EMPIRICAL DYNAMIC ASSET PRICING

Model Specification and Econometric Assessment

KENNETH J. SINGLETON

Empirical Dynamic Asset Pricing

This page intentionally left blank

Empirical Dynamic Asset Pricing

Model Specification and Econometric Assessment

Kenneth J. Singleton

Princeton University Press
Princeton and Oxford

Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press, 3 Market Place, Woodstock, Oxfordshire OX20 1SY

ISBN-13: 978-0-691-12297-7
ISBN-10: 0-691-12297-0

Library of Congress Control Number: 2005937679

British Library Cataloging-in-Publication Data is available

This book has been composed in New Baskerville by Princeton Editorial Associates, Inc., Scottsdale, Arizona

Printed on acid-free paper. ∞

pup.princeton.edu

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

For my mother, Estelle, and in memory of my father, Harold

Contents

Preface

THIS BOOK EXPLORES the interplay among financial economic theory, the availability of relevant data, and the choice of econometric methodology in the empirical study of dynamic asset pricing models. Given the central roles of all of these ingredients, I have had to compromise on the depth of treatment that could be given to each of them. The end result is a book that presumes readers have had some Ph.D.-level exposure to basic probability theory and econometrics, and to discrete- and continuous-time asset pricing theory.

This book is organized into three blocks of chapters that, to a large extent, can be treated as separate modules. Chapters 1 to 6 of Part I provide an in-depth treatment of the econometric theory that is called upon in our discussions of empirical studies of dynamic asset pricing models. Readers who are more interested in the analysis of pricing models and wish to skip over this material may nevertheless find it useful to read Chapters 1 and 5. The former introduces many of the estimators and associated notation used throughout the book, and the latter introduces affine processes, which are central to much of the literature covered in the last module. The final chapter of Part I, Chapter 7, introduces a variety of parametric descriptive models for asset prices that accommodate stochastic volatility and jumps. Some of the key properties of the implied conditional distributions of these models are discussed, with particular attention given to the second through fourth moments of security returns. This material serves as background for our discussion of the econometric analysis of dynamic asset pricing models.

Part II begins with a more formal introduction to the concept of a "pricing kernel" and relates this concept to both preference-based and no-arbitrage models of asset prices. Chapter 9 examines the linear asset pricing relations—restrictions on the conditional means of returns—derived by restricting agents' preferences or imposing distributional assumptions on the joint distributions of pricing kernels and asset returns. It is in this chapter that we discuss the vast literature on testing for serial correlation in asset returns.

Chapter 10 discusses the econometric analyses of pricing relations based directly on the first-order conditions associated with agents' intertemporal consumption and investment decisions. Chapter 11 examines so-called beta representations of conditional expected excess returns, covering both their economic foundations and the empirical evidence on their goodness-of-fit.

Part III covers the literature on no-arbitrage pricing models. Readers wishing to focus on this material will find Chapter 8 on pricing kernels to be useful background. Chapters 12 and 13 explore the specification and goodness-of-fit of dynamic term structure models for default-free bonds. Defaultable bonds, particularly corporate bonds and credit default swaps, are taken up in Chapter 14. Chapters 15 and 16 cover the empirical literature on equity and fixed-income option pricing models.

Acknowledgments

This book is an outgrowth of many years of teaching advanced econometrics and empirical finance to doctoral students at Carnegie Mellon and Stanford Universities. I am grateful to the students in these courses who have challenged my own thinking about econometric modeling of asset price behavior and thereby have influenced the scope and substance of this book.

My way of approaching the topics addressed here, and indeed my understanding of many of the issues, have been shaped to a large degree by discussions and collaborations with Lars Hansen and Darrell Duffie starting in the 1980s. Their guidance has been invaluable as I have wandered through the maze of dynamic asset pricing models.

More generally, readers will recognize that I draw heavily from published work with several co-authors. Chapters 3 and 4 on the properties of econometric estimators and statistical inference draw from joint work with Lars Hansen. Chapter 6 on simulation-based estimators draws from my joint work with Darrell Duffie on simulated method of moments estimation. Chapter 5 on affine processes draws from joint work with Qiang Dai, Darrell Duffie, Anh Le, and Jun Pan. Chapters 10 and 11 on preference-based pricing models and beta models for asset returns draw upon joint work with Lars Hansen, Scott Richard, and Marty Eichenbaum. Chapters 12 and 13 draw upon joint work with Qiang Dai, Anh Le, and Wei Yang. The discussion of defaultable security pricing in Chapter 14 draws upon joint work with Darrell Duffie, Lasse Pedersen, and Jun Pan. Portions of Chapter 16 are based on joint work with Qiang Dai and Len Umantsev. I am sincerely grateful to these colleagues for the opportunities to have worked with them and, through these collaborations, for their contributions to this effort. They are, of course, absolved of any responsibility for remaining confusion on my part.

I am also grateful to Mikhail Chernov, Michael Johannes, and Stefan Nagel for their helpful comments on drafts of this book, particularly on the chapters covering equity returns and option prices.

Throughout the past 20 years I have benefited from working with many conscientious research assistants. Their contributions are sprinkled throughout my research, and recent assistants have been helpful in preparing material for this book. In addition, I thank Linda Bethel for extensive assistance with the graphs and tables, and with related LaTeX issues that arose during the preparation of the manuscript.

Completing this project would not have been possible without the support of and encouragement from Fumi, Shauna, and Yuuta.

Empirical Dynamic Asset Pricing

This page intentionally left blank

1
Introduction

A DYNAMIC ASSET pricing model is refutable empirically if it restricts the joint distribution of the observable asset prices or returns under study. A wide variety of economic and statistical assumptions have been imposed to arrive at such testable restrictions, depending in part on the objectives and scope of a modeler's analysis. For instance, if the goal is to price a given cash-flow stream based on agents' optimal consumption and investment decisions, then a modeler typically needs a fully articulated specification of agents' preferences, the available production technologies, and the constraints under which agents optimize. On the other hand, if a modeler is concerned with the derivation of prices as discounted cash flows, subject only to the constraint that there be no "arbitrage" opportunities in the economy, then it may be sufficient to specify how the relevant discount factors depend on the underlying risk factors affecting security prices, along with the joint distribution of these factors.

An alternative, typically less ambitious, modeling objective is that of testing the restrictions implied by a particular "equilibrium" condition arising out of an agent's consumption/investment decision. Such tests can often proceed by specifying only portions of an agent's intertemporal portfolio problem and examining the implied restrictions on moments of subsets of variables in the model. With this narrower scope often comes some "robustness" to potential misspecification of components of the overall economy that are not directly of interest.

Yet a third case is one in which we do not have a well-developed theory for the joint distribution of prices and other variables and are simply attempting to learn about features of their joint behavior. This case arises, for example, when one finds evidence against a theory, is not sure about how to formulate a better-fitting, alternative theory, and, hence, is seeking a better understanding of the historical relations among key economic variables as guidance for future model construction.

As a practical matter, differences in model formulation and the decision to focus on a "preference-based" or "arbitrage-free" pricing model may also be influenced by the availability of data. A convenient feature of financial data is that it is sampled frequently, often daily and increasingly intraday as well. On the other hand, macroeconomic time series and other variables that may be viewed as determinants of asset prices may only be reported monthly or quarterly. For the purpose of studying the relation between asset prices and macroeconomic series, it is therefore necessary to formulate models and adopt econometric methods that accommodate these data limitations. In contrast, those attempting to understand the day-to-day movements in asset prices—traders or risk managers at financial institutions, for example—may wish to design models and select econometric methods that can be implemented with daily or intraday financial data alone.

Another important way in which data availability and model specification often interact is in the selection of the decision interval of economic agents. Though available data are sampled at discrete intervals of time—daily, weekly, and so on—it need not be the case that economic agents make their decisions at the same sampling frequency. Yet it is not uncommon for the available data, including their sampling frequency, to dictate a modeler's assumption about the decision interval of the economic agents in the model. Almost exclusively, two cases are considered: *discrete-time* models typically match the sampling and decision intervals—monthly sampled data mean monthly decision intervals, and so on—whereas *continuous-time* models assume that agents make decisions continuously in time and then implications are derived for discretely sampled data. There is often no sound economic justification for either the coincidence of timing in discrete-time models, or the convenience of continuous decision making in continuous-time models. As we will see, analytic tractability is often a driving force behind these timing assumptions.

Both of these considerations (the degree to which a complete economic environment is specified and data limitations), as well as the computational complexity of solving and estimating a model, may affect the choice of estimation strategy and, hence, the econometric properties of the estimator of a dynamic pricing model. When a model provides a full characterization of the joint distribution of its variables, a historical sample is available, and fully exploiting this information in estimation is computationally feasible, then the resulting estimators are "fully efficient" in the sense of exploiting all of the model-implied restrictions on the joint distribution of asset prices. On the other hand, when any one of these conditions is not met, researchers typically resort, by choice or necessity, to making compromises on the degree of model complexity (the richness of the economic environment) or the computational complexity of the estimation strategy (which often means less econometric efficiency in estimation).

With these differences in modelers' objectives, practical constraints on model implementation, and computational considerations in mind, this book: (1) characterizes the nature of the restrictions on the joint distributions of asset returns and other economic variables implied by dynamic asset pricing models (DAPMs); (2) discusses the interplay between model formulation and the choice of econometric estimation strategy and analyzes the large-sample properties of the feasible estimators; and (3) summarizes the existing, and presents some new, empirical evidence on the fit of various DAPMs.

We briefly expand on the interplay between model formulation and econometric analysis to set the stage for the remainder of the book.

