in \Theta$, $E^{\mathcal{P}}[(\hat{f}_{s,j}^{\beta n} - f_{s,j}^{\beta n} \mid \mathcal{G}_0)^{2q}] < \Delta < \infty$ *for some* $q \geq 1$, *all* j, *and all* s.

Assumption (4) restricts the moments of the terms of the martingale difference sequence and is needed for the law of large numbers below.

Let $\epsilon^k = (\epsilon_1, \ldots, \epsilon_k)$ denote the sequence of random variables used to construct the simulated state variable process $X^{\beta n}(s)$ and view $f_s^{\beta n} = E^{\mathcal{P}}(\hat{f}_s^{\beta n}|\mathcal{F}_s)$ as a function from $\Re^{pk} \times \Theta$ to $\Re^B$, i.e. $f_s^{\beta n} = f_s^*(\epsilon^k, \beta)$,

Definition 1 *The family* $\{f_s^{\beta n}\} = \{f_s^*(\epsilon^k,\beta) : \beta \in \Theta; s = 1,2,...\}$ *is locally Lipschitz-smooth if (i) for each* $\alpha \in \Theta$ *there is a* $\delta > 0$ *and Lipschitz constant* $K_t(\epsilon^k)$ *such that, with* $||\beta - \alpha|| \leq \delta$,

$$||f_s^*(\epsilon^k,\beta) - f_s^*(\epsilon^k,\alpha)|| \leq K_s(\epsilon^k)||\beta - \alpha||,$$

where $K_s(\epsilon^k)$ *may depend on* α, *and the sequence* $\{\mathcal{T}^{-1}\sum_{s=1}^{\mathcal{T}} K_s(\epsilon^k)\}_{s=1}^{\infty}$ *is bounded in probability.*

Lemma 1 *Suppose, for each* $\beta \in \Theta$, *that* $\{X_s^n\}$ *is ergodic and that* $E(|f_s^{\beta n}|) < \infty$. *Suppose, in addition, that* $E(f_\infty^{\beta n})$ *is a continuous function of* β *and the family* $\{f_s^{\beta n}\}$ *is Lipschitz-smooth. Then, for any* $\epsilon > 0$,

$$\lim_{\mathcal{T}\to\infty} P\left[\sup_{\beta\in\Theta}\left|E^P(f_\infty^{\beta n}) - \frac{1}{\mathcal{T}}\sum_{s=1}^{\mathcal{T}} f_s^{\beta n}\right| > \epsilon\right] = 0.$$

The proof may be found in DS.

For $q \in (0,\infty)$, let $||X||_q = [E(||X||^q)]^{1/q}$ denote the L^q norm of the random variable X.

Assumption 5 *For all* $\beta \in \Theta$, $\{||f_s^{\beta n}||_{2+\delta};\ t = 1,2,\ldots\}$ *is bounded for some* $\delta > 0$. *The family* $\{f_s^{\beta n}\}$ *is Lipschitz-smooth and* $E^{\mathcal{P}}(f_\infty^{\beta n})$ *is a continuous function of* β.

Assumption 6 *For all $\beta \in \Theta$, the state variable processes $\{X_t^n\}$ given by (27) and the simulated state variable process $\{X_s^n\}$ given by (22) are geometrically ergodic.*

Sufficient conditions for the geometric ergodicity of a process are discussed in DS.

Define

$$\tilde{\Sigma}^n \equiv \sum_{j=-\infty}^{\infty} E^{\tilde{P}}([\tilde{f}_t - E^{\tilde{P}}(\tilde{f}_t)][\tilde{f}_{t-j} - E^{\tilde{P}}(\tilde{f}_{t-j})]')$$

and

$$\hat{\Sigma}^n \equiv \sum_{j=-\infty}^{\infty} E^{\mathcal{P}}([\hat{f}_s^{\beta_0 n} - E^{\mathcal{P}}(\hat{f}_s^{\beta_0 n})][\hat{f}_{s-j}^{\beta_0 n} - E^{\mathcal{P}}(\hat{f}_{s-j}^{\beta_0 n})]').$$

Assumption (6) guarantees the convergence of these sums.

Assumption 7 *$T/\mathcal{T}(T) \to c$ as $T \to \infty$ and $W_T \to W_0 = (\tilde{\Sigma} + c\hat{\Sigma}^n)^{-1}$, in probability.*

Assumption 8 *$m_\infty^{n\mathcal{T}(T)}(\beta_0)' W_0 m_\infty^{n\mathcal{T}(T)}(\beta_0) < m_\infty^{n\mathcal{T}(T)}(\beta)' W_0 m_\infty^{n\mathcal{T}(T)}(\beta)$ for $\beta \in \Theta$, $\beta \neq \beta_0$.*

Let E denote expectation with respect to the probability on the product space formed from $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{F}}, \tilde{P})$ and $(\Omega \times \left(\times_{s=0}^{\mathcal{T}}\left(\times_{i=1}^{M}\Omega_i^s\right)\right), \mathcal{G}, \mathbf{G}, \mathcal{P})$, let $f_s^{\beta n} = E^{\mathcal{P}}(\hat{f}_s^{\beta n}|\mathcal{G})$, and define $\psi_s^n \equiv \hat{f}_s^{\beta n} - f_s^{\beta n}$. We have the following theorem.

Theorem 1 *Under assumptions 1–8, the extended simulated moments estimator converges in probability to the true parameter β_0 as $T \to \infty$.*

PROOF. We have

$$\begin{aligned}
&\left|\left(\frac{1}{T}\sum_{t=0}^{T}\tilde{f}_t - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}\hat{f}_s^{\beta n}\right) - (E\tilde{f}_\infty - Ef_\infty^{\beta n})\right| \\
\leq\ & \left|\frac{1}{T}\sum_{t=0}^{T}\tilde{f}_t - E\tilde{f}_\infty\right| + \left|Ef_\infty^{\beta n} - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}\hat{f}_s^{\beta n}\right| \\
=\ & \left|\frac{1}{T}\sum_{t=0}^{T}\tilde{f}_t - E\tilde{f}_\infty\right| + \left|Ef_\infty^{\beta n} - \left(\frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}(f_s^{\beta n} + \psi_s^n)\right)\right| \\
\leq\ & \left|\frac{1}{T}\sum_{t=0}^{T}\tilde{f}_t - E\tilde{f}_\infty\right| + \left|Ef_\infty^{\beta n} - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}f_s^{\beta n}\right| + \left|\frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}\psi_s^n\right|.
\end{aligned}$$

The first two terms on the right-hand side converge to zero in probability as in DS Theorem 1.

The third term on the right-hand side converges to zero a.s. and hence in probability from Assumption 3 and the fact that $\{\psi_s, \mathcal{G}_{s+\tau}\}$ form a martingale difference sequence.

The balance of the proof follows the proof of Theorem 1 in DS.

