Essays on Consumption and Expected Returns

A thesis presented

by

Motohiro Yogo

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

May 2004

Abstract

Essays on Consumption and Expected Returns

Thesis Advisor: Professor John Y. Campbell Author: Motohiro Yogo

This thesis consists of three essays on the relationship between consumption, the business cycle, and expected returns on financial assets.

The first essay is “A Consumption-Based Explanation of Expected Stock Returns”. When utility is non-separable in nondurable and durable consumption and the elasticity of substitution between the goods is high, marginal utility rises when durable consumption falls. The model explains both the cross-sectional variation in expected stock returns and the time variation in the equity premium. Small stocks and value stocks deliver relatively low returns during recessions when durable consumption falls, which explains their high average returns relative to big stocks and growth stocks. Stock returns are unexpectedly low at business-cycle troughs when durable consumption falls sharply, which explains the counter-cyclical variation in the equity premium.

The second essay is “Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak”. In the instrumental variables (IV) regression model, weak instruments can lead to bias in estimators and size distortion in hypothesis tests.

This essay examines how weak instruments affect the identification of the elasticity of intertemporal substitution (EIS) through the linearized Euler equation. Conventional IV methods result in an empirical puzzle that the EIS is significantly less than one while its inverse is not different from one. This essay shows that weak instruments can explain the puzzle and reports valid confidence intervals for the EIS using pivotal statistics. The EIS is less than one and not significantly different from zero for eleven developed countries.

The third essay is "Efficient Tests of Stock Return Predictability". Tests of the predictability of stock returns may be invalid when the predictor variable is persistent and its innovations are highly correlated with returns. This essay develops a pretest to determine whether the conventional t-test leads to incorrect inference and an efficient test of predictability that always leads to correct inference. Although the conventional t-test is highly misleading for the dividend-price and the smoothed earnings-price ratios, we find evidence for predictability using our test. We also find evidence for predictability with the short rate and the long-short yield spread, for which the conventional t-test leads to correct inference.

Contents

List of Tables

List of Figures

Acknowledgments

I became interested in the relationship between consumption, the business cycle, and asset prices as an undergraduate at Princeton. Working on a senior thesis on that topic, I had read many of John Campbell's papers (e.g. Campbell and Shiller (1988a), Campbell (1996), and Campbell and Cochrane (1999)). Through his work, I became captivated by the intimate connection between theory and evidence at the frontier of asset pricing.

Although I agonized quite a bit on the choice of graduate programs, I ultimately chose Harvard in order to work with John. I do not regret that decision. John has led me step by step in the process of becoming an academic: first equipping me with the necessary tools through a research assistantship, then teaching me to carry out a research project from start to finish through a joint paper (Chapter 3), and finally navigating me through my "job market paper" (Chapter 1). Although I can never repay my enormous debt, I hope to acknowledge my intellectual roots to future students in the same way that Bob Shiller's books are "visionary", his articles "seminal", and his algebra "clever" in John's graduate classes at Harvard.

Jim Stock has contributed significantly to my intellectual development, particulary in econometrics. What began as a course project for his time series class in my first year developed into a very productive three-year research collaboration on weak instruments. This work led me to write Chapter 2 of the thesis as my Economics 3000 paper. Chapter 3 of the thesis is inspired by Jim's earlier work on nonstationary time series (Cavanagh, Elliott, and Stock 1995, Elliott, Rothenberg, and Stock 1996). I thank Jim for his constant encouragement and support during my graduate studies.

Tuomo Vuolteenaho has given me invaluable guidance and support on the work contained in the thesis. I also want to thank others who have provided helpful feedback on parts of the thesis. Chapter 1 has benefited from comments by John Cochrane, Borja Larrain, Jonathan Parker, Monika Piazzesi, and seminar participants at Carnegie Mellon GSIA, Federal Reserve Board, Harvard, Harvard Business School, Kennedy School of Government, Michigan Business School, NYU Stern, Northwestern Kellogg, Princeton, Stanford GSB, UCLA Anderson, Wharton, and Yale School of Management. Chapter 2 has benefited from comments by Marcelo Moreira, Luis Viceira, Michael Woodford, and seminar participants at Harvard and the 2002 Econometric Society Australasian Meeting. Chapter 3 has benefited from comments by Andrew Ang, Geert Bekaert, Markku Lanne, Marcelo Moreira, Robert Shiller, Robert Stambaugh, Mark Watson, and seminar participants at Harvard, the Econometric Society Australasian Meeting 2002, and the NBER Summer Institute 2003.

I am very grateful for financial support through the Graduate School of Arts and Sciences Fellowship in 2000–2002 and the National Science Foundation Graduate Research Fellowship in 2002–2004. Without the generous support, I would not have been able to complete my doctoral work in four years.

What I will most miss about graduate school are the daily interactions with interesting and nice colleagues. In particular, I will miss the not always productive conversations with Borja Larrain, Steve Lin, Gustavo Suarez, Adam Szeidl, and Francesco Trebbi. I learned a lot about economics and life from these guys. Graduate school has not been without its ups and downs. Special thanks to Peter Lu and Stanley

Watt for bearing the downs with me through all these years.

Lastly, I thank my parents for a constant source of wisdom and encouragement.