1.1. Model Implied Restrictions

Let $\mathcal{P}_s$ denote the set of "payoffs" at date s that are to be priced at date t, for $s > t$, by an economic model (e.g., next period's cum-dividend stock price, cash flows on bonds, and so on),[1] and let $\pi_t : \mathcal{P}_s \to \mathbb{R}$ denote the pricing function, where $\mathbb{R}^n$ denotes the n-dimensional Euclidean space. Most DAPMs maintain the assumption of no arbitrage opportunities on the set of securities being studied: for any $q_{t+1} \in \mathcal{P}_{t+1}$ for which $\Pr\{q_{t+1} \geq 0\} = 1$, $\Pr(\{\pi_t(q_{t+1}) \leq 0\} \cap \{q_{t+1} > 0\}) = 0$.[2] In other words, nonnegative payoffs at $t+1$ that are positive with positive probability have positive prices at date t. A key insight underlying the construction of DAPMs is that the absence of arbitrage opportunities on a set of payoffs $\mathcal{P}_s$ is essentially equivalent to the existence of a special payoff, a *pricing kernel* q_s^*, that is strictly positive ($\Pr\{q_s^* > 0\} = 1$) and represents the pricing function π_t as

$$\pi_t(q_s) = E\left[q_s q_s^* \mid \mathcal{I}_t\right], \tag{1.1}$$

for all $q_s \in \mathcal{P}_s$, where $\mathcal{I}_t$ denotes the information set upon which expectations are conditioned in computing prices.[3]

[1] At this introductory level we remain vague about the precise characteristics of the payoffs investors trade. See Harrison and Kreps (1979), Hansen and Richard (1987), and subsequent chapters herein for formal definitions of payoff spaces.

[2] We let $\Pr\{\cdot\}$ denote the probability of the event in brackets.

[3] The existence of a pricing kernel q^* that prices all payoffs according to (1.1) is equivalent to the assumption of no arbitrage opportunities when uncertainty is generated by discrete random variables (see, e.g., Duffie, 2001). More generally, when $\mathcal{I}_t$ is generated by continuous random variables, additional structure must be imposed on the payoff space and pricing function π_t for this equivalence (e.g., Harrison and Kreps, 1979, and Hansen and Richard, 1987). For now, we focus on the pricing relation (1.1), treating it as being equivalent to the absence of arbitrage. A more formal development of pricing kernels and the properties of q^* is taken up in Chapter 8 using the framework set forth in Hansen and Richard (1987).

This result by itself does not imply testable restrictions on the prices of payoffs in $\mathcal{P}_{t+1}$, since the theorem does not lead directly to an empirically observable counterpart to the benchmark payoff. Rather, overidentifying restrictions are obtained by restricting the functional form of the pricing kernel q_s^* or the joint distribution of the elements of the *pricing environment* $(\mathcal{P}_s, q_s^*, \mathcal{I}_t)$. It is natural, therefore, to classify DAPMs according to the types of restrictions they impose on the distributions of the elements of $(\mathcal{P}_s, q_s^*, \mathcal{I}_t)$. We organize our discussions of models and the associated estimation strategies under four headings: preference-based DAPMs, arbitrage-free pricing models, "beta" representations of excess portfolio returns, and linear asset pricing relations. This classification of DAPMs is not mutually exclusive. Therefore, the organization of our subsequent discussions of specific models is also influenced in part by the choice of econometric methods typically used to study these models.

1.1.1. Preference-Based DAPMs

The approach to pricing that is most closely linked to an investor's portfolio problem is that of the preference-based models that directly parameterize an agent's intertemporal consumption and investment decision problem. Specifically, suppose that the economy being studied is comprised of a finite number of infinitely lived agents who have identical endowments, information, and preferences in an uncertain environment. Moreover, suppose that $\mathcal{A}_t$ represents the agents' information set and that the representative consumer ranks consumption sequences using a von Neumann-Morgenstern utility functional

$$E\left[\sum_{t=0}^{\infty}\beta^t U(c_t)\,\middle|\,\mathcal{A}_0\right]. \tag{1.2}$$

In (1.2), preferences are assumed to be time separable with period utility function U and the subjective discount factor $\beta \in (0, 1)$. If the representative agent can trade the assets with payoffs $\mathcal{P}_s$ and their asset holdings are interior to the set of admissible portfolios, the prices of these payoffs in equilibrium are given by (Rubinstein, 1976; Lucas, 1978; Breeden, 1979)

$$\pi_t(q_s) = E\left[m_s^{s-t} q_s \mid A_t\right], \tag{1.3}$$

where $m_s^{s-t} = \beta U'(c_s)/U'(c_t)$ is the intertemporal marginal rate of substitution of consumption (MRS) between dates t and s. For a given parameterization of the utility function $U(c_t)$, a preference-based DAPM allows the association of the pricing kernel q_s^* with m_s^{s-t}.

To compute the prices $\pi_t(q_s)$ requires a parametric assumption about the agent's utility function $U(c_t)$ and sufficient economic structure to determine the joint, conditional distribution of m_s^{s-t} and q_s. Given that prices are set as part of the determination of an equilibrium in goods and securities markets, a modeler interested in pricing must specify a variety of features of an economy outside of securities markets in order to undertake preference-based pricing. Furthermore, limitations on available data may be such that some of the theoretical constructs appearing in utility functions or budget constraints do not have readily available, observable counterparts. Indeed, data on individual consumption levels are not generally available, and aggregate consumption data are available only for certain categories of goods and, at best, only at a monthly sampling frequency.

For these reasons, studies of preference-based models have often focused on the more modest goal of attempting to evaluate whether, for a particular choice of utility function $U(c_t)$, (1.3) does in fact "price" the payoffs in $\mathcal{P}_s$. Given observations on a candidate m_s^{s-t} and data on asset returns $\mathcal{R}_s \equiv \{q_s \in \mathcal{P}_s : \pi_t(q_s) = 1\}$, (1.3) implies testable restrictions on the joint distribution of $\mathcal{R}_s$, m_t^{s-t}, and elements of $\mathcal{A}_t$. Namely, for each s-period return r_s, $E[m_s^{s-t} r_s - 1 | \mathcal{A}_t] = 0$, for any $r_s \in \mathcal{R}_s$ (see, e.g., Hansen and Singleton, 1982). An immediate implication of this moment restriction is that $E[(m_s^{s-t} r_s - 1) x_t] = 0$, for any $x_t \in A_t$.[4] These unconditional moment restrictions can be used to construct method-of-moments estimators of the parameters governing m_s^{s-t} and to test whether or not m_s^{s-t} prices the securities with payoffs in $\mathcal{P}_s$. This illustrates the use of restrictions on the moments of certain functions of the observed data for estimation and inference, when complete knowledge of the joint distribution of these variables is not available.

An important feature of preference-based models of frictionless markets is that, assuming agents optimize and rationally use their available information $\mathcal{A}_t$ in computing the expectation (1.3), there will be *no arbitrage opportunities* in equilibrium. That is, the absence of arbitrage opportunities is a consequence of the equilibrium price-setting process.

1.1.2. Arbitrage-Free Pricing Models

An alternative approach to pricing starts with the presumption of no arbitrage opportunities (i.e., this is not derived from equilibrium behavior). Using the principle of "no arbitrage" to develop pricing relations dates back at least to the key insights of Black and Scholes (1973), Merton (1973), Ross

[4] This is an implication of the "law of iterated expectations," which states that $E[y_s] = E[E(y_s | \mathcal{A}_t)]$, for any conditioning information set $\mathcal{A}_t$.

(1978), and Harrison and Kreps (1979). Central to this approach is the observation that, under weak regularity conditions, pricing can proceed "as if" agents are risk neutral. When time is measured continuously and agents can trade a default-free bond that matures an "instant" in the future and pays the (continuously compounded) rate of return r_t, discounting for risk-neutral pricing is done by the default-free "roll-over" return $e^{-\int_t^s r_u\,du}$. For example, if uncertainty about future prices and yields is generated by a continuous-time Markov process Y_t (so, in particular, the conditioning information set $\mathcal{I}_t$ is generated by Y_t), then the price of the payoff q_s is given equivalently by

$$\pi_t(q_s) = E\left[q_s^* q_s \mid Y_t\right] = E^{\mathbb{Q}}\left[e^{-\int_t^s r_u\,du}\, q_s \mid Y_t\right], \tag{1.4}$$

where $E_t^{\mathbb{Q}}$ denotes expectation with regard to the "risk-neutral" conditional distribution of Y. The term risk-neutral is applied because prices in (1.4) are expressed as the expected value of the payoff q_s *as if agents are neutral toward financial risks.*

As we will see more formally in subsequent chapters, the risk attitudes of investors are implicit in the exogenous specification of the pricing kernel q^* as a function of the state Y_t and, hence, in the change of probability measure underlying the risk-neutral representation (1.4). Leaving preferences and technology in the "background" and proceeding to parameterize the distribution of q^* directly facilitates the computation of security prices. The parameterization of $(\mathcal{P}_s, q_s^*, Y_t)$ is chosen so that the expectation in (1.4) can be solved, either analytically or through tractable numerical methods, for $\pi_t(q_s)$ as a function of $Y_t : \pi_t(q_s) = P(Y_t)$. This is facilitated by the adoption of continuous time (continuous trading), special structure on the conditional distribution of Y, and constraints on the dependence of q^* on Y so that the second expectation in (1.4) is easily computed. However, similarly tractable models are increasingly being developed for economies specified in discrete time and with discrete decision/trading intervals.

Importantly, though knowledge of the risk-neutral distribution of Y_t is sufficient for pricing through (1.4), this knowledge is typically not sufficient for econometric estimation. For the purpose of estimation using historical price or return information associated with the payoffs $\mathcal{P}_s$, we also need information about the distribution of Y under its *data-generating* or *actual* measure. What lie between the actual and risk-neutral distributions of Y are adjustments for the "market prices of risk"—terms that capture agents' attitudes toward risk. It follows that, throughout this book, when discussing arbitrage-free pricing models, we typically find it necessary to specify the distributions of the state variables or risk factors under both measures.