∎

Assumption 9 *(i) β_0 and the estimators $\{b_T\}$ are interior to Θ. (ii) $\hat{f}_s^{\beta n}$ is continuously differentiable with respect to β for all s. (iii) $D_0 \equiv E(\partial f_\infty^{\beta_0 n}/\partial\beta)$ exists, is finite, and has full rank.*

Assumption 10 *The family $\{\frac{\partial}{\partial\beta}\hat{f}_s^{\beta n} : \beta \in \Theta, s = 1, 2, \ldots\}$ is Lipschitz-smooth. For all $\beta \in \Theta$, $E(|\frac{\partial}{\partial\beta}\hat{f}_\infty^{\beta n}|) < \infty$, and $E(\frac{\partial}{\partial\beta}\hat{f}_\infty^{\beta n})$ is a continuous function of β.*

Then

Theorem 2 *Suppose $T/\mathcal{T}(T) \to c$ as $T \to \infty$, and let $\Sigma^n = \tilde{\Sigma} + c\hat{\Sigma}^n$. Under Assumptions 1–10, $\sqrt{T}G_{nT}^{\mathcal{T}(T)}(\beta_0)$ converges in distribution to a normal random vector with mean zero and covariance matrix $\Sigma^n = \tilde{\Sigma} + c\hat{\Sigma}^n$.*

PROOF. From the definition of $m_{nT}^{\mathcal{T}(T)}(\beta)$,

$$\sqrt{T}m_{nT}^{\mathcal{T}(T)}(\beta_0) = \left(\frac{1}{\sqrt{T}}\sum_{t=0}^{T}[\tilde{f}_t - E(\tilde{f}_\infty)]\right) - \frac{\sqrt{T}}{\sqrt{\mathcal{T}(T)}}\left(\frac{1}{\sqrt{\mathcal{T}(T)}}\sum_{s=0}^{\mathcal{T}(T)}[\hat{f}^{\beta_0 n} - E(\hat{f}_\infty^{\beta_0 n})]\right)$$

The two terms on the right hand side are independent. As in DS Theorem 2 (p. 15), the limiting distribution of the first term is $N[0, \tilde{\Sigma}]$. For the second term,

$$\frac{\sqrt{T}}{\sqrt{\mathcal{T}(T)}}\left(\frac{1}{\sqrt{\mathcal{T}(T)}}\sum_{s=0}^{\mathcal{T}(T)}[\hat{f}^{\beta_0} - E(\hat{f}_\infty^{\beta_0})]\right) \Rightarrow N[0, c\hat{\Sigma}^n].$$

Together, we have

$$\sqrt{T}m_{nT}^{\mathcal{T}(T)}(\beta_0) \to N[0, \Sigma^n].$$

∎

Corollary 1 *Under the assumptions of Theorem 2, $\sqrt{T}(b_T - \beta_0)$ converges in distribution to a normal random vector with mean zero and covariance matrix*

$$\Lambda = (D_0'(\Sigma^n)^{-1}D_0)^{-1}.$$

PROOF. See the discussion in DS pp. 14–16. ∎

The covariance matrix is larger than the one obtained in DS because the functions $f_s^{\beta n}$ are replaced by estimates $\hat{f}_s^{\beta n}$. Still, if the simulated sample size $\mathcal{T}(T)$ is large relative to T, it is theoretically possible to obtain the covariance matrix of the GMM estimator. More computation will be required because $\tilde{\Sigma} + c\hat{\Sigma}^n > \tilde{\Sigma}$. Nonetheless, this estimator eliminates the need to evaluate the asset pricing function at each draw of the simulation.

5 Simulated Moments Estimation of a Class of Bond Pricing Models

A limitation of the preceding strategy is that it was assumed that the discrete approximation is the true probability model. This assumption was made because the mean square convergence of the discrete approximation to the underlying diffusion is, in general, known only for a fixed interval $[0, \mathcal{T}]$, while in the asymptotic theory we must let $\mathcal{T} \to \infty$. In this section we show the mean square convergence of the discrete approximation to the underlying diffusion process for all times $s \in [0, \infty)$ for a class of bond pricing models. This allows us to prove the consistency and asymptotic normality of the estimator for the parameters of the underlying continuous time model.

I consider the following class of bond pricing models. The rate of interest $\tilde{r}(t)$ follows the process

$$d\tilde{r}(t) = \kappa_0(\theta_0 - \tilde{r}(t))dt + \sigma(\tilde{r}(t))d\widetilde{W}(t), \tag{31}$$

where $\widetilde{W}(t)$ is a scalar Brownian Motion and the diffusion coefficient $\sigma(\tilde{r})$ satsifies a growth condition and a uniform Lipschitz condition with Lipschitz constant $\bar{\sigma}$. The price at time t of a bond with maturity τ, $\widetilde{H}(\tilde{r}, \tau, \beta_0)$, satisfies

$$1/2\sigma_0^2\tilde{r}\widetilde{H}_{\tilde{r}\tilde{r}} + (\kappa_0\theta_0 - \kappa_0\tilde{r} - \lambda(\tilde{r}))\widetilde{H}_{\tilde{r}} - \widetilde{H}_\tau - \widetilde{H} = 0 \tag{32}$$

with boundary condition $\widetilde{H}(\tilde{r}(t+\tau), \beta_0) = 1$. Here the function $\lambda(\tilde{r})$ is assumed to satisfy growth and Lipschitz conditions.[14] The price at time t of a bond with maturity τ is given by the expectation

$$\widetilde{H}(\tilde{r}(t), \beta_0) = E^{Q^t}\left(e^{-\int_t^{t+\tau}\tilde{r}(u)du} \mid \widetilde{\mathcal{F}}_t\right),$$

where the probability Q^t is defined as in Section 2.

The first goal of this section is to approximate the simulated state variable process

$$dr(s) = \kappa(\theta r(s) - r(s))ds + \sigma(r(s))dW(s) \tag{33}$$

[14]It is relatively easy to choose state variable processes in the equilibrium model of Cox, Ingersoll, and Ross (1985a) such that these assumptions are satisfied. For example, if one makes all of the assumptions in CIR (1985b) except for the assumption made about the dynamics of the state variable Y, and instead assumes that the state variable follows a process

$$dY(t) = (\zeta Y(t) + \xi)dt + \nu(Y(t))dW(t),$$

where $\nu(Y) = \nu Y$ for $Y \leq \overline{Y}$ and $\nu(Y) = \nu\overline{Y}$ for $Y > \overline{Y}$ then the interest rate is a constant multiple of Y and the function $\lambda(\tilde{r})$ satisfies a growth condition and a uniform Lipschitz condition.

corresponding to (31) in mean square on the interval $[0,\infty)$.[15] That is, I want to show that $E^P[(r^n(s_k)-r(s_k))^2] \leq Kh$ for some constant K as $h \to 0$. I assume $2\kappa > \bar{\sigma}^2$.

5.1 Convergence of the Discrete Approximation to the Diffusion Process

Consider the stochastic difference equation given by the Euler approximation

$$\begin{aligned} r^n(s_0) &= r(0), \\ r^n(s_{k+1}) &= \begin{cases} 0 & \text{if } r^n(s_k)+\Delta r^n(s_k) < 0, \\ r^n(s_k)+\Delta r^n(s_k) & \text{if } 0 \leq r^n(s_k)+\Delta r^n(s_k), \end{cases} \\ \Delta r^n(s_k) &= h(\kappa\theta - \kappa r^n(s_k)) + \sqrt{h}\sigma(r^n(s_k)))\epsilon(s_{k+1}). \end{aligned} \tag{34}$$

It is well known that the local error of this approximation is $O(h^2)$ and that the error over a fixed interval of time $[0,\mathcal{T}]$ is $O(h)$. That is, as $h \to 0$, $E^P[(r^n(s_{k+1})-r(s_k)^2] \leq Kh^2$ if $r^n(s_k) = r(s_k)$ and $E^P[(r^n(s_k)-r(s_k))^2] \leq Kh$ if $kh \leq \mathcal{T}$ (Milshtein (1974), Pardoux and Talay (1985), Gard (1988)). However, in the asymptotic theory one must let $\mathcal{T} \to \infty$, and the existing results are not adequate because the constant K depends upon $\mathcal{T}$. However, for the process (33) above and the approximating process (34) it is relatively easy to show that $E^P[(r^n(s_k)-r(s_k))^2] \leq Kh$ for all $s \in [0,\infty)$, i.e. the bound K is not a function of $\mathcal{T}$.