Chapter 1

A Consumption-Based Explanation of Expected Stock Returns

1.1 Introduction

Explaining the variation in expected returns across stocks and the variation in the equity premium through time as a tradeoff between risk and return is a challenge for financial economists. In his review article on market efficiency, Fama (1991, p. 1610) concludes

> In the end, I think we can hope for a coherent story that (1) relates the cross-section properties of expected returns to the variation of expected returns through time, and (2) relates the behavior of expected returns to the real economy in a rather detailed way. Or we can hope to convince ourselves that no such story is possible.

This essay proposes a "coherent story" that satisfies both criteria.

A well-known empirical fact in finance is the high average returns of small stocks relative to big stocks (i.e. low relative to high market equity stocks) and value stocks relative to growth stocks (i.e. high relative to low book-to-market equity stocks). The evidence suggests that there are size and value premia in the cross section of expected stock returns. In an equilibrium asset pricing model, cross-sectional variation in expected returns must be explained by cross-sectional variation in risk. The Capital Asset Pricing Model (CAPM), where risk is measured by market beta, fails to explain the size and value premia (see Fama and French (1992) and references therein). The Consumption CAPM (CCAPM), where risk is measured by nondurable consumption beta, also fails to explain the cross section of expected stock returns (Mankiw and Shapiro 1986b, Breeden, Gibbons, and Litzenberger 1989).

Another well-known empirical fact is the predictability of stock returns by variables that are informative about the business cycle (see Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988b), and Fama and French (1988, 1989)). The evidence suggests that the equity premium is time varying, that it is higher at business-cycle troughs than at peaks. In an equilibrium asset pricing model, time variation in the equity premium must be explained by time variation in the price or quantity of risk. Although there is some evidence for time variation in risk, it cannot be reconciled with the evidence for expected returns in a way that offers a consistent description of the time-varying tradeoff between risk and return (see Harvey (1989) for evidence on the CAPM and Kandel and Stambaugh (1990) for the CCAPM).

This essay proposes a simple consumption-based explanation of both the cross-sectional variation in expected stock returns and the counter-cyclical variation in the equity premium. I use a representative household model, where intraperiod utility is a constant elasticity of substitution (CES) function of nondurable and durable consumption. It nests the CCAPM as a special case when utility is separable in the two consumption goods.

In the language of macroeconomics, the main findings can be summarized as follows. When the elasticity of substitution between nondurable and durable consumption is high, the marginal utility of consumption rises when durable consumption falls. First, small stocks and value stocks deliver low returns when marginal utility rises, that is during recessions when durable consumption falls. Investors must therefore be rewarded with high expected returns to hold these risky stocks. Second, stocks deliver unexpectedly low returns when marginal utility rises sharply, that is at business-cycle troughs when durable consumption falls sharply relative to nondurable consumption. Investors must therefore be rewarded with high expected returns to hold stocks during recessions.

In the language of finance, the main findings can be summarized as follows. When utility is non-separable in nondurable and durable consumption, optimal portfolio allocation implies a two-factor model in nondurable and durable consumption growth. The risk price for durable consumption is positive, provided that the elasticity of substitution between the two goods is high. First, small stocks and value stocks have higher durable consumption betas than big stocks and growth stocks. Simply put, the returns on small stocks and value stocks are more pro-cyclical, explaining their high average returns. Second, the covariance of stock returns with durable consumption growth is higher at business-cycle troughs than at peaks. The equity premium is therefore counter-cyclical because the quantity of risk, measured by the conditional covariance of returns with durable consumption growth, is counter-cyclical.

Previous papers that have tested the representative household model with durable consumption include Dunn and Singleton (1986), Eichenbaum and Hansen (1990), and Ogaki and Reinhart (1998). These papers test the conditional moment restrictions implied by the model using T-bill returns and instruments. This essay instead tests the unconditional moment restrictions using a large cross section of stock returns, and the conditional moment restrictions using stock returns and instruments that predict returns. Because both nondurable and durable consumption are smooth, the model requires large risk aversion to fit the high level and volatility of expected stock returns. This essay shows that the model can successfully explain the cross-sectional and time variation in expected stock returns, conditional on an "equity premium puzzle" (Mehra and Prescott 1985).

In related work, Pakoš (2003) considers a representative household model with non-homothetic utility in nondurable and durable consumption goods. He focuses on the Leontief model, where the elasticity of substitution between the two goods is zero. Since the consumption of durables relative to nondurables is pro-cyclical, a low elasticity of substitution between the goods implies pro-cyclical marginal utility. The Leontief model therefore cannot explain the value premium (since value stocks are more pro-cyclical than growth stocks) or the counter-cyclical variation in the equity premium. In contrast, I estimate a high elasticity of substitution between the goods, implying counter-cyclical marginal utility.

The rest of the essay is organized as follows. Section 1.2 reviews the household's consumption and portfolio choice problem in the presence of durable consumption goods. In Section 1.3, I linearize the unconditional Euler equation to obtain a two-factor model in nondurable and durable consumption growth. I show that the linear factor model can be estimated by Generalized Method of Moments (GMM). In Section 1.4, I linearize the conditional Euler equation to obtain a conditional factor model in nondurable and durable consumption growth. I show that its conditional moments can be estimated by an instrumental variables methodology (Campbell 1987, Harvey 1989).

Section 1.5 provides a description of the consumption data used in the empirical work. The service flow for "durable goods" (as defined in the national accounts) is more cyclical than the service flow for "nondurable goods" and "services". The high cyclicality of the service flow, rather than durability of the good, is the key ingredient in explaining the known facts about expected stock returns.