If the conditional distribution of Y_t given Y_{t-1} is known (i.e., derivable from knowledge of the continuous-time specification of Y), then so typically is the conditional distribution of the observed market prices $\pi_t(q_s)$. The completeness of the specification of the pricing relations (both the distribution of Y and the functional form of P_s) in this case implies that one can in principle use "fully efficient" maximum likelihood methods to estimate the unknown parameters of interest, say θ_0. Moreover, this is feasible using market price data alone, even though the risk factors Y may be latent (unobserved) variables. This is a major strength of this modeling approach since, in terms of data requirements, one is constrained only by the availability of financial market data.

Key to this strategy for pricing is the presumption that the burden of computing $\pi_t(q_s) = P_s(Y_t)$ is low. For many specifications of the distribution of the state Y_t, the pricing relation $P_s(Y_t)$ must be determined by numerical methods. In this case, the computational burden of solving for P_s while simultaneously estimating θ_0 can be formidable, especially as the dimension of Y gets large. Have these considerations steered modelers to simpler data-generating processes (DGPs) for Y_t than they might otherwise have studied? Surely the answer is yes and one might reasonably be concerned that such compromises in the interest of computational tractability have introduced model misspecification.

We will see that, fortunately, in many cases there are alternative estimation strategies for studying arbitrage-free pricing relations that lessen the need for such compromises. In particular, one can often compute the moments of prices or returns implied by a pricing model, even though the model-implied likelihood function is unknown. In such cases, method-of-moments estimation is feasible. Early implementations of method-of-moments estimators typically sacrificed some econometric efficiency compared to the maximum likelihood estimator in order to achieve substantial computational simplification. More recently, however, various approximate maximum likelihood estimators have been developed that involve little or no loss in econometric efficiency, while preserving computational tractability.

1.1.3. Beta Representations of Excess Returns

One of the most celebrated and widely applied asset pricing models is the static capital-asset pricing model (CAPM), which expresses expected excess returns in terms of a security's beta with a benchmark portfolio (Sharpe, 1964; Mossin, 1968). The traditional CAPM is static in the sense that agents are assumed to solve one-period optimization problems instead of multi-period utility maximization problems. Additionally, the CAPM beta pricing relation holds only under special assumptions about either the distributions of asset returns or agents' preferences.

Nevertheless, the key insights of the CAPM carry over to richer stochastic environments in which agents optimize over multiple periods. There is an analogous "single-beta" representation of expected returns based on the representation (1.1) of prices in terms of a pricing kernel q^*, what we refer to as an *intertemporal* CAPM or ICAPM.[5] Specifically, setting $s = t + 1$, the benchmark return $r^*_{t+1} = q^*_{t+1}/\pi_t(q^*_{t+1})$ satisfies[6]

$$E\left[r^*_{t+1}(r_{t+1} - r^*_{t+1}) \mid \mathcal{I}_t\right] = 0, \; r_{t+1} \in \mathcal{R}_{t+1}. \tag{1.5}$$

Equation (1.5) has several important implications for the role of r^*_{t+1} in asset return relations, one of which is that r^*_{t+1} is a benchmark return for a single-beta representation of excess returns (see Chapter 11):

$$E\left[r_{j,t+1} \mid \mathcal{I}_t\right] - r^f_t = \beta_{jt}\left(E\left[r^*_{t+1} \mid \mathcal{I}_t\right] - r^f_t\right), \tag{1.6}$$

where

$$\beta_{jt} = \frac{\text{Cov}\left[r_{j,t+1}, r^*_{t+1} \mid \mathcal{I}_t\right]}{\text{Var}\left[r^*_{t+1} \mid \mathcal{I}_t\right]}, \tag{1.7}$$

and r^f_t is the interest rate on one-period riskless loans issued at date t. In words, the excess return on a security is proportional to the excess return on the benchmark portfolio, $E[r^*_{t+1} - r^f_t \mid \mathcal{I}_t]$, with factor of proportionality β_{jt}, for all securities j with returns in $\mathcal{R}_{t+1}$.

It turns out that the beta representation (1.6), together with the representation of r^f in terms of q^*_{t+1},[7] constitute exactly the same information as the basic pricing relation (1.1). Given one, we can derive the other, and vice versa. At first glance, this may seem surprising given that econometric tests of beta representations of asset returns are often not linked to pricing kernels. The reason for this is that most econometric tests of expressions like (1.6) are in fact *not* tests of the joint restriction that $r^f_t = 1/E[q^*_{t+1}|\mathcal{I}_t]$ and r^*_{t+1} satisfies (1.6). Rather tests of the ICAPM are tests of whether a proposed candidate benchmark return r^β_{t+1} satisfies (1.6) alone, for a given information set $\mathcal{I}_t$. There are an infinite number of returns r^β_t that satisfy (1.6) (see Chapter 11). The return r^*_{t+1}, on the other hand, is the unique

[5] By defining a benchmark return that is explicitly linked to the marginal rate of substitution, Breeden (1979) has shown how to obtain a single-beta representation of security returns that holds in continuous time. The following discussion is based on the analysis in Hansen and Richard (1987).

[6] Hansen and Richard (1987) show that when the pricing function π_t is nontrivial, $\Pr\{\pi_t(q^*_{t+1}) = 0\} = 0$, so that r^*_{t+1} is a well-defined return. Substituting r^* into (1.1) gives $E[r^*_{t+1}r_{t+1} \mid \mathcal{I}_t] = \{E[q^{*2}_{t+1} \mid \mathcal{I}_t]\}^{-1}$, for all $r_{t+1} \in \mathcal{R}_{t+1}$. Since r^*_{t+1} is one such return, (1.5) follows.

[7] The interest rate r^f_t can be expressed as $1/E[q^*_{t+1}|\mathcal{I}_t]$ by substituting the payoff $q_{t+1} = 1$ into (1.1) with $s = t + 1$.

return (within a set that is formally defined) satisfying (1.5). Thus, tests of single-beta ICAPMs are in fact tests of weaker restrictions on return distributions than tests of the pricing relation (1.1).

Focusing on a candidate benchmark return r^{β}_{t+1} and relation (1.6) (with r^{β}_{t+1} in place of r^{*}_{t+1}), once again the choices made regarding estimation and testing strategies typically involve trade-offs between the assumptions about return distributions and the robustness of the empirical analysis. Taken by itself, (1.6) is a restriction on the conditional first and second moments of returns. If one specifies a parametric family for the joint conditional distribution of the returns $r_{j,t+1}$ and r^{β}_{t+1} and the state Y_t, then estimation can proceed imposing the restriction (1.6). However, such tests may be compromised by misspecification of the higher moments of returns, even if the first two moments are correctly specified. There are alternative estimation strategies that exploit less information about the conditional distribution of returns and, in particular, that are based on the first two conditional moments for a given information set $\mathcal{I}_t$, of returns.

1.1.4. Linear Pricing Relations

Historically, much of the econometric analysis of DAPMs has focused on linear pricing relations. One important example of a linear DAPM is the version of the ICAPM obtained by assuming that β_{jt} in (1.6) is constant (not state dependent), say β_j. Under this additional assumption, β_j is the familiar "beta" of the jth common stock from the *CAPM*, extended to allow both expected returns on stocks and the riskless interest rate to change over time. The mean of

$$u_{j,t+1} \equiv \left(r_{j,t+1} - r^{f}_{t}\right) - \beta_j \left(r^{\beta}_{t+1} - r^{f}_{t}\right) \tag{1.8}$$

conditioned on $\mathcal{I}_t$ is zero for all admissible r_j. Therefore, the expression in (1.8) is uncorrelated with any variable in the information set $\mathcal{I}_t$; $E[u_{j,t+1}x_t] = 0$, $x_t \in \mathcal{I}_t$. Estimators of the β_j and tests of (1.6) can be constructed based on these moment restrictions.

This example illustrates how additional assumptions about one feature of a model can make an analysis more robust to misspecification of other features. In this case, the assumption that β_j is constant permits estimation of β_j and testing of the null hypothesis (1.6) without having to fully specify the information set $\mathcal{I}_t$ or the functional form of the conditional means of $r_{j,t+1}$ and r^{β}_{t+1}. All that is necessary is that the candidate elements x_t of $\mathcal{I}_t$ used to construct moment restrictions are indeed in $\mathcal{I}_t$.[8]

[8] We will see that this simplification does not obtain when the β_{jt} are state dependent. Indeed, in the latter case, we might not even have readily identifiable benchmark returns r^{β}_{t+1}.

Another widely studied linear pricing relation was derived under the presumption that in a well-functioning—some say *informationally efficient*—market, holding-period returns on assets must be unpredictable (see, e.g., Fama, 1970). It is now well understood that, in fact, the optimal processing of information by market participants is not sufficient to ensure unpredictable returns. Rather, we should expect returns to evidence some predictability, either because agents are risk averse or as a result of the presence of a wide variety of market frictions.

Absent market frictions, then, one sufficient condition for returns to be unpredictable is that agents are risk neutral in the sense of having linear utility functions, $U(c_t) = u_0 + u_c c_t$. Then the MRS is $m_s^{s-t} = \beta^s$, where β is the subjective discount factor, and it follows immediately from (1.3) that

$$E[r_s|I_t] = 1/\beta^s, \tag{1.9}$$

for an admissible return r_s. This, in turn, implies that r_s is unpredictable in the sense of having a constant conditional mean. The restrictions on returns implied by (1.9) are, in principle, easily tested under only minimal additional auxiliary assumptions about the distributions of returns. One simply checks to see whether $r_s - 1/\beta^s$ is uncorrelated with variables dated t or earlier that might be useful for forecasting future returns. However, as we discuss in depth in Chapter 9, there is an enormous literature examining this hypothesis. In spite of the simplicity of the restriction (1.9), whether or not it is true in financial markets remains an often debated question.