Let $\delta(s_k) \equiv r^n(s_k) - r(s_k)$ denote the approximation error at the k-th time step. I am interested in the expectation of its square, $E^P\delta(s_k)^2$.

First, note that if $r^n(s_k) = r(s_k)$ then, as $h \to 0$, $E^P(\delta(s_k)^2) \leq Kh^2$ for some constant K. The next step is to bound the approximation error $E^P\delta(s_{k+1})^2$ in terms of $E^P\delta(s_k)^2$. The error at the $(k+1)$-th time-step $\delta(s_k) = r^n(s_{k+1}) - r(s_{k+1})$ is given by the sum of the preceding period's error and the error this period,[16] or

$$\begin{aligned} \delta(s_{k+1}) &= \delta(s_k) + h(\kappa\theta - \kappa r^n(s_k)) + \sqrt{h}\sigma(r^n(s_k))\epsilon(s_{k+1}) \\ &- \int_{s_k}^{s_{k+1}} \kappa(\theta - r(u))du - \int_{s_k}^{s_{k+1}} \sigma(r(u))dW(u). \end{aligned}$$

This can be rewritten

$$\begin{aligned} \delta(s_{k+1}) &= \delta(s_k)(1-\kappa h) + h(\kappa\theta - \kappa r(s_k)) + \sqrt{h}\sigma(r(s_k))\epsilon(s_{k+1}) + \sqrt{h}\zeta(s_k)\epsilon(s_{k+1}) \\ & \int_{s_k}^{s_{k+1}} \kappa(\theta - r(u))du - \int_{s_k}^{s_{k+1}} \sigma(r(u))dW(u), \end{aligned} \tag{35}$$

[15] Mean square convergence of the discrete approximations (23) to the continuous time processes (20) is (subject to certain conditions) given by standard results because the process (20) is defined on a finite interval $[s, s+\tau]$.

[16] This analysis does not consider explicitly the "trimming" of the approximating process at zero in equation (34) above. However, this trimming always reduces the magnitude of the approximation error.

where $\zeta(s_k) = \sigma(r^n(s_k) - \sigma(r(s_k))$ and $|\zeta(s_k)| \le \bar{\sigma}|\delta(s_k)|$. This last inequality follows from the Lipschitz condition satisfied by the function $\sigma(r)$. The four terms

$$h(\kappa\theta - \kappa r(s_k)) + \sqrt{h}\sigma(r(s_k))\epsilon(s_{k+1}) - \int_{s_k}^{s_{k+1}} \kappa(\theta - r(u))du - \int_{s_k}^{s_{k+1}} \sigma(r(u))dW(u)$$

comprise the "local error" at the $(k+1)$-th time-step. Squaring (35) and taking the expectation at $s = 0$,

$$E^P\delta(s_{k+1})^2 \le \left(1 - \left(2\kappa - \bar{\sigma}^2\right)h + \kappa^2h^2\right)\delta(s_k)^2 + 2K_1h^{3/2}\sqrt{E^P\delta(s_k)^2} + K_2h^2. \quad (36)$$

as $h \to 0$. I claim that $E^P\delta(s_k)^2 = O(h)$.

The assumption that $-2\kappa + \bar{\sigma}^2 < 0$ implies that there exists a constant $c \ge 1$ such that

$$-2\kappa + \bar{\sigma}^2 + \frac{\kappa^2}{c} + \frac{2K_1}{\sqrt{c}} + \frac{K_2}{c} < 0.$$

(i) First, suppose $E^P\delta(s_k)^2 \ge ch$ for some time s_k. Then it follows from (36) that

$$E^P\delta(s_{k+1})^2 < E^P(\delta(s_k)^2)\left(1 - \left(2\kappa - \bar{\sigma}^2\right)h + \frac{\kappa^2}{c}h + \frac{2K_1}{\sqrt{c}}h + \frac{K_2}{c}h\right)$$

and $E^P\delta(s_{k+1})^2 < E^P\delta(s_k)^2$.

(ii) Alternatively, suppose that $E^P\delta(s_k)^2 < ch$ for some time s_k. Then, from (36), as $h \to 0$,

$$\begin{aligned}
E^P\delta(s_{k+1})^2 &< E^P\delta(s_k)^2\left(1 - \left(2\kappa - \bar{\sigma}^2\right)h + c\kappa^2h^3 + 2\sqrt{c}K_1h^2 + cK_2h^2\right) \\
&< E^P\delta(s_k)^2 + (c\kappa^2 + 2\sqrt{c}K_1 + cK_2)h \\
&< (c + c\kappa^2 + 2\sqrt{c}K_1 + cK_2)h \\
&< Kh,
\end{aligned}$$

where $K = c + c\kappa^2 + 2\sqrt{c}K_1 + cK_2$.

Parts (i) and (ii) together show that, as $h \to 0$, $E^P\delta(s_{k+1})^2 < (c + c\kappa^2 + 2K_1\sqrt{c} + cK_2)h = O(h)$ regardless of the magnitude of $E^P\delta(s_k)^2$. This proves the following proposition.

Proposition 1 *Let $\delta(s_k) \equiv r^n(s_k) - r(s_k)$ where $r(s_k)$ is the value of the process (33) at time $t = kh$ and $r^n(s_k)$ is the value of the Euler approximation (34). Then, as $h \to 0$, $E^P(\delta(s_k)^2) \le Kh$ for all $s \in [0, \infty)$.*

A similar proposition can be proved for the approximation

$$\begin{aligned}
r^n(t_0) &= r(0), \\
r^n(s_{k+1}) &= \begin{cases} 0 & \text{if } r^n(s_k) + \Delta r^n(s_k) < 0, \\ r^n(s_k) + \Delta r^n(s_k) & \text{if } 0 \le r^n(s_k) + \Delta r^n(s_k), \end{cases} \\
\Delta r^n(s_k) &= r^n(s_k)e^{-\kappa h} + \theta(1 - e^{-\kappa h}) - r^n(s_k) + \sigma(r^n(s_k)))\epsilon(s_{k+1}).
\end{aligned} \quad (37)$$

This approximation agrees with (34) in mean square to $O(h^2)$ when $r^n(s_k) = r(s_k)$. The difference is that with this approximation $E^P(r^n(s_{k+1})|r^n(s_k)$ equals the expectation of the original continuous time process when $r^n(s_k) = r(s_k)$, so that one source of approximation error is eliminated. With this approximation one obtains

$$\begin{aligned}\delta(s_{k+1}) &= \delta(s_k)e^{-\kappa h} + \theta(1-e^{-\kappa h}) + \sigma(r^n(s_k))\epsilon(s_{k+1}) \\ &- \int_{s_k}^{s_{k+1}} \kappa(\theta - r(s))ds - \int_{s_k}^{s_{k+1}} \sigma(r(s))dW(s),\end{aligned}$$

and, similar to (36) above,

$$E^P\delta(s_{k+1})^2 \leq \left(e^{-2\kappa h} + \bar{\sigma}^2 h\right) E^P\delta(s_k)^2 + 2K_1 h^{3/2}\sqrt{E^P\delta(s_k)^2} + K_2 h^2. \tag{38}$$

Observing that $\exists h > 0$ such that $e^{-2\kappa h} + \bar{\sigma}^2 h < 1$ provided $-2\kappa + \bar{\sigma}^2 < 0$ and arguing as before I obtain the following result.