Section 1.6 reports the cross-sectional tests. I find that the durable consumption model explains the variation in average returns across the 25 Fama-French (1993) portfolios better than the Fama-French three-factor model; the R^2 for these models are 77% and 66%, respectively. The durable consumption model is not rejected by the test that the pricing errors are jointly zero, while the CAPM, the three-factor model, and the CCAPM are all rejected. The model also explains returns on portfolios sorted by book-to-market equity within industry and portfolios sorted by risk (i.e. pre-formation betas).

Section 1.7 reports the time series tests of the model. I estimate the conditional Euler equation by GMM, using excess stock returns and instruments. The test of overidentifying restrictions fails to reject the model. The CCAPM, which is a restriction that utility be separable in nondurable and durable consumption, is strongly rejected (Hansen and Singleton 1982). To connect these results to the predictability of stock returns, I jointly estimate the conditional mean and variance of stock returns and its conditional covariance with nondurable and durable consumption growth. I find that much of the counter-cyclical variation in the equity premium is driven by counter-cyclical variation in the conditional covariance of returns with durable (rather than nondurable) consumption growth, explaining the failure of the CCAPM.

The large risk aversion required to explain stock returns not only implies a high riskfree rate (Weil 1989), but high volatility in the riskfree rate due to the high persistence of durable consumption growth. In Section 1.8, I show that this "riskfree rate puzzle" can be resolved by separating risk aversion from the elasticity of intertemporal substitution (EIS) with more general preferences.

Section 1.9 offers some conclusions. Supplementary derivations and results are contained in Appendices A.1–A.3, referenced throughout the text.

1.2 Household Optimization with Durable Consumption Goods

1.2.1 Euler Equations

Consider the canonical consumption and portfolio choice problem of a household. In each period t, the household purchases C_t units of nondurable consumption goods and E_t units of durable consumption goods. P_t is the price of durable goods in units of nondurable goods. Nondurable goods are entirely consumed in the period of purchase, whereas durable goods provide service flows for more than one period. The household's stock of durable goods D_t is related to its expenditures by the law of motion

$$D_t = (1 - \delta)D_{t-1} + E_t, \tag{1.1}$$

where $\delta \in (0, 1)$ is the depreciation rate.

There are $N + 1$ tradeable assets in the economy, indexed by $i = 0, 1, \ldots, N$. In period $t - 1$, the household invests $B_{i,t-1}$ units of wealth W_{t-1} in asset i, which realizes the gross rate of return R_{it} in period t. Given the initial level of wealth, W_0, and the initial stock of durable goods, D_{-1}, the household chooses the sequence

$\{C_t, E_t, B_{0t}, \ldots, B_{Nt}\}_{t=0}^{\infty}$ to maximize

$$\mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t u(C_t, D_t) \tag{1.2}$$

subject to the intertemporal budget constraint

$$W_t = \sum_{i=0}^{N} B_{i,t-1} R_{it}, \tag{1.3}$$

$$\sum_{i=0}^{N} B_{it} = W_t - C_t - P_t E_t. \tag{1.4}$$

$\beta > 0$ is the household's subjective discount factor, and $u_t = u(C_t, D_t)$ is its period utility, which depends on the consumption of nondurable goods and the stock of durable goods.

Let u_{Ct} and u_{Dt} denote the marginal utility of C_t and D_t, respectively. The household's first-order conditions and the envelope theorem imply the pair of Euler equations

$$u_{C,t-1} = \mathbf{E}_{t-1}[\beta u_{Ct} R_{it}], \tag{1.5}$$

$$u_{D,t-1} = \mathbf{E}_{t-1}[\beta u_{Ct}(P_{t-1} R_{it} - (1-\delta) P_t)]. \tag{1.6}$$

Define the stochastic discount factor (SDF) as $M_t = \beta u_{Ct}/u_{C,t-1}$. Equations (1.5) and (1.6) together imply an intratemporal first-order condition (FOC) of the form

$$\frac{u_{D,t-1}}{u_{C,t-1}} = P_{t-1} - (1-\delta)\mathbf{E}_{t-1}[M_t P_t] = Q_{t-1}. \tag{1.7}$$

Since a unit of the durable consumption good costs P_{t-1} today and can be sold for $(1-\delta)P_t$ tomorrow, after depreciation, Q_{t-1} has a natural interpretation as the user cost of the service flow for the durable good.

1.2.2 Asset Pricing

Equation (1.5) can be written as

$$\mathbf{E}_{t-1}\left[M_t R_{it}\right] = 1. \tag{1.8}$$

The excess return on asset i then satisfies

$$\mathbf{E}_{t-1}[M_t(R_{it} - R_{0t})] = 0. \tag{1.9}$$

Equation (1.8) is the basis for consumption-based asset pricing. The marginal utility of consumption is the appropriate measure of risk for an investor who cares about consumption. Assets that deliver low returns when marginal utility is high must have high expected returns to reward the investor for bearing risk. On the other hand, assets that deliver high returns when marginal utility is high provides a good hedge for consumption risk and must consequently have low expected returns.

Equation (1.8) was derived here in the context of a household optimization problem, but it holds more generally by a well-known existence theorem. In the absence of arbitrage, there exists a strictly positive SDF, M_t, which satisfies equation (1.8) for all tradable assets $i = 0, 1, \ldots, N$ (see Cochrane (2001, Chapter 4.2)). Various asset pricing models correspond to particular forms of the SDF. In the consumption-based model, the SDF is the marginal rate of substitution in consumption.