1.2. Econometric Estimation Strategies

While the specification of a DAPM logically precedes the selection of an estimation strategy for an empirical analysis, we begin Part I with an overview of econometric methods for analyzing DAPMs. Applications of these methods are then taken up in the context of the discussions of specific DAPMs. To set the stage for Part I, we start by viewing the model construction stage as leading to a family of models or pricing relations describing features of the distribution of an observed vector of variables z_t. This vector may include asset prices or returns, possibly other economic variables, as well as lagged values of these variables. Each model is indexed by a K-dimensional vector of parameters θ in an admissible parameter space $\Phi \in \mathbb{R}^K$. We introduce Φ

For instance, if $\mathcal{I}_t$ is taken to be agents' information set $\mathcal{A}_t$, then the contents of $\mathcal{I}_t$ may not be known to the econometrician. In this case the set of returns that satisfy (1.6) may also be unknown. It is of interest to ask then whether or not there are similar risk-return relations with moments conditioned on an observable subset of $\mathcal{A}_t$, say $\mathcal{I}_t$, for which benchmark returns satisfying an analogue to (1.6) are observable. This is among the questions addressed in Chapter 11.

because, for each of the DAPMs indexed by θ to be well defined, it may be necessary to constrain certain parameters to be larger than some minimum value (e.g., variances or risk aversion parameters), or DAPMs may imply that certain parameters are functionally related. The basic premise of an econometric analysis of a DAPM is that there is a unique $\theta_0 \in \Phi$ (a unique pricing relation) consistent with the population distribution of z. A primary objective of the econometric analysis is to construct an estimator of θ_0.

More precisely, we view the selection of an *estimation strategy* for θ_0 as the choice of:

- A sample of size T on a vector z_t of observed variables, $\vec{z}_T \equiv (z_T, z_{T-1}, \ldots, z_1)'$.
- An *admissible* parameter space $\Phi \subseteq \mathbb{R}^K$ that includes θ_0.
- A K-vector of functions $\mathcal{D}(z_t; \theta)$ with the property that θ_0 is the unique element of Φ satisfying

$$E[\mathcal{D}(z_t; \theta_0)] = 0. \tag{1.10}$$

What ties an estimation strategy to the particular DAPM of interest is the requirement that θ_0 be the unique element of Φ that satisfies (1.10) for the chosen function $\mathcal{D}$. Thus, we view (1.10) as summarizing the implications of the DAPM that are being used directly in estimation. Note that, while the estimation strategy is premised on the economic theory of interest implying that (1.10) is satisfied, there is no presumption that this theory implies a unique $\mathcal{D}$ that has mean zero at θ_0. In fact, usually, there is an uncountable infinity of admissible choices of $\mathcal{D}$.

For many of the estimation strategies considered, $\mathcal{D}$ can be reinterpreted as the first-order condition for maximizing a nonstochastic *population estimation objective* or *criterion* function $Q_0(\theta) : \Phi \to \mathbb{R}$. That is, at θ_0,

$$\frac{\partial Q_0}{\partial \theta}(\theta_0) = E[\mathcal{D}(z_t; \theta_0)] = 0. \tag{1.11}$$

Thus, we often view a choice of estimation strategy as a choice of criterion function Q_0. For well-behaved Q_0, there is always a θ^* that is the global maximum (or minimum, depending on the estimation strategy) of the criterion function Q_0. Therefore, for Q_0 to be a sensible choice for the model at hand we require that θ^* be unique and equal to the population parameter vector of interest, θ_0. A necessary step in verifying that $\theta^* = \theta_0$ is verifying that $\mathcal{D}$ satisfies (1.10) at θ_0.

So far we have focused on constraints on the population moments of z derived from a DAPM. To construct an estimator of θ_0, we work with the sample counterpart of $Q_0(\theta)$, $Q_T(\theta)$, which is a known function of $\vec{z}_T$. (The subscript T is henceforth used to indicate dependence on the entire sample.)

The sample-dependent θ_T that minimizes $Q_T(\theta)$ over Φ is the *extremum* estimator of θ_0. When the first-order condition to the population optimum problem takes the form (1.11), the corresponding first-order condition for the sample estimation problem is[9]

$$\frac{\partial Q_T}{\partial \theta}(\theta_T) = \frac{1}{T}\sum_{t=1}^{T} \mathcal{D}(z_t; \theta_T) = 0. \tag{1.12}$$

The sample relation (1.12) is obtained by replacing the population moment in (1.11) by its sample counterpart and choosing θ_T to satisfy these sample moment equations. Since, under regularity, sample means converge to their population counterparts [in particular, $Q_T(\cdot)$ converges to $Q_0(\cdot)$], we expect θ_T to converge to θ_0 (the parameter vector of interest and the unique minimizer of Q_0) as $T \to \infty$.

As noted previously, DAPMs often give rise to moment restrictions of the form (1.10) for more than one $\mathcal{D}$, in which case there are multiple feasible estimation strategies. Under regularity, all of these choices of $\mathcal{D}$ have the property that the associated θ_T converge to θ_0 (they are *consistent* estimators of θ_0). Where they differ is in the variance-covariance matrices of the implied large-sample distributions of θ_T. One paradigm, then, for selecting among the feasible estimation strategies is to choose the $\mathcal{D}$ that gives the most econometrically efficient estimator in the sense of having the smallest asymptotic variance matrix. Intuitively, the later estimator is the one that exploits the most information about the distribution of $\vec{z}_T$ in estimating θ_0.

Once a DAPM has been selected for study and an estimation strategy has been chosen, one is ready to proceed with an empirical study. At this stage, the econometrician/modeler is faced with several new challenges, including:

1. The choice of computational method to find a global optimum to $Q_T(\theta)$.
2. The choice of statistics and derivation of their large-sample properties for testing hypotheses of interest.
3. An assessment of the actual small-sample distributions of the test statistics and, thus, of the reliability of the chosen inference procedures.

The computational demands of maximizing Q_T can be formidable. When the methods used by a particular empirical study are known, we occasionally comment on the approach taken. However, an in-depth exploration of

[9] In subsequent chapters we often find it convenient to define Q_T more generally as $1/T\sum_{t=1}^{T} \mathcal{D}_T(z_t; \theta_T) = 0$, where $\mathcal{D}_T(z_t; \theta)$ is chosen so that it converges (almost surely) to $\mathcal{D}(z_t; \theta)$, as $T \to \infty$, for all $\theta \in \Phi$.

alternative algorithms for finding the optimum of Q_T is beyond the scope of this book.

With regard to points (2) and (3), there are many approaches to testing hypotheses about the goodness-of-fit of a DAPM or the values of the parameters θ_0. The criteria for selecting a test procedure (within the classical statistical paradigm) are virtually all based on large-sample considerations. In practice, however, the actual distributions of estimators in finite samples may be quite different than their large-sample counterparts. To a limited degree, Monte Carlo methods have been used to assess the small-sample properties of estimators θ_T. We often draw upon this literature, when available, in discussing the empirical evidence.

Part I

Econometric Methods for Analyzing DAPMs

This page intentionally left blank

2

Model Specification and Estimation Strategies

A DAPM MAY: (1) provide a complete characterization of the joint distribution of all of the variables being studied; or (2) imply restrictions on some moments of these variables, but not reveal the form of their joint distribution. A third possibility is that there is not a well-developed theory for the joint distribution of the variables being studied. Which of these cases obtains for the particular DAPM being studied determines the feasible estimation strategies; that is, the feasible choices of $\mathcal{D}$ in the definition of an *estimation strategy*. This chapter introduces the maximum likelihood (ML), generalized method of moments (GMM), and linear least-squares projection (LLP) estimators and begins our development of the interplay between model formulation and the choice of an estimation strategy discussed in Chapter 1.

2.1. Full Information about Distributions

Suppose that a DAPM yields a complete characterization of the joint distribution of a sample of size T on a vector of variables y_t, $\vec{y}_T \equiv \{y_1, \ldots, y_T\}$. Let $L_T(\beta) = L(\vec{y}_T; \beta)$ denote the family of joint density functions of $\vec{y}_T$ implied by the DAPM and indexed by the K-dimensional parameter vector β. Suppose further that the admissible parameter space associated with this DAPM is $\Theta \subseteq \mathbb{R}^K$ and that there is a unique $\beta_0 \in \Theta$ that describes the true probability model generating the asset price data.

In this case, we can take $L_T(\beta)$ to be our sample criterion function—called the *likelihood function* of the data—and obtain the *maximum likelihood* (ML) estimator b_T^{ML} by maximizing $L_T(\beta)$. In ML estimation, we start with the joint density function of $\vec{y}_T$, evaluate the random variable $\vec{y}_T$ at the realization comprising the observed historical sample, and then maximize the value of this density over the choice of $\beta \in \Theta$. This amounts to maximizing, over all admissible β, the "likelihood" that the realized sample was drawn from the density $L_T(\beta)$. ML estimation, when feasible, is the most econometrically efficient estimator within a large class of consistent estimators (Chapter 3).

In practice, it turns out that studying L_T is less convenient than working with a closely related objective function based on the conditional density function of y_t. Many of the DAPMs that we examine in later chapters, for which ML estimation is feasible, lead directly to knowledge of the density function of y_t conditioned on $\vec{y}_{t-1}$, $f_t(y_t|\vec{y}_{t-1};\beta)$ and imply that

$$f_t(y_t|\vec{y}_{t-1};\beta) = f\left(y_t\middle|\vec{y}^{\,J}_{t-1};\beta\right), \tag{2.1}$$

where $\vec{y}^{\,J}_t \equiv (y_t, y_{t-1}, \ldots, y_{t-J+1})$, a J-history of y_t. The right-hand side of (2.1) is *not* indexed by t, implying that the conditional density function does not change with time.[1] In such cases, the likelihood function L_T becomes

$$L_T(\beta) = \prod_{t=J+1}^{T} f\left(y_t\middle|\vec{y}^{\,J}_{t-1};\beta\right) \times f_m(\vec{y}_J;\beta), \tag{2.2}$$

where $f_m(\vec{y}_J)$ is the marginal, joint density function of $\vec{y}_J$. Taking logarithms gives the *log-likelihood* function $l_T \equiv T^{-1}\log L_T$,

$$l_T(\beta) = \frac{1}{T}\sum_{t=J+1}^{T}\log f\left(y_t\middle|\vec{y}^{\,J}_{t-1};\beta\right) + \frac{1}{T}\log f_m(\vec{y}_J;\beta). \tag{2.3}$$

Since the logarithm is a monotonic transformation, maximizing l_T gives the same ML estimator b_T^{ML} as maximizing L_T.