Proposition 2 *Let $\delta(s_k) \equiv r^n(s_k) - r(s_k)$ where now $r^n(s_k)$ denotes the value of the approximating process (38). Then, as $h \to 0$, $E^P(\delta(s_k)^2) \leq Kh$ for all $s \in [0, \infty)$.*

Similar propositions can be proved for multivariate process that satisfy the appropriate Lipschitz continuity conditions and parameter restrictions.

5.2 Approximation of the functions $\hat{f}_s^\beta$

I now turn to showing that the mean square convergence of the discretized state variable processes to the underlying diffusion processes implies the convergence of $\hat{f}_s^{\beta n}$ to $\hat{f}_s^\beta$. Let r_i^s denote the risk-adjusted process corresponding to (20), and let $r_i^{s,n}$ denote the approximating process corresponding to (23). The next result bounds the discretization error $\hat{f}_s^{\beta n} - \hat{f}_s^\beta$.

Proposition 3 *Let $\xi_s^n \equiv \hat{f}_s^{\beta n} - \hat{f}_s^\beta$ denote the discretization error in computing $\hat{f}_s^\beta$. As $h \to 0$, $E^P[(\xi_s^n)^2] \leq Kh$.*

PROOF. It suffices to show that $E^{\mathcal{P}}\left(e^{-\int_s^{s+\tau} r_i^{n,s}(u)du} - e^{-\int_s^{s+\tau} r_i^s(u)du}\right)^2 \leq Kh$ as $h \to 0$.

First, it follows immediately from Proposition 1 and standard results on approximating stochastic differential equations (see, e.g., Gard (1988), Theorem 7.2) that, as $h \to 0$, $E^{\mathcal{P}}(r^{n,s}(s_k) - r^s(s_k))^2 \leq Kh$ for each s_k such that $s \leq s_k \leq s+\tau$. Next,

$$E^{\mathcal{P}}\left[\left(\int_s^{s+\tau}(r_i^{s,n}(u) - r_i^s(u))du\right)^2\right]$$

$$= E^{\mathcal{P}}\left[\left(\int_s^{s+\tau}(r_i^{s,n}(u) - q(u))du + \int_s^{s+\tau}(q(u) - r_i^s(u))du\right)^2\right]$$
$$\leq E^{\mathcal{P}}\left[\left(\int_s^{s+\tau}(r_i^{s,n}(u) - q(u))du\right)^2\right] + E^{\mathcal{P}}\left[\left(\int_s^{s+\tau}(q(u) - r_i^s(u))du\right)^2\right]$$

where $q(u)$ is the simple function given by $q(u)=r_i^s(s_k)$ for $s_k \leq u < s_{k+1}$.

Considering the first term to the right of the inequality, let

$$d^n \equiv \max_{s_k\in\{s,s+1/n,\ldots,s+\tau\}} [\mid r_i^{s,n}(s_k) - q(s_k) \mid].$$

Then

$$\begin{aligned} E^{\mathcal{P}}\left[\left(\int_s^{s+\tau}(r_i^{s,n}(u) - q(u))du\right)^2\right] &\leq E^{\mathcal{P}}\left[\left(\int_s^{s+\tau} d^n\right)^2\right] \\ &\leq E^{\mathcal{P}}\left[(\tau d^n)^2\right] \\ &\leq \tau^2 E^{\mathcal{P}}[(d^n)^2]. \end{aligned}$$

The fact above that $E^{\mathcal{P}}(r^{s,n}(s_k) - r^s(s_k))^2 \leq Kh$ as $h \to 0$ for each s_k such that $s \leq s_k \leq s+\tau$ implies that $E^{\mathcal{P}}[(d^n)^2] \leq Kh$. Therefore, as $h \to 0$, there exists some (different) constant K such that

$$E^P\left[\left(\int_s^{s+\tau}(r_i^{s,n}(u) - q(u))du\right)^2\right] \leq Kh.$$

The mean square convergence to zero of the second term to the right of the inequality is a standard result. We conclude that there exists a constant K such that, as $h \to 0$,

$$E^P\left[\left(\int_s^{s+\tau}(r_i^{s,n}(u) - r_i^s(u))du\right)^2\right] \leq Kh.$$

Finally, define $\eta \equiv \int_s^{s+\tau} r_i^{s,n}(u)du - \int_s^{s+\tau} r_i^s(u)du$. By Taylor's theorem,

$$\begin{aligned} \exp\left[-\int_s^{s+\tau} r_i^{s,n}(u)du\right] &= \exp\left[-\left(\int_s^{s+\tau} r_i^s(u)du + \eta\right)\right] \\ &= \exp\left[-\int_s^{s+\tau} r_i^s(u)du\right] - \exp\left[-\left(\int_s^{s+\tau} r_i^s(u)du + x\right)\right]\eta, \end{aligned}$$

where $x \in [\int_s^{s+\tau} r_i^s(u)du, \int_s^{s+\tau} r_i^s(u)du + \eta]$ if $\eta \geq 0$ and $x \in [\int_s^{s+\tau} r_i^s(u)du + \eta, \int_s^{s+\tau} r_i^s(u)du]$ if $\eta < 0$. Now $\exp[-(\int_s^{s+\tau} r_i^s(u)du + x)]$ is bounded if $\int_s^{s+\tau} r_i^s(u)du + x \geq 0$ (it is), and I conclude that, as $h \to 0$, $E^{\mathcal{P}}\left[\left(e^{-\int_s^{s+\tau} r_i^{n,s}(u)du} - e^{-\int_s^{s+\tau} r_i^s(u)du}\right)^2\right] \leq Kh$ for some constant K.

∎

With this proposition, I have shown that the moment conditions can be constructed using the discrete approximations of the stochastic differential equations.

5.3 Consistency and Asymptotic Normality

Define

$$\tilde{\Sigma} \equiv \sum_{j=-\infty}^{\infty} E^{\tilde{P}}([f_t - E^{\tilde{P}}(f_t)][f_{t-j} - E^{\tilde{P}}(f_{t-j})]')$$

and

$$\hat{\Sigma} \equiv \sum_{j=-\infty}^{\infty} E^{\mathcal{P}}([\hat{f}_s^{\beta_0} - E^{\mathcal{P}}(\hat{f}_s^{\beta_0})][\hat{f}_{s-j}^{\beta_0} - E^{\mathcal{P}}(\hat{f}_{s-j}^{\beta_0})]').$$

Let E denote expectation with respect to the probability on the product space formed from $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{F}}, \tilde{P})$ and $(\Omega \times \left(\times_{s=0}^{T}\left(\times_{i=1}^{M}\Omega_i^s\right)\right), \mathcal{G}, \mathbf{G}, \mathcal{P})$, let $f_s^\beta = E^{\mathcal{P}}(\hat{f}_s^\beta|\mathcal{G}_s)$, define $\psi_s \equiv \hat{f}_s^\beta - f_s^\beta$, and define $\xi_s^n \equiv \hat{f}_s^{\beta n} - \hat{f}_s^\beta$.