1.2.3 CES Utility

I now specify a particular form of utility that is used in the empirical work. The period utility takes the form

$$u(C, D) = \frac{v(C, D)^{1-\gamma}}{1-\gamma}, \tag{1.10}$$

where $\gamma > 0$ is the coefficient of relative risk aversion with respect to intraperiod utility flow. The intraperiod utility takes the CES form

$$v(C, D) = [(1-\alpha)C^{\rho} + \alpha D^{\rho}]^{1/\rho}, \tag{1.11}$$

where $\alpha \in (0, 1)$ and $\rho \leq 1$. The elasticity of substitution between nondurable and durable consumption goods is $1/(1-\rho)$. Implicit in this specification is the assumption that the service flow for the durable good is linear in the stock of the durable good. I therefore use the words "stock" and "consumption" interchangeably in regard to durable goods, hopefully without confusion. The utility specification (1.10) and (1.11) has been used previously in related empirical work by Dunn and Singleton (1986) and Ogaki and Reinhart (1998).[1]

When $\rho = 1 - \gamma$, the period utility is separable in nondurable and durable consumption,

$$u(C, D) = (1-\alpha)\frac{C^{1-\gamma}}{1-\gamma} + \alpha\frac{D^{1-\gamma}}{1-\gamma}. \tag{1.12}$$

The marginal utility of nondurable consumption takes the simple form $u_C = (1-\alpha)C^{-\gamma}$. The additively separable model is the leading case in macroeconomics and finance applications. It provides a useful reference point for the general model with

[1]Dunn and Singleton (1986) use Cobb-Douglas intraperiod utility, which corresponds to the special case $\rho = 0$.

CES intraperiod utility.

In the general case, the marginal utility of nondurable consumption is

$$u_C = (1-\alpha)C^{-\gamma}\left[1+\alpha\left(\left(\frac{D}{C}\right)^{\rho}-1\right)\right]^{\frac{1-\gamma-\rho}{\rho}}. \tag{1.13}$$

The marginal utility under the additively separable model is now multiplied by a function of the ratio of durable to nondurable consumption, D/C. Figure 1.1 illustrates the dependence of the marginal utility on D/C. For a given level of nondurable consumption, marginal utility decreases in D/C if $\rho > 1-\gamma$. Intuitively, low nondurable consumption can be offset by high durable consumption provided that the elasticity of substitution between the two goods is sufficiently high. On the other hand, relatively high durable consumption increases the marginal utility of nondurable consumption if the elasticity is low (i.e. $\rho < 1-\gamma$). The additively separable model, where $\rho = 1-\gamma$, is the knife-edge case when the marginal utility is independent of durable consumption.

The SDF for the durable consumption model is

$$M_t = \beta\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\left(\frac{1+\alpha[(D_t/C_t)^{\rho}-1]}{1+\alpha[(D_{t-1}/C_{t-1})^{\rho}-1]}\right)^{\frac{1-\gamma-\rho}{\rho}}. \tag{1.14}$$

The intratemporal FOC (1.7) takes the form

$$\frac{\alpha}{1-\alpha}\left(\frac{D_{t-1}}{C_{t-1}}\right)^{\rho-1} = P_{t-1}-(1-\delta)\mathbf{E}_{t-1}[M_t P_t]. \tag{1.15}$$

In the empirical work, I assume that there is a representative household, so that assets can be priced by the SDF (1.14) using aggregate consumption data.

At the microeconomic level, there may be lumpiness in the adjustment of durable consumption, which can cause aggregate durable consumption to deviate from the