The first-order conditions for the sample criterion function (2.3) are

$$\frac{\partial l_T}{\partial\beta}\left(b_T^{\mathrm{ML}}\right) = \frac{1}{T}\sum_{t=J+1}^{T}\frac{\partial\log f}{\partial\beta}\left(y_t|\vec{y}^{\,J}_{t-1};b_T^{\mathrm{ML}}\right) + \frac{1}{T}\frac{\partial\log f_m}{\partial\beta}\left(\vec{y}_J;b_T^{\mathrm{ML}}\right) = 0, \tag{2.4}$$

where it is presumed that, among all estimators satisfying (2.4), b_T^{ML} is the one that maximizes l_T.[2] Choosing $z_t' = (y_t', \vec{y}^{\,J\prime}_{t-1})$ and

[1] A sufficient condition for this to be true is that the time series $\{y_t\}$ is a strictly stationary process. Stationarity does not preclude time-varying conditional densities, but rather just that the functional form of these densities does not change over time.

[2] It turns out that b_T^{ML} need not be unique for fixed T, even though β_0 is the unique minimizer of the population objective function Q_0. However, this technical complication need not concern us in this introductory discussion.

$$\mathcal{D}(z_t;\beta) \equiv \frac{\partial \log f}{\partial \beta}\left(y_t|\vec{y}^{\,J}_{t-1};\beta\right) \tag{2.5}$$

as the function defining the moment conditions to be used in estimation, it is seen that (2.4) gives first-order conditions of the form (1.12), except for the last term in (2.4).[3] For the purposes of large-sample arguments developed more formally in Chapter 3, we can safely ignore the last term in (2.3) since this term converges to zero as $T\to\infty$.[4] When the last term is omitted from (2.3), this objective function is referred to as the *approximate* log-likelihood function, whereas (2.3) is the *exact* log-likelihood function. Typically, there is no ambiguity as to which likelihood is being discussed and we refer simply to the log-likelihood function l.

Focusing on the approximate log-likelihood function, fixing $\beta\in\Theta$, and taking the limit as $T\to\infty$ gives, under the assumption that sample moments converge to their population counterparts, the associated population criterion function

$$Q_0(\beta) = E\left[\log f\left(y_t|\vec{y}^{\,J}_{t-1};\beta\right)\right]. \tag{2.6}$$

To see that the β_0 generating the observed data is a maximizer of (2.6), and hence that this choice of Q_0 underlies a sensible estimation strategy, we observe that since the conditional density integrates to 1,

$$\begin{aligned} 0 &= \frac{\partial}{\partial\beta}\int_{-\infty}^{\infty} f\left(y_t|\vec{y}^{\,J}_{t-1};\beta_0\right)dy_t \\ &= \int_{-\infty}^{\infty}\frac{\partial \log f}{\partial\beta}\left(y_t|\vec{y}^{\,J}_{t-1};\beta_0\right)f\left(y_t|\vec{y}^{\,J}_{t-1};\beta_0\right)dy_t \\ &= E\left[\frac{\partial \log f}{\partial\beta}(y_t|\vec{y}^{\,J}_{t-1};\beta_0)\,\middle|\,\vec{y}^{\,J}_{t-1}\right], \end{aligned} \tag{2.7}$$

which, by the law of iterated expectations, implies that

$$\frac{\partial Q_0}{\partial\beta}(\beta_0) = E\left[\frac{\partial \log f}{\partial\beta}(y_t|\vec{y}^{\,J}_{t-1};\beta_0)\right] = E\left[\mathcal{D}(z_t;\beta_0)\right] = 0. \tag{2.8}$$

Thus, for ML estimation, (2.8) is the set of constraints on the joint distribution of $\vec{y}_T$ used in estimation, the ML version of (1.10). Critical to (2.8)

[3] The fact that the sum in (2.4) begins at $J+1$ is inconsequential, because we are focusing on the properties of b_T^{ML} (or θ_T) for large T, and J is fixed a priori by the asset pricing theory.

[4] There are circumstances where the small-sample properties of b_T^{ML} may be substantially affected by inclusion or omission of the term $\log f_m(\vec{y}_J;\beta)$ from the likelihood function. Some of these are explored in later chapters.

being satisfied by β_0 is the assumption that the conditional density f implied by the DAPM is in fact the density from which the data are drawn.

An important special case of this estimation problem is where $\{y_t\}$ is an independently and identically distributed (i.i.d.) process. In this case, if $f_m(y_t;\beta)$ denotes the density function of the vector y_t evaluated at β, then the log-likelihood function takes the simple form

$$l_T(\beta) \equiv T^{-1} \log L_T(\beta) = \frac{1}{T} \sum_{t=1}^{T} \log f_m(y_t;\beta). \tag{2.9}$$

This is an immediate implication of the independence assumption, since the joint density function of $\vec{y}_T$ factors into the product of the marginal densities of the y_t. The ML estimator of β_0 is obtained by maximizing (2.9) over $\beta \in \Theta$. The corresponding population criterion function is $Q_0(\beta) = E[\log f_m(y_t;\beta)]$.

Though the simplicity of (2.9) is convenient, most dynamic asset pricing theories imply that at least some of the observed variables y are not independently distributed over time. Dependence might arise, for example, because of mean reversion in an asset return or persistence in the volatility of one or more variables (see the next example). Such time variation in conditional moments is accommodated in the formulation (2.1) of the conditional density of y_t, but not by (2.9).

Example 2.1. *Cox, Ingersoll, and Ross [Cox et al., 1985b] (CIR) developed a theory of the term structure of interest rates in which the instantaneous short-term rate of interest, r, follows the mean reverting diffusion*

$$dr = \kappa(\bar{r} - r)\, dt + \sigma\sqrt{r}\, dB. \tag{2.10}$$

An implication of (2.10) is that the conditional density of r_{t+1} given r_t is

$$f(r_{t+1}|r_t;\beta_0) = ce^{-u_t - v_{t+1}} \left(\frac{v_{t+1}}{u_t}\right)^{\frac{q}{2}} I_q\big(2(u_t v_{t+1})^{\frac{1}{2}}\big), \tag{2.11}$$

where

$$c = \frac{2\kappa}{\sigma^2(1-e^{-\kappa})}, \tag{2.12}$$

$$u_t = \frac{2\kappa}{\sigma^2(1-e^{-\kappa})} e^{-\kappa} r_t, \tag{2.13}$$

$$v_{t+1} = \frac{2\kappa}{\sigma^2(1-e^{-\kappa})} r_{t+1}, \tag{2.14}$$

$q = 2\kappa\bar{r}/\sigma^2 - 1$, and I_q is the modified Bessel function of the first kind of order q. This is the density function of a noncentral χ^2 with $2q + 2$ degrees of freedom and noncentrality parameter $2u_t$. For this example, ML estimation would proceed by substituting (1.11) into (2.4) and solving for b_T^{ML}. The short-rate process (2.10) is the continuous time version of an interest-rate process that is mean reverting to a long-run mean of $\bar{r}$ and that has a conditional volatility of $\sigma\sqrt{r}$. This process is Markovian and, therefore, $\vec{y}_t^{\,J} = y_t$, which explains the single lag in the conditioning information in (1.11).

Though desirable for its efficiency, ML may not be, and indeed typically is not, a feasible estimation strategy for DAPMs, as often they do not provide us with complete knowledge of the relevant conditional distributions. Moreover, in some cases, even when these distributions are known, the computational burdens may be so great that one may want to choose an estimation strategy that uses only a portion of the available information. This is a consideration in the preceding example given the presence of the modified Bessel function in the conditional density of r. Later in this chapter we consider the case where only limited information about the conditional distribution is known or, for computational or other reasons, is used in estimation.

2.2. No Information about the Distribution

At the opposite end of the knowledge spectrum about the distribution of $\vec{y}_T$ is the case where we do not have a well-developed DAPM to describe the relationships among the variables of interest. In such circumstances, we may be interested in learning something about the joint distribution of the vector of variables z_t (which is presumed to include some asset prices or returns). For instance, we are often in a situation of wondering whether certain variables are correlated with each other or if one variable can predict another. Without knowledge of the joint distribution of the variables of interest, researchers typically proceed by *projecting* one variable onto another to see if they are related. The properties of the estimators in such projections are examined under this case of no information.[5] Additionally, there are occasions when we reject a theory and a replacement theory that explains the rejection has yet to be developed. On such occasions, many have resorted to projections of one variable onto others with the hope of learning more about the source of the initial rejection. Following is an example of this second situation.

[5] Projections, and in particular linear projections, are a simple and often informative first approach to examining statistical dependencies among variables. More complex, nonlinear relations can be explored with nonparametric statistical methods. The applications of nonparametric methods to asset pricing problems are explored in subsequent chapters.

Example 2.2. *Several scholars writing in the 1970s argued that, if foreign currency markets are informationally efficient, then the forward price for delivery of foreign exchange one period hence (F_t^1) should equal the market's best forecast of the spot exchange rate next period (S_{t+1}):*

$$F_t^1 = E[S_{t+1}|I_t], \tag{2.15}$$

where I_t denotes the market's information at date t. This theory of exchange rate determination was often evaluated by projecting $S_{t+1} - F_t^1$ onto a vector x_t and testing whether the coefficients on x_t are zero (e.g., Hansen and Hodrick, 1980). The evidence suggested that these coefficients are not zero, which was interpreted as evidence of a time-varying market risk premium $\lambda_t \equiv E[S_{t+1}|I_t] - F_t^1$ (see, e.g., Grauer et al., 1976, and Stockman, 1978). Theory has provided limited guidance as to which variables determine the risk premiums or the functional forms of premiums. Therefore, researchers have projected the spread $S_{t+1} - F_t^1$ onto a variety of variables known at date t and thought to potentially explain variation in the risk premium. The objective of the latter studies was to test for dependence of λ_t on the explanatory variables, say x_t.