Theorem 3 *Suppose $T/\mathcal{T}(T) \to c$ and $T/n^2 \to 0$ as $(T,n) \to \infty$, and suppose $W_T \to W_0 = (\tilde{\Sigma} + c\hat{\Sigma})^{-1}$, in probability. The extended simulated moments estimator converges in probability to the true parameter β_0 as $(T,n) \to \infty$.*

PROOF. Note that

$$\begin{aligned}
&\left|\left(\frac{1}{T}\sum_{t=0}^{T}\tilde{f}_t - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}\hat{f}_s^{\beta n}\right) - (E\tilde{f}_\infty - Ef_\infty^\beta)\right| \\
&\quad\le \left|\frac{1}{T}\sum_{t=0}^{T}\tilde{f}_t - E\tilde{f}_\infty\right| + \left|Ef_\infty^\beta - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}\hat{f}_s^{\beta n}\right| \\
&\quad= \left|\frac{1}{T}\sum_{t=0}^{T}\tilde{f}_t - E\tilde{f}_\infty\right| + \left|Ef_\infty^\beta - \left(\frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}(f_s^\beta + \psi_s + \xi_s^n)\right)\right| \\
&\quad\le \left|\frac{1}{T}\sum_{t=0}^{T}\tilde{f}_t - E\tilde{f}_\infty\right| + \left|Ef_\infty^\beta - \frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}f_s^\beta\right| \\
&\qquad + \left|\frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}\psi_s\right| + \left|\frac{1}{\mathcal{T}(T)}\sum_{s=0}^{\mathcal{T}(T)}\xi_s^n\right|.
\end{aligned}$$

The first two terms on the right-hand side converge to zero in probability as in DS (Theorem 1, pp. 13–14).

The third term on the right-hand side converges to zero a.s. and hence in probability because all moments of ψ_s are bounded and $\{\psi_s, \mathcal{G}_{s+\tau}\}$ forms a martingale difference sequence.

By Proposition 3, each of the terms ξ_s^n converges to zero in mean square, and hence in probability, as $n \to \infty$. Therefore the last term on the right-hand side converges to zero in probability

The rest of the proof follows the proof of Theorem 1 in DS.

∎

One also has

Theorem 4 *Suppose $T/\mathcal{T}(T) \to c$ and $T/n^2 \to 0$ as $(T,n) \to \infty$, and let $\Sigma = \tilde{\Sigma} + c\hat{\Sigma}$. Then $\sqrt{T} m_{nT}^{\mathcal{T}(T)}(\beta_0)$ converges in distribution to a normal random vector with mean zero and covariance matrix Σ.*

PROOF. Note that $T/\mathcal{T}(T) \to c$ and $T/n^2 \to 0$ as $(T,n) \to \infty$ imply $\mathcal{T}(T)/n^2 \to \infty$. From the definition of $m_{nT}^{\mathcal{T}(T)}(\beta)$,

$$\begin{aligned} \sqrt{T} m_{nT}^{\mathcal{T}(T)}(\beta_0) &= \left(\frac{1}{\sqrt{T}} \sum_{t=0}^{T} [\tilde{f}_t - E(\tilde{f}_\infty)]\right) - \frac{\sqrt{T}}{\sqrt{\mathcal{T}(T)}} \left(\frac{1}{\sqrt{\mathcal{T}(T)}} \sum_{s=0}^{\mathcal{T}(T)} [\hat{f}^{\beta_0 n} - E(\hat{f}_\infty^{\beta_0 n})]\right) \\ &+ \sqrt{T}(E(f_\infty^{\beta_0}) - E(f_\infty^{\beta_0 n})) \end{aligned}$$

From Proposition 3 and the assumption that $T/n^2 \to 0$, $\lim_{T\to\infty} \sqrt{T}(E(f_\infty^{\beta_0}) - E(f_\infty^{\beta_0 n})) = 0$ and the limiting distribution of $\sqrt{T} m_{nT}^{\mathcal{T}(T)}(\beta_0)$ is equal to the limiting distribution of

$$\left(\frac{1}{\sqrt{T}} \sum_{t=0}^{T} [\tilde{f}_t - E(\tilde{f}_\infty)]\right) - \frac{\sqrt{T}}{\sqrt{\mathcal{T}(T)}} \left(\frac{1}{\sqrt{\mathcal{T}(T)}} \sum_{s=0}^{\mathcal{T}(T)} [\hat{f}^{\beta_0 n} - E(\hat{f}_\infty^{\beta_0 n})]\right).$$

These two terms are independent. As in DS Theorem 2 (p. 15), the limiting distribution of the first term is $N[0, \tilde{\Sigma}]$. For the second term,

$$\begin{aligned} \frac{1}{\sqrt{\mathcal{T}(T)}} \sum_{s=0}^{\mathcal{T}(T)} [\hat{f}^{\beta_0 n} - E(\hat{f}_\infty^{\beta_0 n})] &= \frac{1}{\sqrt{\mathcal{T}(T)}} \sum_{s=0}^{\mathcal{T}(T)} [\hat{f}^{\beta_0} - E(\hat{f}_\infty^{\beta_0}) + \xi_s^n] \\ &= \frac{1}{\sqrt{\mathcal{T}(T)}} \sum_{s=0}^{\mathcal{T}(T)} [\hat{f}^{\beta_0} - E(\hat{f}_\infty^{\beta_0})] + \frac{1}{\sqrt{\mathcal{T}(T)}} \sum_{s=0}^{\mathcal{T}(T)} \xi_s^n \end{aligned}$$

where $E[(\xi_s^n)^2] \leq K/n$. Since $\mathcal{T}(T)/n^2 \to 0$, for $\mathcal{T}(T)$ sufficiently large

$$E\left(\frac{1}{\sqrt{\mathcal{T}(T)}} \sum_{s=0}^{\mathcal{T}(T)} \xi_s^n\right)^2 \leq K \frac{\sqrt{\mathcal{T}(T)}}{n}$$

and

$$E\left(\frac{1}{\sqrt{\mathcal{T}(T)}} \sum_{s=0}^{\mathcal{T}(T)} \xi_s^n\right) \to 0$$

in mean square and therefore in probability. Hence the limiting distribution of

$$\frac{1}{\sqrt{\mathcal{T}(T)}} \sum_{s=0}^{\mathcal{T}(T)} [\hat{f}^{\beta_0 n} - E(\hat{f}_\infty^{\beta_0 n})]$$

is identical to that of

$$\frac{1}{\sqrt{\mathcal{T}(T)}}\sum_{s=0}^{\mathcal{T}(T)}[\hat{f}^{\beta_0}-E(\hat{f}^{\beta_0}_{\infty})]$$

and

$$\frac{\sqrt{T}}{\sqrt{\mathcal{T}(T)}}\left(\frac{1}{\sqrt{\mathcal{T}(T)}}\sum_{s=0}^{\mathcal{T}(T)}[\hat{f}^{\beta_0}-E(\hat{f}^{\beta_0}_{\infty})]\right)\Rightarrow N[0,c\hat{\Sigma}].$$

Together,

$$\sqrt{T}m_{nT}^{\mathcal{T}(T)}(\beta_0)\rightarrow N[0,\Sigma].$$

∎

Suppose $D_0\equiv E[\partial\hat{f}^{\beta_0}/\partial\beta]$ exists, is finite, and has full rank. Then

Corollary 2 *Under the assumptions of Theorem 4, $\sqrt{T}(b_T-\beta_0)$ converges in distribution to a normal random vector with mean zero and covariance matrix*

$$\Lambda=(D_0'\Sigma^{-1}D_0)^{-1}.$$

PROOF. See the discussion in DS pp. 14–16. ∎

The following proposition allows one to construct an estimate of W_0 using an estimate of $\hat{\Sigma}^n$.