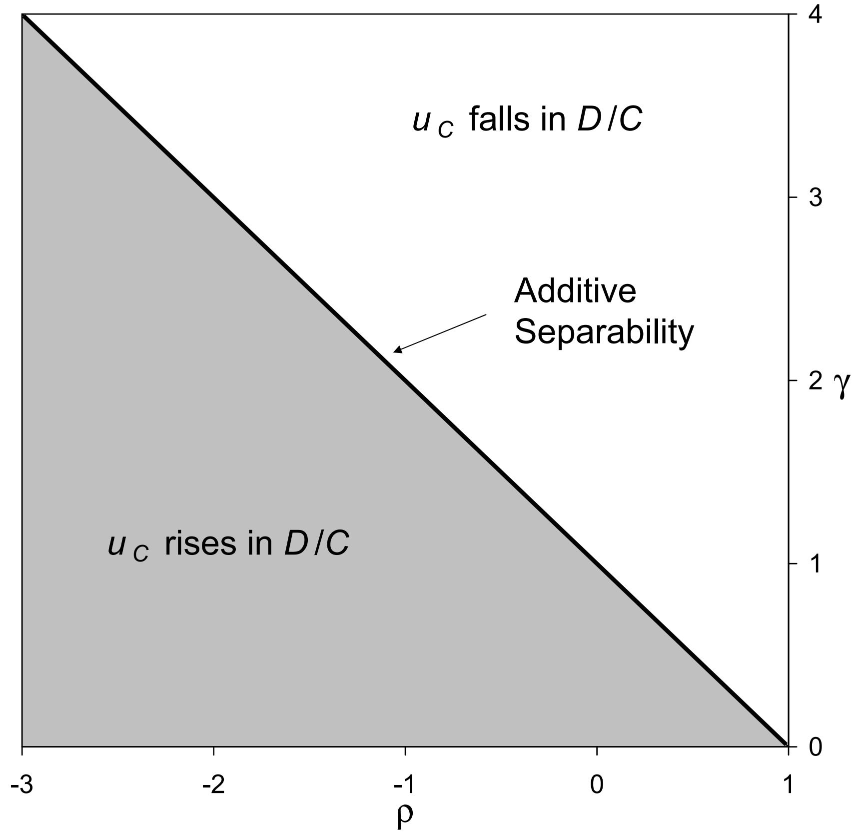

Figure 1.1: Marginal Utility of Nondurable Consumption

optimal behavior implied by the frictionless model (see Grossman and Laroque (1990) and Caballero (1993)). As long as nondurable consumption still adjusts in a way that the Euler equation (1.5) holds, the slow adjustment of durable consumption does not pose a problem for asset pricing.[2] However, the Euler equation for durable consumption (1.6), and consequently, the intratemporal FOC (1.15) may not hold due to frictions in the adjustment of durable consumption.

1.2.4 A Log-Linear Approximation of the SDF

I now introduce a linear approximation to the log of the SDF (1.14), which is convenient for transforming the asset pricing equation (1.8) into a linear factor model. Let lowercase letters denote the logs of the corresponding uppercase variables. Taking the log of both sides of (1.14) and approximating around the special case of Cobb-Douglas intraperiod utility (i.e. $\rho = 0$),

$$m_t \approx \log \beta - \gamma \Delta c_t + \alpha(1 - \gamma - \rho)(\Delta d_t - \Delta c_t). \tag{1.16}$$

The approximation is exact when $\rho = 0$. Let $\overline{r} = -\log \beta$, $f_t = (\Delta c_t, \Delta d_t)'$, and

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \gamma + \alpha(1 - \gamma - \rho) \\ -\alpha(1 - \gamma - \rho) \end{bmatrix}. \tag{1.17}$$

Then equation (1.16) can be written more compactly as

$$-m_t \approx \overline{r} + b' f_t = \overline{r} + b_1 \Delta c_t + b_2 \Delta d_t. \tag{1.18}$$

[2]Ogaki and Reinhart (1998) make a similar argument to motivate their empirical methodology.

$\overline{r}$ is the rate of time preference. b_1 and b_2 are the risk prices for the two risk factors, nondurable and durable consumption growth, respectively.

The risk price of durable consumption is positive when $\rho > 1 - \gamma$, that is when the elasticity of substitution between the two consumption goods is greater than the inverse of risk aversion. Using quarterly data in the sample 1951:1–1983:4, Ogaki and Reinhart (1998, Table 2) estimate a 95% confidence interval of $[-0.03, 0.27]$ for ρ. On the other hand, the literature on the equity premium puzzle suggests that γ is large (see Campbell (2003) for a survey). Therefore, the assumption of additive separability is not supported by the data. Moreover, durable consumption is potentially an important risk factor that carries a large positive risk price.

An intuitive way to think about the risk prices for nondurable and durable consumption is the approximation

$$\begin{aligned} b_1 &= (1-\alpha)\gamma + \alpha(1-\rho) \approx (1-\alpha)\gamma, \\ b_2 &= \alpha\gamma - \alpha(1-\rho) \approx \alpha\gamma. \end{aligned}$$

The approximation holds in the empirically relevant case where $\rho \approx 0$ and γ is large. The sum of the risk prices is total risk aversion γ. The fraction of risk attributed to nondurable consumption is $1 - \alpha$, which is the budget share of nondurable consumption under Cobb-Douglas intraperiod utility.

1.3 Linear Factor Models

Taking the unconditional expectation of equation (1.9),

$$\mathbf{E}[M_t(R_{it} - R_{0t})] = 0. \tag{1.19}$$

Suppose the SDF is linear in a vector f_t of F underlying factors, that is

$$-\frac{M_t}{\mathbf{E}[M_t]} = k + b' f_t.$$

Let $\mu_f = \mathbf{E}[f_t]$, $\Sigma_{ff} = \mathbf{E}[(f_t - \mu_f)(f_t - \mu_f)']$, and $\Sigma_{fi} = \mathbf{E}[(f_t - \mu_f)(R_{it} - R_{0t})]$. Equation (1.19) can then be written as a linear factor model

$$\mathbf{E}[R_{it} - R_{0t}] = b' \Sigma_{fi}. \tag{1.20}$$

This equation says that the premium on asset i is the price of risk b times its quantity of risk Σ_{fi}.

Define the "beta" of asset i as $\beta_i = \Sigma_{ff}^{-1} \Sigma_{fi}$, which can be interpreted as the coefficient vector in a multiple regression of R_{it} onto f_t. The linear factor model can be written as a beta pricing model

$$\mathbf{E}[R_{it} - R_{0t}] = \lambda' \beta_i, \tag{1.21}$$

where $\lambda = \Sigma_{ff} b$ is the factor risk premium.

1.3.1 Fama-French Three-Factor Model

In response to the failures of the CAPM and the CCAPM, Fama and French (1993) proposed an influential three-factor model. The three factors are excess returns on the market portfolio, returns on the SMB (Small Minus Big stocks) portfolio, and returns on the HML (High Minus Low book-to-market stocks) portfolio. The Fama-French three-factor model nests the static CAPM (Sharpe 1964, Lintner 1965) as a special case where the risk prices for SMB and HML are restricted to zero.