To be more precise about what is meant by a *projection*, let L^2 denote the set of (scalar) random variables that have finite second moments:

$$L^2 = \{\text{random variables } x \text{ such that } Ex^2 < \infty\}. \tag{2.16}$$

We define an inner product on L^2 by

$$\langle\, x \,|\, y \,\rangle \equiv E(xy), \qquad x, y \in L^2, \tag{2.17}$$

and a norm by

$$\|\, x \,\| = [\langle\, x \,|\, x \,\rangle]^{\frac{1}{2}} = \sqrt{E(x^2)}. \tag{2.18}$$

We say that two random variables x and y in L^2 are *orthogonal* to each other if $E(xy) = 0$. Note that being orthogonal is not equivalent to being uncorrelated as the means of the random variables may be nonzero.

Let A be the closed linear subspace of L^2 generated by all linear combinations of the K random variables $\{x_1, x_2, \ldots, x_K\}$. Suppose that we want to project the random variable $y \in L^2$ onto A in order to obtain its best linear predictor. Letting $\delta' \equiv (\delta_1, \ldots, \delta_K)$, the best linear predictor is that element of A that minimizes the distance between y and the linear space A:

$$\min_{z \in A} \|\, y - z \,\| \quad \Leftrightarrow \quad \min_{\delta \in \mathbb{R}^K} \|\, y - \delta_1 x_1 - \ldots - \delta_K x_K \,\|. \tag{2.19}$$

The *orthogonal projection theorem*[6] tells us that the *unique* solution to (2.19) is given by the $\delta_0 \in \mathbb{R}^K$ satisfying

$$E\left[(y - x'\delta_0)x\right] = 0, \quad x' = (x_1, \ldots, x_K); \tag{2.20}$$

that is, the forecast error $u \equiv (y - x'\delta_0)$ is orthogonal to all linear combinations of x. The solution to the first-order condition (2.20) is

$$\delta_0 = E[xx']^{-1}E[xy]. \tag{2.21}$$

In terms of our notation for criterion functions, the population criterion function associated with least-squares projection is

$$Q_0(\delta) = E\left[(y_t - x_t'\delta)^2\right], \tag{2.22}$$

and this choice is equivalent to choosing $z_t' = (y_t, x_t')$ and the function $\mathcal{D}$ as

$$\mathcal{D}(z_t; \delta) = (y_t - x_t'\delta)x_t. \tag{2.23}$$

The interpretation of this choice is a bit different than in most estimation problems, because our presumption is that one is proceeding with estimation in the absence of a DAPM from which restrictions on the distribution of (y_t, x_t) can be deduced. In the case of a least-squares projection, we view the moment equation

$$E\left[\mathcal{D}(y_t, x_t; \delta_0)\right] = E\left[(y_t - x_t'\delta_0)x_t\right] = 0 \tag{2.24}$$

as the moment restriction that *defines* δ_0.

The sample least-squares objective function is

$$Q_T(\delta) = \frac{1}{T}\sum_{t=1}^{T}(y_t - x_t'\delta)^2, \tag{2.25}$$

with minimizer

$$\delta_T = \left[\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right]^{-1} \frac{1}{T}\sum_{t=1}^{T} x_t y_t. \tag{2.26}$$

[6] The orthogonal projection theorem says that if L is an inner product space, M is a closed linear subspace of L, and y is an element of L, then $z^* \in M$ is the unique solution to

$$\min_{z \in M} \| y - z \|$$

if and only if $y - z^*$ is orthogonal to all elements of M. See, e.g., Luenberger (1969).

The estimator δ_T is also obtained directly by replacing the population moments in (2.21) by their sample counterparts.

In the context of the pricing model for foreign currency prices, researchers have projected $(S_{t+1} - F_t^1)$ onto a vector of explanatory variables x_t. The variable being predicted in such analyses, $(S_{t+1} - F_t^1)$, is not the risk premium, $\lambda_t = E[(S_{t+1} - F_t^1)|I_t]$. Nevertheless, the resulting predictor in the population, $x_t'\delta_0$, is the same regardless of whether λ_t or $(S_{t+1} - F_t^1)$ is the variable being forecast. To see this, we digress briefly to discuss the difference between *best linear* and *best* prediction.

The predictor $x_t'\delta_0$ is the best linear predictor, which is defined by the condition that the projection error $u_t = y_t - x_t'\delta_0$ is orthogonal to all linear combinations of x_t. Predicting y_t using linear combinations of x_t is only one of many possible approaches to prediction. In particular, we could also consider prediction based on both linear and nonlinear functions of the elements of x_t. Pursuing this idea, let V denote the closed linear subspace of L^2 generated by all random variables $g(x_t)$ with finite second moments:

$$V = \left\{ g(x_t) : g : \mathbb{R}^K \to \mathbb{R}, \text{ and } g(x_t) \in L^2 \right\}. \tag{2.27}$$

Consider the new minimization problem $\min_{z \in V} \| y_t - z_t \|$. By the orthogonal projection theorem, the unique solution z_t^* to this problem has the property that $(y_t - z_t^*)$ is orthogonal to all $z_t \in V$. One representation of z^* is the conditional expectation $E[y_t|x_t]$. This follows immediately from the properties of conditional expectations: the error $\epsilon_t = y_t - E[y_t|x_t]$ satisfies

$$E[\epsilon_t g(x_t)] = E\left[(y_t - E[y_t|x_t]) g(x_t) \right] = 0, \tag{2.28}$$

for all $g(x_t) \in V$. Clearly, $A \subseteq V$ so the best predictor is at least as good as the best linear predictor. The precise sense in which best prediction is better is that, whereas ϵ_t is orthogonal to *all* functions of the conditioning information x_t, u_t is orthogonal to only linear combinations of x_t.

There are circumstances where best and best linear predictors coincide. This is true whenever the conditional expectation $E[y_t|x_t]$ is linear in x_t. One well-known case where this holds is when (y_t, x_t') is distributed as a multivariate normal random vector. However, normality is not necessary for best and best linear predictors to coincide. For instance, consider again Example 2.1. The conditional mean $E[r_{t+\Delta}|r_t]$ for positive time interval Δ is given by (Cox et al., 1985b)

$$\mu_{rt}(\Delta) \equiv E[r_{t+\Delta}|r_t] = r_t e^{-\Delta\kappa} + \bar{r}(1 - e^{-\Delta\kappa}), \tag{2.29}$$

which is linear in r_t, yet neither the joint distribution of $(r_t, r_{t-\Delta})$ nor the distribution of r_t conditioned on $r_{t-\Delta}$ is normal. (The latter is noncentral chi-square.)

With these observations in mind, we can now complete our argument that the properties of risk premiums can be studied by linearly projecting $(S_{t+1} - F_t^1)$ onto x_t. Letting $\mathrm{Proj}[\cdot|x_t]$ denote linear least-squares projection onto x_t, we get

$$\begin{aligned}\mathrm{Proj}[\lambda_t|x_t] &= \mathrm{Proj}\left[\left(S_{t+1} - F_t^1\right) - \epsilon_{t+1}\middle|x_t\right]\\ &= \mathrm{Proj}\left[\left(S_{t+1} - F_t^1\right)\middle|x_t\right],\end{aligned} \tag{2.30}$$

where $\epsilon_{t+1} \equiv (S_{t+1} - F_t^1) - \lambda_t$. The first equality follows from the definition of the risk premium as $E[S_{t+1} - F_t^1|I_t]$ and the second follows from the fact that ϵ_{t+1} is orthogonal to all functions of x_t including linear functions.

2.3. Limited Information: GMM Estimators

In between the cases of full information and no information about the joint distribution of $\vec{y}_T$ are all of the intermediate cases of *limited information.* Suppose that estimation of a parameter vector θ_0 in the admissible parameter space $\Phi \subset \mathbb{R}^K$ is to be based on a sample $\vec{z}_T$, where z_t is a subvector of the complete set of variables y_t appearing in a DAPM.[7] The restrictions on the distribution of $\vec{z}_T$ to be used in estimating θ_0 are summarized as a set of restrictions on the moments of functions of z_t. These moment restrictions may be either *conditional* or *unconditional.*

2.3.1. Unconditional Moment Restrictions

Consider first the case where a DAPM implies that the unconditional moment restriction

$$E[h(z_t;\theta_0)] = 0 \tag{2.31}$$

is satisfied uniquely by θ_0, where h is an M-dimensional vector with $M \geq K$. The function h may define standard central or noncentral moments of asset returns, the orthogonality of forecast errors to variables in agents' information sets, and so on. Illustrations based on Example 2.1 are presented later in this section.

To develop an estimator of θ_0 based on (2.31), consider first the case of $K = M$; the number of moment restrictions equals the number of parameters to be estimated. The function $H_0 : \Phi \to \mathbb{R}^M$ defined by $H_0(\theta) =$

[7] There is no requirement that the dimension of Φ be as large as the dimension of the parameter space Θ considered in full information estimation; often Φ is a lower-dimensional subspace of Θ, just as z_t may be a subvector of y_t. However, for notational convenience, we always set the dimension of the parameter vector of interest to K, whether it is θ_0 or β_0.

$E[h(z_t;\theta)]$ satisfies $H_0(\theta_0) = 0$. Therefore, a natural estimation strategy for θ_0 is to replace H_0 by its sample counterpart,

$$H_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} h(z_t;\theta), \tag{2.32}$$

and choose the estimator θ_T to set (2.32) to zero. If H_T converges to its population counterpart as T gets large, $H_T(\theta) \to H_0(\theta)$, for all $\theta \in \Phi$, then under regularity conditions we should expect that $\theta_T \to \theta_0$. The estimator θ_T is an example of what Hansen (1982b) refers to as a generalized method-of-moments, or GMM, estimator of θ_0.