Proposition 4 *Let $(\mathcal{T}(T),n)\rightarrow\infty$. Then $\hat{\Sigma}^n\rightarrow\hat{\Sigma}$ in probability.*

PROOF. It is enough to show that $\lim_{n\rightarrow\infty}E([\hat{f}^{\beta n}_s-E(\hat{f}^{\beta n}_s)][\hat{f}^{\beta n}_s-E(\hat{f}^{\beta n}_s)]')=E([\hat{f}^{\beta}_s-E(\hat{f}^{\beta}_s)][\hat{f}^{\beta}_s-E(\hat{f}^{\beta}_s)]')$. Write $\hat{f}^{\beta n}_s=\hat{f}^{\beta}_s+\xi^n_s$, where $E[(\xi^n_s)^2]\leq Kh$ as $h\rightarrow 0$ by Proposition 3. Letting $\zeta^n\equiv\xi^n_s-E(\xi^n_s)$, $[E(\xi^n_s)]^2\leq Kh$ by Jensen's inequality and $E\zeta^n\zeta^{n\prime}\leq Kh$. Then

$$\begin{aligned}
&\lim_{n\rightarrow\infty}E([\hat{f}^{\beta n}_s-E(\hat{f}^{\beta n}_s)][\hat{f}^{\beta n}_s-E(\hat{f}^{\beta n}_s)]')\\
&=\lim_{n\rightarrow\infty}E([\hat{f}^{\beta}_s-E(\hat{f}^{\beta}_s)+\zeta^n][\hat{f}^{\beta}_s-E(\hat{f}^{\beta}_s)+\zeta^n]')\\
&=\lim_{n\rightarrow\infty}E([\hat{f}^{\beta}_s-E(\hat{f}^{\beta}_s)][\hat{f}^{\beta}_s-E(\hat{f}^{\beta}_s)]')+2E([\hat{f}^{\beta}_s-E(\hat{f}^{\beta}_s)]\zeta^{n\prime})+E(\zeta^n\zeta^{n\prime})\\
&=E([\hat{f}^{\beta}_s-E(\hat{f}^{\beta}_s)][\hat{f}^{\beta}_s-E(\hat{f}^{\beta}_s)]').
\end{aligned}$$

∎

6 Computational Feasibility of the ESME

Although the ESME described above requires only that the econometrician generate a simulated sequence of estimates of the bond prices and therefore does not require that the econometrician evaluate the bond pricing function, it nonetheless requires a non-trivial amount of computation. In this section I use the ESME to estimate the bond pricing model due to CIR (1985b) discussed in Section 2 in order to demonstrate its computational feasibility. For this exercise I interpret the model as a model of the prices of nominal bonds. This particular model was selected because the steady state moments of bond prices can be calculated, allowing one to estimate the model by the GMM and compare these estimates to those obtained from the ESME.[17]

The data set consists of the Treasury bill data taken from the CRSP Government Bond Master File. In particular, for each month between December 1971 and December 1986 I took from the Government Bond Master File the price (and other data) for the (just auctioned) 13 and 26-week bills. The prices used consist of the mean of the bid and asked prices on the last business day of the month. By including the (just auctioned) 13 and 26-week bills, the sample includes the most actively traded discount bonds. It seems reasonable to treat the observed prices of these securities as approximations of the equilibrium prices.

The following four moment conditions are used in estimation.

$$\frac{1}{T}\sum_{t=0}^{T}\tilde{H}(\tilde{r}(t),\tau_1,\beta_0) - \frac{1}{T(T)}\sum_{s=0}^{T(T)}\hat{H}_1(r(s),\tau_1,\beta),$$

$$\frac{1}{T}\sum_{t=0}^{T}\tilde{H}(\tilde{r}(t),\tau_2,\beta_0) - \frac{1}{T(T)}\sum_{s=0}^{T(T)}\hat{H}_2(r(s),\tau_2,\beta)$$

$$\frac{1}{T}\sum_{t=0}^{T}\tilde{H}(\tilde{r}(t),\tau_1,\beta_0)\tilde{H}(\tilde{r}(t),\tau_2,\beta_0) - \frac{1}{T(T)}\sum_{s=0}^{T(T)}\hat{H}_1(r(s),\tau_1,\beta_0)\hat{H}_2(r(s),\tau_2,\beta),$$

$$\frac{1}{T}\sum_{t=0}^{T}\tilde{H}(\tilde{r}(t),\tau_1,\beta_0)\tilde{H}(\tilde{r}(t-1),\tau_1,\beta_0) - \frac{1}{T(T)}\sum_{s=0}^{T(T)}\hat{H}_1(r(s),\tau_1,\beta)\hat{H}_2(r(s-1),\tau_1,\beta)$$

where $\tau_1 = 3$ months and $\tau_2 = 6$ months.

Results for the ESME with several simulation sample sizes and time steps of $n = 1$ and $n = 5$ (per month) along with GMM results are shown in Table 1. Analytic expressions for the moments used in the GMM estimation are derived in the appendix. With four moment conditions, the parameters are exactly identified, and the estimates were computed using the identity matrix as the weighting matrix. The estimates of the standard errors were computed

[17]This particular bond pricing model does not satisfy the assumptions of Section (5). Specifically, the diffusion coefficient of the state variable process does not satisfy a uniform Lipschitz condition. In this section I ignore this issue.

using an estimate of $W_0 = (\tilde{\Sigma}^n + c\,\hat{\Sigma}^n)^{-1}$. The matrix $\tilde{\Sigma}$ was calculated using the Newey-West (1987) procedure with twelve autocovariances, $\hat{\Sigma}^n$ was estimated using twelve autocovariances of the simulated data at the estimates, and $c = T/\mathcal{T}(T)$.

Table 1
Comparison of Estimates
Treasury Bill Monthly Data 1972–1986[18]

	$\mathcal{T}(T)$	κ	σ	θ	λ
ESME	2400	0.90790	0.18958	0.07756	-0.30860
$n = 1$		(0.5330)	(0.0676)	(0.1303)	(0.2715)
	4800	0.52191	0.12939	0.07937	-0.23647
		(0.4356)	(0.0489)	(0.1225)	(0.2713)
	9600	0.40619	0.09804	0.07559	-0.23296
		(0.3971)	(0.0391)	(0.1378)	(0.2533)
	19200	0.33590	0.08943	0.08266	-0.19153
		(0.2593)	(0.0371)	(0.1319)	(0.2671)
ESME	2400	0.93452	0.18838	0.07690	-0.34414
$n = 5$		(0.5669)	(0.0601)	(0.1275)	(0.2733)
	4800	0.52101	0.12839	0.07982	-0.24046
		(0.4581)	(0.0487)	(0.1465)	(0.2743)
	9600	0.39625	0.10369	0.08093	-0.20509
		(0.3593)	(0.0399)	(0.1213)	(0.2657)
	19200	0.36080	0.08645	0.07844	-0.20593
		(0.2788)	(0.0373)	(0.1307)	(0.2571)
GMM	∞	0.32066	0.08396	0.07913	-0.19570
$n = \infty$		(0.1293)	(0.0184)	(0.0675)	(0.1387)

The distinctive feature of these estimates is the relatively large simulation sample size $\mathcal{T}(T)$ needed to obtain estimates of κ and σ reasonably close to those obtained from the GMM. It is possible that the need for such large simulation sample sizes is related to the apparent weak mean reversion of the interest rate process,[19] but establishing this would require an additional study. Regardless, I interpret this limited set of computations as demonstrating the computational feasibility of the estimator. The simulation required to obtain estimates which approximate the GMM estimates can be performed on a personal computer.