Although the model is an empirical success, it falls short of a satisfactory understanding of the underlying risk reflected in stock returns. "Without a theory that specifies the exact form of the state variables or common factors in returns, the choice of any particular version of the factors is somewhat arbitrary." (Fama and French 1993, p. 53) As emphasized by Cochrane (2001, Chapter 9), a satisfactory factor model must ultimately connect the factors to the marginal utility of consumption.

1.3.2 Consumption-Based Model

A nonlinear SDF, M_t, can be approximated by first-order log-linear approximation as

$$\frac{M_t}{\mathbf{E}[M_t]} \approx 1 + m_t - \mathbf{E}[m_t].$$

Using equation (1.18), the SDF (1.14) of the durable consumption model can be approximated as

$$-\frac{M_t}{\mathbf{E}[M_t]} \approx k + b_1 \Delta c_t + b_2 \Delta d_t, \tag{1.22}$$

where $k = -1 - b_1 \mathbf{E}[\Delta c_t] - b_2 \mathbf{E}[\Delta d_t]$. The corresponding linear factor model (1.20) is

$$\mathbf{E}[R_{it} - R_{0t}] = b_1 \mathrm{Cov}(\Delta c_t, R_{it} - R_{0t}) + b_2 \mathrm{Cov}(\Delta d_t, R_{it} - R_{0t}). \tag{1.23}$$

When $\rho = 1 - \gamma$ (i.e. additive separability), this equation reduces to

$$\mathbf{E}[R_{it} - R_{0t}] = \gamma \mathrm{Cov}(\Delta c_t, R_{it} - R_{0t}), \tag{1.24}$$

which is the familiar CCAPM (Rubinstein 1976, Breeden and Litzenberger 1978, Breeden 1979).

Equation (1.23) says that an asset with high nondurable consumption beta,

$\text{Cov}(\Delta c_t, R_{it} - R_{0t})/\text{Var}(\Delta c_t)$, must have high expected returns. Likewise, an asset with high durable consumption beta, $\text{Cov}(\Delta d_t, R_{it} - R_{0t})/\text{Var}(\Delta d_t)$, must have high expected returns when $b_2 > 0$. In equilibrium, differences in expected returns across assets must reflect differences in the quantity of risk across assets, measured by the covariance of returns with nondurable and durable consumption growth.

1.3.3 An Approximation of the Intratemporal FOC

Suppose the intraperiod utility is Cobb-Douglas (i.e. $\rho = 0$). Substituting the linearized SDF (1.22) in the intratemporal FOC (1.15) and taking the expectation of both sides of the equation yields

$$\begin{aligned}\frac{b_2}{b_1 - 1}\mathbf{E}\left[\frac{C_{t-1}}{P_{t-1}D_{t-1}}\right] &= 1 - a\left[\mathbf{E}\left[\frac{P_t}{P_{t-1}}\right]\right. \\ &\qquad \left. -b_1\text{Cov}\left(\Delta c_t, \frac{P_t}{P_{t-1}}\right) - b_2\text{Cov}\left(\Delta d_t, \frac{P_t}{P_{t-1}}\right)\right], \quad (1.25)\end{aligned}$$

where $a = (1-\delta)\mathbf{E}[M_t]$. Note that the parameters in this equation are the risk prices b_1 and b_2, rather than the preference parameters γ, ρ, and α. Hence, the equation is a useful way of imposing the intratemporal FOC in estimating the linear factor model (1.23).

1.3.4 GMM Estimation of Linear Factor Models

Since the linear factor model is a set of moment restrictions on asset returns, GMM is a natural way to estimate and test the model.[3] Since my focus is on consumption-based models, I base estimation on the covariance representation (1.20), rather than the beta representation (1.21) of the model. The coefficients b of the covariance representation

[3] See Cochrane (2001, Chapter 13) for a textbook treatment of GMM for linear factor models.

are immediately interpretable as preference parameters, unlike the coefficients λ of the beta representation. In Appendix A.1, I relate the GMM estimator to an estimator of the risk prices based on a cross-sectional regression.

Define the parameter space $\Theta \subset \mathbb{R}^{2F}$ with a generic element $\theta = (b', \mu_f')'$. Let R_{0t}, $R_t = (R_{1t}, \ldots, R_{Nt})'$, and f_t be the time t observation on the reference return, the vector of N test asset returns, and the vector of F factors, respectively. Stack the variables in a vector as $z_t = (R_{0t}, R_t', f_t')'$. Let ι be an $N \times 1$ vector of ones. Consider the $(N + F) \times 1$ moment function

$$e(z_t, \theta) = \begin{bmatrix} R_t - R_{0t}\iota - (R_t - R_{0t}\iota)(f_t - \mu_f)'b \\ f_t - \mu_f \end{bmatrix}. \tag{1.26}$$

The moment function satisfies the moment restriction $\mathbf{E}[e(z_t, \theta_0)] = 0$, for some $\theta_0 \in \Theta$, through equation (1.20). A necessary condition for identification is that $N \geq F$. A sufficient condition for identification is that the $F \times N$ matrix $[\Sigma_{f1} \cdots \Sigma_{fN}]$ has rank F. This condition assures that θ_0 is a unique solution to $\mathbf{E}[e(z_t, \theta)] = 0$, so that the key identification condition for GMM is satisfied (see Wooldridge (1994, Theorem 7.1)). Intuitively, the factors cannot be perfectly correlated in order for the factor risk prices to be identified.