Next suppose that $M > K$. Then there is not in general a unique way of solving for the K unknowns using the M equations $H_T(\theta) = 0$, and our strategy for choosing θ_T must be modified. We proceed to form K linear combinations of the M moment equations to end up with K equations in the K unknown parameters. That is, letting $\bar{\mathcal{A}}$ denote the set of $K \times M$ (constant) matrices of rank K, we select an $A \in \bar{\mathcal{A}}$ and set

$$\mathcal{D}^A(z_t;\theta) = Ah(z_t;\theta), \tag{2.33}$$

with this choice of $\mathcal{D}^A$ determining the estimation strategy. Different choices of $A \in \bar{\mathcal{A}}$ index (lead to) different estimation strategies. To arrive at a sample counterpart to (2.33), we select a possibly sample-dependent matrix A_T with the property that $A_T \to A$ (almost surely) as sample size gets large. Then the $K \times 1$ vector θ_T^A (the superscript A indicating that the estimator is A-dependent) is chosen to satisfy the K equations $\sum_t \mathcal{D}_T(z_t, \theta_T^A) = 0$, where $\mathcal{D}_T(z_t, \theta_T^A) = A_T h(z_t;\theta_T^A)$. Note that we are now allowing $\mathcal{D}_T$ to be sample dependent directly, and not only through its dependence on θ_T^A. This will frequently be the case in subsequent applications.

The construction of GMM estimators using this choice of $\mathcal{D}_T$ can be related to the approach to estimation involving a criterion function as follows: Let $\{a_T : T \geq 1\}$ be a sequence of $s \times M$ matrices of rank s, $K \leq s \leq M$, and consider the function

$$Q_T(\theta) = |a_T H_T(\theta)|, \tag{2.34}$$

where $|\cdot|$ denotes the Euclidean norm. Then

$$\underset{\theta}{\operatorname{argmin}}\, |a_T H_T(\theta)| = \underset{\theta}{\operatorname{argmin}}\, |a_T H_T(\theta)|^2 = \underset{\theta}{\operatorname{argmin}}\, H_T(\theta)' a_T' a_T H_T(\theta), \tag{2.35}$$

and we can think of our criterion function Q_T as being the quadratic form

$$Q_T(\theta) = H_T'(\theta) W_T H_T(\theta), \tag{2.36}$$

where $W_T \equiv a_T' a_T$ is often referred to as the *distance matrix*. This is the GMM criterion function studied by Hansen (1982b). The first-order conditions for this minimization problem are

$$\frac{\partial H_T}{\partial \theta}(\theta_T)' W_T H_T(\theta_T) = 0. \tag{2.37}$$

By setting

$$A_T = [\partial H_T(\theta_T)'/\partial \theta]\, W_T, \tag{2.38}$$

we obtain the $\mathcal{D}_T(z_t; \theta)$ associated with Hansen's GMM estimator.

The population counterpart to Q_T in (2.36) is

$$Q_0(\theta) = E[h(z_t; \theta)]' W_0 E[h(z_t; \theta)]. \tag{2.39}$$

The corresponding population $\mathcal{D}_0(z_t, \theta)$ is given by

$$\mathcal{D}_0(z_t, \theta) = E\left[\frac{\partial h}{\partial \theta}(z_t; \theta_0)'\right] W_0 h(z_t; \theta) \equiv A_0 h(z_t; \theta), \tag{2.40}$$

where W_0 is the (almost sure) limit of W_T as T gets large. Here $\mathcal{D}_0$ is not sample dependent, possibly in contrast to $\mathcal{D}_T$.

Whereas the first-order conditions to (2.36) give an estimator in the class $\bar{\mathcal{A}}$ [with A defined by (2.40)], not all GMM estimators in $\bar{\mathcal{A}}$ are the first-order conditions from minimizing an objective function of the form (2.36). Nevertheless, it turns out that the *optimal* GMM estimators in $\bar{\mathcal{A}}$, in the sense of being asymptotically most efficient (see Chapter 3), can be represented as the solution to (2.36) for appropriate choice of W_T. Therefore, the large-sample properties of GMM estimators are henceforth discussed relative to the sequence of objective functions $\{Q_T(\cdot) : T \geq 1\}$ in (2.36).

2.3.2. Conditional Moment Restrictions

In some cases, a DAPM implies the stronger, conditional moment restrictions

$$E[h(z_{t+n}; \theta_0)|I_t] = 0, \qquad \text{for given } n \geq 1, \tag{2.41}$$

where the possibility of $n > 1$ is introduced to allow the conditional moment restrictions to apply to asset prices or other variables more than one period in the future. Again, the dimension of h is M, and the information set I_t may be generated by variables other than the history of z_t.

To construct an estimator of θ_0 based on (2.41), we proceed as in the case of unconditional moment restrictions and choose K sample moment equations in the K unknowns θ. However, because $h(z_{t+n}; \theta_0)$ is orthogonal to any random variable in the information set I_t, we have much more flexibility in choosing these moment equations than in the preceding case. Specifically, we introduce a class of $K \times M$ full-rank "instrument" matrices $\mathcal{A}_t$ with each $A_t \in \mathcal{A}_t$ having elements in I_t. For any $A_t \in \mathcal{A}_t$, (2.41) implies that

$$E[A_t h(z_{t+n}; \theta_0)] = 0 \tag{2.42}$$

at $\theta = \theta_0$. Therefore, we can define a family of GMM estimators indexed by $A \in \mathcal{A}$, θ_T^A, as the solutions to the corresponding sample moment equations,

$$\frac{1}{T} \sum_t A_t h\big(z_{t+n}; \theta_T^A\big) = 0. \tag{2.43}$$

If the sample mean of $A_t h(z_{t+n}; \theta)$ in (2.43) converges to its population counterpart in (2.42), for all $\theta \in \Phi$, and A_t and h are chosen so that θ_0 is the unique element of Φ satisfying (2.42), then we might reasonably expect θ_T^A to converge to θ_0 as T gets large. The large-sample distribution of θ_T^A depends, in general, on the choice of A_t.[8]

The GMM estimator, as just defined, is not the extreme value of a specific criterion function. Rather, (2.42) defines θ_0 as the solution to K moment equations in K unknowns, and θ_T solves the sample counterpart of these equations. In this case, $\mathcal{D}_0$ is chosen directly as

$$\mathcal{D}_0(z_{t+n}, A_t; \theta) = \mathcal{D}_T(z_{t+n}, A_t; \theta) = A_t h(z_{t+n}; \theta). \tag{2.44}$$

Once we have chosen an A_t in $\mathcal{A}_t$, we can view a GMM estimator constructed from (2.41) as, trivially, a special case of an estimator based on unconditional moment restrictions. Expression (2.42) is taken to be the basic K moment equations that we start with. However, the important distinguishing feature of the class of estimators $\mathcal{A}_t$, compared to the class $\bar{\mathcal{A}}$, is that the former class offers much more flexibility in choosing the weights on h. We will see in Chapter 3 that the most efficient estimator in the class $\mathcal{A}$ is often more efficient than its counterpart in $\bar{\mathcal{A}}$. That is, (2.41) allows one to exploit more information about the distribution of z_t than (2.31) in the estimation of θ_0.

[8] As is discussed more extensively in the context of subsequent applications, this GMM estimation strategy is a generalization of the instrumental variables estimators proposed for classical simultaneous equations models by Amemiya (1974) and Jorgenson and Laffont (1974), among others.

2.3.3. Linear Projection as a GMM Estimator

Perhaps the simplest example of a GMM estimator based on the moment restriction (2.31) is linear least-squares projection. Suppose that we project y_t onto x_t. Then the best linear predictor is defined by the moment equation (2.20). Thus, if we define

$$h\big(y_t, x_t; \delta\big) = \big(y_t - x_t'\delta\big)\, x_t, \tag{2.45}$$

then by construction δ_0 satisfies $E[h(y_t, x_t; \delta_0)] = 0$.

One might be tempted to view linear projection as special case of a GMM estimator in $\mathcal{A}_t$ by choosing $n = 0$,

$$A_t = x_t \quad \text{and} \quad h\big(y_t, x_t; \delta\big) = \big(y_t - x_t'\delta\big). \tag{2.46}$$

However, importantly, we are not free to select among other choices of $A_t \in \mathcal{A}_t$ in constructing a GMM estimator of the linear predictor $x_t'\delta_0$. Therefore, least-squares projection is appropriately viewed as a GMM estimator in $\bar{\mathcal{A}}$.

Circumstances change if a DAPM implies the stronger moment restriction

$$E\big[\big(y_t - x_t'\delta_0\big)\big| x_t\big] = 0. \tag{2.47}$$

Now we are no longer in an environment of complete ignorance about the distribution of (y_t, x_t), as it is being assumed that $x_t'\delta_0$ is the best, not just the best linear, predictor of y_t. In this case, we are free to choose

$$A_t = g(x_t) \quad \text{and} \quad h\big(y_t, x_t; \delta\big) = \big(y_t - x_t'\delta\big), \tag{2.48}$$

for any $g : \mathbb{R}^K \rightarrow \mathbb{R}^K$. Thus, the assumption that the best predictor is linear puts us in the case of conditional moment restrictions and opens up the possibility of selecting estimators in $\mathcal{A}$ defined by the functions g.

2.3.4. Quasi-Maximum Likelihood Estimation

Another important example of a limited information estimator that is a special case of a GMM estimator is the *quasi-maximum likelihood* (QML) estimator. Suppose that $n = 1$ and that I_t is generated by the J-history $\vec{y}_t^{\,J}$ of a vector of observed variables y_t.[9] Further, suppose that the functional

[9] We employ the usual, informal notation of letting I_t or $\vec{y}_t^{\,J}$ denote the σ-algebra (information set) used to construct conditional moments and distributions.

forms of the population mean and variance of y_{t+1}, conditioned on I_t, are known and let θ denote the vector of parameters governing these first two conditional moments. Then ML estimation of θ_0 based on the classical normal conditional likelihood function gives an estimator that converges to θ_0 and is normally distributed in large samples (see, e.g., Bollerslev and Wooldridge, 1992).