[18] Asymptotic standard errors of the estimates are in parenthesis.
[19] The GMM estimate of $\kappa = 0.32066$ implies that the half-life of the interest rate process is 2.16 years.

7 Conclusion

This paper presents an estimated simulated moments estimator (ESME) which extends the SME of DS to cases where an analytic expression for the asset pricing function is not known and demonstrates the computational feasibility of the ESME. The estimator requires that the underlying state variable process be geometrically ergodic. This assumption is satisfied in a number of the bond pricing models that have appeared in the finance literature, and these models and generalizations of them could be estimated using the ESME. A simple example of this appears in Section 6.

The prices of common stocks are usually assumed to follow processes that are not geometrically ergodic, and at first glance it seems that the ESME is not well suited to option pricing models. However, appropriate transformation of the data can make certain stock and option price processes geometrically ergodic, and the domain of the ESME is not as limited as the discussion in terms of a bond pricing model in Sections 2 and 5 and the application in Section 6 might suggest. Moreover, the use of this estimation strategy is not restricted to continuous-time financial models. Asset prices can be written as the conditional expectations of the products of their payoffs and ratios of marginal utilities in a number of models that have appeared in the macroeconomic literature, and those models in which sequences of appropriately transformed consumption and asset prices are geometrically ergodic may be estimated using the ESME.

Consistency of the ESME for the parameters of the underlying continuous-time model requires that the number of steps per unit time in the stochastic difference equations used to approximated the underlying state variable process must grow with the simulation sample size. A limitation of my analysis is that I have shown that the required rate of growth to be reasonable (i.e., polynomial) only for a limited class of bond pricing models. While a similar result could be obtained for certain (appropriately transformed) option pricing models, I have not provided a general set of conditions sufficient to ensure that only polynomial growth is required. The alternate strategy of taking the approximating process as the true probability model seems like a reasonable approach in applications, but does not provide consistent estimates of the parameters of the parameters of the underlying continous time model.

Appendix

A Bond Price Moments

This appendix presents the calculation of the moments $E(P(t,\tau_j))$, $E(P(t,\tau_j)P(t,\tau_k))$, and $E(P(t,\tau_j)P(s,\tau_k))$, for $s > t$, for the one-factor model in CIR (1985b). For this model,

$$P(t,\tau) = A(\tau)e^{-B(\tau)r}, \tag{39}$$

where

$$\begin{aligned} A(\tau) &\equiv \left[\frac{2\gamma e^{(\gamma+\lambda+\kappa)\tau/2}}{(\gamma+\lambda+\kappa)(e^{\gamma\tau}-1)+2\gamma}\right]^{2\kappa\theta/\sigma^2}, \\ B(\tau) &\equiv \frac{2(e^{\gamma\tau}-1)}{(\gamma+\lambda+\kappa)(e^{\gamma\tau}-1)+2\gamma}, \\ \gamma &\equiv ((\kappa+\lambda)^2+2\sigma^2)^{1/2}. \end{aligned}$$

First moment

The steady state distribution of the interest rate is gamma, with density function

$$f(r) = \frac{\omega^\nu}{\Gamma(\nu)} r^{\nu-1} e^{-\omega r}, \tag{40}$$

where

$$\begin{aligned} \omega &\equiv 2\kappa/\sigma^2, \\ \nu &\equiv 2\kappa\theta/\sigma^2. \end{aligned}$$

The first moment is

$$\begin{aligned} E(P(t,\tau)) &= \int_0^\infty A(\tau)e^{-B(\tau)r} f(r) dr \\ &= \int_0^\infty A(\tau)e^{-B(\tau)r} \frac{\omega^\nu}{\Gamma(\nu)} r^{\nu-1} e^{-\omega r} dr \\ &= A(\tau)\left(\frac{\omega}{\omega+B(\tau)}\right)^\nu \int_0^\infty (\omega+B(\tau))^\nu \Gamma(\nu)^{-1} r^{\nu-1} e^{-(\omega+B(\tau))r} dr \\ &= A(\tau)\left(\frac{\omega}{\omega+B(\tau)}\right)^\nu. \end{aligned} \tag{41}$$

The last equality follows from the fact that the integrand is the density function of the gamma distribution.

Second moment

The second moment is

$$
\begin{aligned}
E(P(t,\tau)^2) &= \int_0^\infty A(\tau_j)A(\tau_k)e^{-(B(\tau_j)+B(\tau_k))r}f(r)dr \\
&= \int_0^\infty A(\tau_j)A(\tau_k)e^{-(B(\tau_j)+B(\tau_k))r}\frac{\omega^\nu}{\Gamma(\nu)}r^{\nu-1}e^{-\omega r}dr \\
&= A(\tau_j)A(\tau_k)\left(\frac{\omega}{\omega+B(\tau_j)+B(\tau_k)}\right)^\nu \\
&\qquad \times \int_0^\infty (\omega+B(\tau_j)+B(\tau_k))^\nu \Gamma(\nu)^{-1} r^{\nu-1} e^{-(\omega+B(\tau_j)+B(\tau_k))r}dr \\
&= A(\tau_j)A(\tau_k)\left(\frac{\omega}{\omega+B(\tau_j)+B(\tau_k)}\right)^\nu .
\end{aligned} \tag{42}
$$

Again, the last equality follows from the fact that the integrand is the density function of the gamma distribution.

Lagged second moment

The lagged second moment is

$$
\begin{aligned}
E(P(t,\tau_j)P(s,\tau_k)) &= \int_0^\infty \int_0^\infty P(t,\tau_j)P(s,\tau_k)f(r_t,r_s)dr_s dr_t \\
&= \int_0^\infty \int_0^\infty P(t,\tau_j)P(s,\tau_k)f(r_t)f(r_s|r_t)dr_s dr_t \\
&= \int_0^\infty \left(P(t,\tau_j)f(r_t)\int_0^\infty P(s,\tau_k)f(r_s|r_t)dr_s\right) dr_t,
\end{aligned}
$$

where r_t and r_s denote $r(t)$ and $r(s)$, respectively. The inner integral can be evaluated as follows. The density for r_s conditional on r_t, for $s > t$, is (CIR (1985b))

$$
f(r_s \mid r_t) = ce^{-u-v}\left(\frac{v}{u}\right)^{q/2} I_q(2(uv)^{1/2}), \tag{43}
$$

where

$$
\begin{aligned}
c &\equiv \frac{2\kappa}{\sigma^2(1-e^{-\kappa(s-t)})}, \\
u &\equiv cr_t e^{-\kappa(s-t)}, \\
v &\equiv cr_s, \\
q &\equiv \frac{2\kappa\theta}{\sigma^2} - 1,
\end{aligned}
$$

and $I_q(\cdot)$ denotes the modified Bessel function of the first kind of order q.[20] Then

[20] See Oliver (1965) for the properties of the modified Bessel function.