The overidentifying restrictions of the model can be tested by Hansen's (1982) J-test. The degree of overidentification is $N - F$, or $N - F + 1$ for the durable consumption model when the intratemporal FOC (1.25) is imposed as an additional moment restriction. The J-test tests the null hypothesis that the pricing errors are jointly zero across the N test assets. The test is conceptually similar to the GRS test (Gibbons, Ross, and Shanken 1989) since the test statistic is a quadratic form in the vector of pricing errors (see Cochrane (2001, Chapters 12–13)).

1.4 Conditional Factor Model

Let R_{0t} be a conditionally riskfree return, so that $R_{0t} = 1/\mathbf{E}_{t-1}[M_t]$. (In practice, R_{0t} is a reference return over which excess returns are computed, such as the 90-day T-bill return.) Let lowercase letters denote the logs of the corresponding uppercase variables (e.g. $m_t = \log M_t$ and $r_{0t} = \log R_{0t}$). For assets $i = 1, \ldots, N$, however, let $r_{it} = \log R_{it} - \log R_{0t}$ be the log return in excess of the log riskfree rate. By a second-order log-linear approximation of equation (1.8) for the riskfree rate (see Campbell (2003)),

$$r_{0t} = -\mathbf{E}_{t-1} m_t - \frac{1}{2}\mathrm{Var}_{t-1}(m_t). \tag{1.27}$$

Similarly, the excess return on asset i can be approximated as

$$\mathbf{E}_{t-1}[r_{it}] + \frac{1}{2}\mathrm{Var}_{t-1}(r_{it}) = -\mathrm{Cov}_{t-1}(m_t, r_{it}). \tag{1.28}$$

Suppose the log SDF is linear in a vector f_t of F underlying factors, that is

$$-m_t = \overline{r} + b' f_t = \overline{r} + \sum_{j=1}^{F} b_j f_{jt}.$$

Then equation (1.27) becomes

$$r_{0t} = \overline{r} + b'\mathbf{E}_{t-1}[f_t] - \frac{1}{2}\mathrm{Var}_{t-1}(b' f_t), \tag{1.29}$$

and equation (1.28) can be written as a conditional factor model

$$\mathbf{E}_{t-1}[r_{it}] + \frac{1}{2}\mathrm{Var}_{t-1}(r_{it}) = \sum_{j=1}^{F} b_j \mathrm{Cov}_{t-1}(f_{jt}, r_{it}). \tag{1.30}$$

This equation says that the premium on an asset is the price of risk b_j times the quantity of risk $\mathrm{Cov}_{t-1}(f_{jt}, r_{it})$, summed over all factors $j = 1, \ldots, F$.

1.4.1 Consumption-Based Model

Using the linear approximation to the log SDF (1.18), the riskfree rate for the durable consumption model is

$$\begin{aligned} r_{0t} &= \overline{r} + b_1 \mathbf{E}_{t-1}[\Delta c_t] + b_2 \mathbf{E}_{t-1}[\Delta d_t] \\ &\quad - \frac{b_1^2}{2}\mathrm{Var}_{t-1}(\Delta c_t) - \frac{b_2^2}{2}\mathrm{Var}_{t-1}(\Delta d_t) - b_1 b_2 \mathrm{Cov}_{t-1}(\Delta c_t, \Delta d_t). \end{aligned} \tag{1.31}$$

The premium on asset i is

$$\mathbf{E}_{t-1}[r_{it}] + \frac{1}{2}\mathrm{Var}_{t-1}(r_{it}) = b_1 \mathrm{Cov}_{t-1}(\Delta c_t, r_{it}) + b_2 \mathrm{Cov}_{t-1}(\Delta d_t, r_{it}). \tag{1.32}$$

When $\rho = 1 - \gamma$ (i.e. additive separability), this equation reduces to

$$\mathbf{E}_{t-1}[r_{it}] + \frac{1}{2}\mathrm{Var}_{t-1}(r_{it}) = \gamma \mathrm{Cov}_{t-1}(\Delta c_t, r_{it}), \tag{1.33}$$

which is the familiar CCAPM.

Equation (1.32) says that the expected return on an asset is high when the covariance of its returns with nondurable consumption growth is high. Likewise, the expected return is high when the covariance of its returns with durable consumption growth is high, provided that $b_2 > 0$. In equilibrium, variation in expected returns through time must reflect variation in the quantity of risk through time, measured by the conditional covariance of returns with nondurable and durable consumption growth.

1.4.2 Estimation of Conditional Moments Using Instruments

I now describe a way to estimate the conditional moments of the conditional factor model (1.30), using a vector x_{t-1} of I instrumental variables known at time $t-1$. The essential idea behind the method is that the representative household's information set can always be conditioned down to the econometrician's information set. Equation (1.30) therefore holds even when the conditioning information is restricted to x_{t-1}. The methodology described here has been used previously in empirical work by Campbell (1987) and Harvey (1989).