Referring back to the introductory remarks in Chapter 1, we see that the function $\mathcal{D}$ $(= \mathcal{D}_0 = \mathcal{D}_T)$ determining the moments used in estimation in this case is

$$\mathcal{D}(z_t; \theta) = \frac{\partial \log f_N}{\partial \theta}\big(y_t \big| \vec{y}^{\,J}_{t-1}; \theta\big), \tag{2.49}$$

where $z_t' = (y_t', \vec{y}^{\,J\prime}_{t-1})$ and f_N is the normal density function conditioned on $\vec{y}^{\,J}_{t-1}$. Thus, for QML to be an admissible estimation strategy for this DAPM it must be the case that θ_0 satisfies

$$E\left[\frac{\partial \log f_N}{\partial \theta}\big(y_t \big| \vec{y}^{\,J}_{t-1}; \theta_0\big)\right] = 0. \tag{2.50}$$

The reason that θ_0 does in fact satisfy (2.50) is that the first two conditional moments of y_t are correctly specified and the normal distribution is fully characterized by its first two moments. This intuition is formalized in Chapter 3. The moment equation (2.50) defines a GMM estimator.

2.3.5. Illustrations Based on Interest Rate Models

Consider again the one-factor interest rate model presented in Example 2.1. Equation (2.29) implies that we can choose

$$h\big(\vec{z}^{\,1}_{t+1}; \theta_0\big) = \big[r_{t+1} - \bar{r}(1 - e^{-\kappa}) - e^{-\kappa} r_t\big], \tag{2.51}$$

where $\vec{z}^{\,2}_{t+1} = (r_{t+1}, r_t)'$. Furthermore, for any 2×1 vector function $g(r_t)$: $\mathbb{R} \to \mathbb{R}^2$, we can set $A_t = g(r_t)$ and

$$E\big[(r_{t+1} - \bar{r}(1 - e^{-\kappa}) - e^{-\kappa} r_t) g(r_t)\big] = 0. \tag{2.52}$$

Therefore, a GMM estimator $\theta_T^{A'} = (\bar{r}_T, \kappa_T)$ of $\theta_0' = (\bar{r}, \kappa)$ can be constructed from the sample moment equations

$$\frac{1}{T}\sum_t \big[r_{t+1} - \bar{r}_T(1 - e^{-\kappa_T}) - e^{-\kappa_T} r_t\big] g(r_t) = 0. \tag{2.53}$$

Each choice of $g(r_t) \in \mathcal{A}_t$ gives rise to a different GMM estimator that in general has a different large-sample distribution. Linear projection of r_t onto r_{t-1} is obtained as the special case with $g(r_{t-1})' = (1, r_{t-1})$, $M = K = 2$, and $\theta' = (\kappa, \bar{r})$.

Turning to the implementation of QML estimation in this example, the mean of $r_{t+\Delta}$ conditioned on r_t is given by (2.29) and the conditional variance is given by (Cox et al., 1985b)

$$\sigma_{rt}^2(\Delta) \equiv Var[r_{t+\Delta}|r_t] = r_t \frac{\sigma^2}{\kappa}(e^{-\Delta\kappa} - e^{-2\Delta\kappa}) + \bar{r}\frac{\sigma^2}{2\kappa}(1 - e^{-\Delta\kappa})^2. \quad (2.54)$$

If we set $\Delta = 1$, it follows that discretely sampled returns $(r_t, r_{t-1}, \ldots)$ follow the model

$$r_{t+1} = \bar{r}(1 - e^{-\kappa}) + e^{-\kappa} r_t + \sqrt{\sigma_{rt}^2}\epsilon_{t+1}, \quad (2.55)$$

where the error term ϵ_{t+1} in (2.55) has (conditional) mean zero and variance one. For this model, $\theta_0 = (\bar{r}, \kappa, \sigma^2)' = \beta_0$ (the parameter vector that describes the entire distribution of r_t), though this is often not true in other applications of QML.

The conditional distribution of r_t is a noncentral χ^2. However, suppose we ignore this fact and proceed to construct a likelihood function based on our knowledge of (2.29) and (2.54), assuming that the return r_t is distributed as a normal conditional on r_{t-1}. Then the log-likelihood function is (l^q to indicate that this is QML)

$$l_T^q(\theta) \equiv \frac{1}{T}\sum_{t=2}^{T}\left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\sigma_{rt-1}^2\right) - \frac{1}{2}\frac{(r_t - \mu_{rt-1})^2}{\sigma_{rt-1}^2}\right). \quad (2.56)$$

Computing first-order conditions gives

$$\frac{\partial l_T^q}{\partial\theta_j}\left(\theta_T^q\right) = \frac{1}{T}\sum_{t=2}^{T} -\frac{1}{2\hat{\sigma}_{rt-1}^2}\frac{\partial\hat{\sigma}_{rt-1}^2}{\partial\theta_j} + \frac{1}{2}\frac{(r_t - \hat{\mu}_{rt-1})^2}{\hat{\sigma}_{rt-1}^4}\frac{\partial\hat{\sigma}_{rt-1}^2}{\partial\theta_j}$$
$$+ \frac{(r_t - \hat{\mu}_{rt-1})}{\hat{\sigma}_{rt-1}^2}\frac{\partial\hat{\mu}_{rt-1}}{\partial\theta_j} = 0, \quad j = 1, 2, 3, \quad (2.57)$$

where θ_T^q denotes the QML estimator and $\hat{\mu}_{rt-1}$ and $\hat{\sigma}_{rt-1}^2$ are μ_{rt-1} and σ_{rt-1}^2 evaluated at θ_T^q. As suggested in the preceding section, this estimation strategy is admissible because the first two conditional moments are correctly specified.

Though one might want to pursue GMM or QML estimation for this interest rate example because of their computational simplicity, this is not the best illustration of a limited information problem because the true likelihood function is known. However, a slight modification of the interest rate process places us in an environment where GMM is a natural estimation strategy.

Example 2.3. *Suppose we extend the one-factor model introduced in Example 2.1 to the following two-factor model:*

$$\begin{aligned} dr &= \kappa(\bar{r} - r)\,dt + \sigma_r\sqrt{v}\,dB_r, \\ dv &= \nu(\bar{v} - v)\,dt + \sigma_v\sqrt{v}\,dB_v. \end{aligned} \tag{2.58}$$

In this two-factor model of the short rate, v plays the role of a stochastic volatility for r. Similar models have been studied by Anderson and Lund (1997a) and Dai and Singleton (2000). The volatility shock in this model is unobserved, so estimation and inference must be based on the sample $\vec{r}_T$ and r_t is no longer a Markov process conditioned on its own past history.

An implication of the assumptions that r mean reverts to the long-run value of $\bar{r}$ and that the conditional mean of r does not depend on v is that (2.29) is still satisfied in this two-factor model. However, the variance of r_t conditioned on r_{t-1} is not known in closed form, nor is the form of the density of r_t conditioned on $\vec{r}^{\,J}_{t-1}$. Thus, neither ML nor QML estimation strategies are easily pursued.[10] Faced with this limited information, one convenient strategy for estimating $\theta_0' \equiv (\bar{r}, \kappa)$ is to use the moment equations (2.52) implied by (2.29).

This GMM estimator of θ_0 ignores entirely the known structure of the volatility process and, indeed, σ_r^2 is not an element of θ_0. Thus, not only are we unable to recover any information about the parameters of the volatility equation using (2.52), but knowledge of the functional form of the volatility equation is ignored. It turns out that substantially more information about $f(r_t|r_{t-1};\theta_0)$ can be used in estimation, but to accomplish this we have to extend the GMM estimation strategy to allow for unobserved state variables. This extension is explored in depth in Chapter 6.

2.3.6. GMM Estimation of Pricing Kernels

As a final illustration, suppose that the pricing kernel in a DAPM is a function of a state vector x_t and parameter vector θ_0. In preference-based DAPMs, the pricing kernel can be interpreted as an agent's intertemporal

[10] Asymptotically efficient estimation strategies based on approximations to the true conditional density function of r have been developed for this model. These are described in Chapter 7.

Table 2.1. Summary of Population and Sample Objective Functions for Various Estimators

	Maximum likelihood	GMM	Least-squares projection
Population objective function	$\max_{\beta \in \Theta} E\left[\log f\left(y_t \,\middle|\, \vec{y}^{\,J}_{t-1}; \beta\right)\right]$	$\min_{\theta \in \Theta} E[h(z_t;\theta)]' W_0 E[h(z_t;\theta)]$	$\min_{\delta \in \mathbb{R}^K} E\left[\left(y_t - x_t'\delta\right)^2\right]$
Sample objective function	$\max_{\beta \in \Theta} \frac{1}{T}\sum_{t=J+1}^{T} \log f\left(y_t \,\middle|\, \vec{y}^{\,J}_{t-1}; \beta\right)$	$\min_{\theta \in \Theta} H_T(\theta)' W_T H_T(\theta)$ $H_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} h(z_t;\theta)$	$\min_{\delta \in \mathbb{R}^K} \frac{1}{T}\sum_{t=1}^{T} \left(y_t - x_t'\delta\right)^2$
Population F.O.C.	$E\left[\frac{\partial \log}{\partial \beta} f\left(y_t \,\middle|\, \vec{y}^{\,J}_{t-1}; \beta_0\right)\right] = 0$	$A_0 E[h(z_t;\theta_0)] = 0$	$E\left[\left(y_t - x_t'\delta_0\right) x_t\right] = 0$
Sample F.O.C.	$\frac{1}{T}\sum_{t=J+1}^{T} \frac{\partial \log}{\partial \beta} f\left(y_t \,\middle|\, \vec{y}^{\,J}_{t-1}; b_T^{ML}\right) = 0$	$A_T \frac{1}{T}\sum_{t=1}^{T} h(z_t;\theta_T) = 0$	$\frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\delta_T\right)x_t = 0$