$$
\begin{aligned}
\int_0^\infty P(s,\tau_k) f(r_s|r_t) dr_s &= \int_0^\infty A(\tau_k) e^{-B(\tau_k) r_s} c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q(2(uv)^{1/2}) dr_s \\
&= \int_0^\infty A(\tau_k) e^{-B(\tau_k) r_s} c e^{-u} e^{-v} \left(\frac{v}{u}\right)^{q/2} \left(\sum_{j=0}^\infty \frac{u^{j+q/2} v^{j+q/2}}{\Gamma(j+1)\Gamma(j+1+q)}\right) dr_s \\
&= \int_0^\infty A(\tau_k) e^{-B(\tau_k) r_s} c e^{-u} e^{-c r_s} \left(\sum_{j=0}^\infty \frac{u^j c^{j+q} r_s^{j+q}}{\Gamma(j+1)\Gamma(j+1+q)}\right) dr_s \\
&= \int_0^\infty A(\tau_k) c e^{-u} \left(\sum_{j=0}^\infty \frac{e^{-(c+B(\tau_k)) r_s} u^j c^{j+q} r_s^{j+q}}{\Gamma(j+1)\Gamma(j+1+q)}\right) dr_s \\
&= \int_0^\infty A(\tau_k) c e^{-u} \\
&\quad \times \left(\sum_{j=0}^\infty \left(\frac{c}{c+B(\tau_k)}\right)^{j+q} \frac{u^j e^{-(c+B(\tau_k)) r_s} ((c+B(\tau_k)) r_s)^{j+q}}{\Gamma(j+1)\Gamma(j+1+q)}\right) dr_s.
\end{aligned}
$$

Exchanging the order of integration and summation,

$$
\begin{aligned}
\int_0^\infty P(s,\tau_k) f(r_s|r_t) dr_s &= A(\tau_k) c \sum_{j=0}^\infty \left(\frac{c}{c+B(\tau_k)}\right)^{j+q} \frac{e^{-u} u^j}{\Gamma(j+1)} \\
&\quad \times \int_0^\infty \frac{e^{-(c+B(\tau_k)) r_s} ((c+B(\tau_k)) r_s)^{j+q}}{\Gamma(j+1+q)} dr_s.
\end{aligned}
$$

Then letting $y = (c + B(\tau_k)) r_s$,

$$
\begin{aligned}
\int_0^\infty P(s,\tau_k) f(r_s|r_t) dr_s &= A(\tau_k) \left(\frac{c}{c+B(\tau_k)}\right) \\
&\quad \times \sum_{j=0}^\infty \left(\frac{c}{c+B(\tau_k)}\right)^{j+q} \frac{e^{-u} u^j}{\Gamma(j+1)} \int_0^\infty \frac{e^{-y} y^{j+q}}{\Gamma(j+1+q)} dy \\
&= A(\tau_k) \left(\frac{c}{c+B(\tau_k)}\right) \sum_{j=0}^\infty \left(\frac{c}{c+B(\tau_k)}\right)^{j+q} \frac{e^{-u} u^j}{\Gamma(j+1)} \\
&= A(\tau_k) \left(\frac{c}{c+B(\tau_k)}\right)^{q+1} e^{-u} \sum_{j=0}^\infty \left(\frac{uc}{c+B(\tau_k)}\right)^j \Big/ \Gamma(j+1) \quad (44) \\
&= A(\tau_k) \left(\frac{c}{c+B(\tau_k)}\right)^{q+1} e^{-u} \exp\left(\frac{uc}{c+B(\tau_k)}\right) \\
&= A(\tau_k) \left(\frac{c}{c+B(\tau_k)}\right)^{q+1} \exp\left(\frac{-uB(\tau_k)}{c+B(\tau_k)}\right) \\
&= A(\tau_k) \left(\frac{c}{c+B(\tau_k)}\right)^{q+1} \exp\left(\frac{-c e^{-\kappa(s-t)} B(\tau_k) r_t}{c+B(\tau_k)}\right) \\
&= k_1(\tau_k) e^{-k_2(\tau_k) r_t},
\end{aligned}
$$

where

$$k_1(\tau_k) \equiv A(\tau_k)\left(\frac{c}{c+B(\tau_k)}\right)^{q+1},$$
$$k_2(\tau_k) \equiv \left(\frac{ce^{-\kappa(s-t)}B(\tau_k)}{c+B(\tau_k)}\right).$$

The second equality follows from the representation of the gamma function

$$\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}dy,$$

and the fourth follows from

$$e^x = \sum_{j=0}^\infty \frac{x^j}{j!}.$$

Using (44), the lagged second moment becomes

$$\begin{aligned}
E(P(t,\tau_j)P(s,\tau_k)) &= \int_0^\infty A(\tau_j)k_1\tau_k)e^{-(B(\tau_j)+k_2(\tau_k))r_t}f(r_t) \\
&= \int_0^\infty A(\tau_j)k_1(\tau_k)e^{-(B(\tau_j)+k_2(\tau_k))r_t}\frac{\omega^\nu}{\Gamma(\nu)}r_t^{\nu-1}e^{-\omega r_t}dr_t \\
&= A(\tau_j)k_1(\tau_k)\left(\frac{\omega}{\omega+B(\tau_j)+k_2(\tau_k)}\right)^\nu \\
&\qquad \times \int_0^\infty (\omega+B(\tau_j)+k_2(\tau_k))^\nu\Gamma(\nu)^{-1}r^{\nu-1}e^{-(\omega+B(\tau_j)+k_2(\tau_k))r_t}dr_t \\
&= A(\tau_j)k_1(\tau_k)\left(\frac{\omega}{\omega+B(\tau_j)+k_2(\tau_k)}\right)^\nu.
\end{aligned}$$

References

Black, F. and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, 637–654.

Bossarts, P., 1989, "Simulation Estimators of Optimal Early Exercise," working paper, Carnegie Mellon University.

Bossarts, P., and P. Hillion, 1989, "Generalized Method of Moments Tests of Contingent Claims Asset Pricing Models," working paper, Carnegie Mellon University.

Bossarts, P., and P. Hillion, 1989, "Method of Simulated Moments Tests of Contingent Claims Asset Pricing Models," working paper, Carnegie Mellon University.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross, 1985a, "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica* 53, 363–384.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross, 1985b, "A Theory of the Term Structure of Interest Rates," *Econometrica* 53, 385–408.

Duffie, D. and K. Singleton, 1989, "Simulated Moments Estimation of Markov Models of Asset Prices," working paper, Stanford University.

Feller, W., 1951, "Two Singular Diffusion Problems," *Annals of Mathematics* 54, 173–182.

Gard, T. C., 1988, *Introduction to Stochastic Differential Equations*, Marcel Dekker Inc., New York.

Gibbons, M. R., and K. Ramaswamy, 1986, "The Term Structure of Interest Rates: Empirical Evidence," working paper, Stanford University.

Gihman, I. I., and A. V. Skorohod, 1972, *Stochastic Differential Equations*, Springer-Verlag, New York.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica* **50**, 1029–1054.

Lee, B.-S., and B. F. Ingram, 1989, "Simulation Estimators of Time Series Models," working paper, University of Iowa, forthcoming in *Journal of Econometrics*.

MacFadden, D., 1989, "Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica* 57, 1989, 995–1026.

Milshtein, G., 1974, "Approximate Integration of Stochastic Differential Equations," *Theory of probability and its applications* 19, 557–562.

Newey, W., and K. West, 1987, "A Simple, Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703–708.

Pakes, A., and D. Pollard, 1989, "Simulation and the Asymptotics of Optimization Estimators," *Econometrica* 57, 1027–1058.

Pardoux, E., and D. Talay, 1985, "Discretization and Simulation of Stochastic Differential Equations," *Acta Applicandae Mathematica* 3, 23–47.

Pearson, N. D., and T-s. Sun, 1989, "A Test of the Cox, Ingersoll, Ross Model of the Term Structure of Interest Rates using the Method of Maximum Likelihood."

White, H., 1984, *Asymptotic Theory for Econometricians*, Academic Press, Inc., New York.