Consider the linear regression model

$$r_{it} = \Pi_i' x_{t-1} + u_{it} \quad (i = 1, \ldots, N), \tag{1.34}$$

$$u_{it} r_{it} = \Gamma_i' x_{t-1} + \epsilon_{it} \quad (i = 1, \ldots, N), \tag{1.35}$$

$$u_{it} f_{jt} = \Upsilon_{ij}' x_{t-1} + \eta_{ijt} \quad (i = 1, \ldots, N; j = 1, \ldots, F). \tag{1.36}$$

Equations (1.34) and (1.35) model the conditional mean and variance of excess log returns, respectively. Equation (1.36) models the conditional covariance of excess log returns with the factors. The model (1.34)–(1.36) is exactly identified under the conditional moment restriction

$$\mathbf{E}[(u_{it}, \epsilon_{it}, \eta_{ijt})' | x_{t-1}] = 0 \quad \forall i, j. \tag{1.37}$$

Define the matrices

$$
\begin{aligned}
\Pi = &\quad [\Pi_1 \cdots \Pi_N] \quad (I \times N), \\
\Gamma = &\quad [\Gamma_1 \cdots \Gamma_N] \quad (I \times N), \\
\Upsilon_j = &\quad [\Upsilon_{1j} \cdots \Upsilon_{Nj}] \quad (I \times N), \\
\Upsilon = &\quad [\Upsilon_1 \cdots \Upsilon_F] \quad (I \times NF).
\end{aligned}
$$

The conditional factor model (1.30) implies NI linear restrictions of the form

$$
\Pi + \frac{1}{2}\Gamma = \sum_{j=1}^{F} b_j \Upsilon_j. \tag{1.38}
$$

Using this equation to substitute out Γ_i in equation (1.35),

$$
u_{it} r_{it} = 2\left(\sum_{j=1}^{F} b_j \Upsilon_{ij} - \Pi_i\right)' x_{t-1} + \epsilon_{it}. \tag{1.39}
$$

Assuming that the vector of risk prices b is known, the model (1.34), (1.39), and (1.36) is overidentified by NI degrees.

Define the parameter space $\Theta \subset \mathbb{R}^{(N+NF)I}$ with a generic element $\theta = (\text{vec}(\Pi)', \text{vec}(\Upsilon)')'$. Let $r_t = (r_{1t}, \ldots, r_{Nt})'$ and f_t be the time t observation on the vector of N excess log returns and the vector of F factors, respectively. Stack the variables and the instruments in a vector as $z_t = (r_t', f_t', x_{t-1})'$. Consider the $(2N + NF)I \times 1$ moment function

$$
e(z_t, \theta; b) = \begin{bmatrix} r_t - \Pi' x_{t-1} \\ \text{diag}((r_t - \Pi' x_{t-1}) r_t') - 2(\sum_{j=1}^{F} b_j \Upsilon_j - \Pi)' x_{t-1} \\ \text{vec}((r_t - \Pi' x_{t-1}) f_t') - \Upsilon' x_{t-1} \end{bmatrix} \otimes x_{t-1}. \tag{1.40}
$$

The moment function satisfies the moment restriction $\mathbf{E}[e(z_t, \theta_0; b)] = 0$, for some $\theta_0 \in \Theta$, through the conditional moment restriction (1.37).

In practice, the vector of risk prices b is not known. It can be estimated jointly with θ using the moment function (1.40), provided that $F \leq NI$. Instead, suppose there is a consistent estimator $\widehat{b}$. Then the GMM estimator for θ based on the moment function $e(z_t, \theta; \widehat{b})$ is consistent and has the same asymptotic distribution as if b were known. This can be verified by checking the sufficient conditions for consistency and asymptotic normality in Newey and McFadden (1994, Theorems 2.1 and 3.2). In the empirical work, I use the estimated preference parameters from GMM estimation of the conditional Euler equation (1.9) to obtain $\widehat{b}$, through equation (1.17). I then estimate θ using the moment function $e(z_t, \theta; \widehat{b})$. This estimation strategy is consistent with the purpose of estimating the conditional factor model (1.30), which is to better understand the dynamics of expected returns implied by asset pricing equation (1.9).

1.5 Consumption Data

1.5.1 Source and Construction

Quarterly consumption data is from the US national accounts. Following convention, nondurable consumption is measured as the sum of real personal consumption expenditures (PCE) on nondurable goods and services.[4] Nondurable consumption includes food, clothing and shoes, housing, utilities, transportation, and medical care. Items such as clothing and shoes are durable at quarterly frequency, but I include them as part of nondurable consumption to be consistent with previous studies of the CCAPM. Similarly, housing is the service flow imputed from the rental value of

[4]See Whelan (2000) for issues concerning aggregation of chained national accounts data.

houses.

Durable consumption consists of items such as motor vehicles, furniture and appliances, and jewelry and watches. The Bureau of Economic Analysis (BEA) publishes year-end estimates of the chained quantity index for the net stock of consumer durable goods. Using quarterly data for real PCE on durable goods, I construct quarterly series for the stock of durables by equation (1.1). Implicit in the data for the stock of durables are the depreciation rates used by the BEA for various components of durable goods. The implied depreciation rate for durable goods as a whole is about 6% per quarter.

Both nondurable consumption and the stock of durables are divided by the population. In matching consumption to returns data, I use "beginning of the period" timing convention, following Campbell (2003). In other words, the consumption data for each quarter is assumed to be the flow on the first, rather than the last, day of the quarter. Although quarterly consumption data is available since 1947, the period immediately after the war experienced unusually high durable consumption growth due to the rapid restocking of durable goods. I therefore use data since 1951, following Ogaki and Reinhart (1998). The resulting sample period is 1951:1–2001:4.

1.5.2 Basic Description

Figure 1.2 is a time series plot of the ratio of the stock of durables to nondurable consumption, that is D/C. The series has an upward trend in the postwar sample, which is consistent with the downward trend in the price of durables relative to nondurables. The shaded regions are recessions, from peak to trough, as defined by the National Bureau of Economic Research (NBER). The ratio D/C rises during booms and falls during recessions, implying strong counter-cyclical movements in

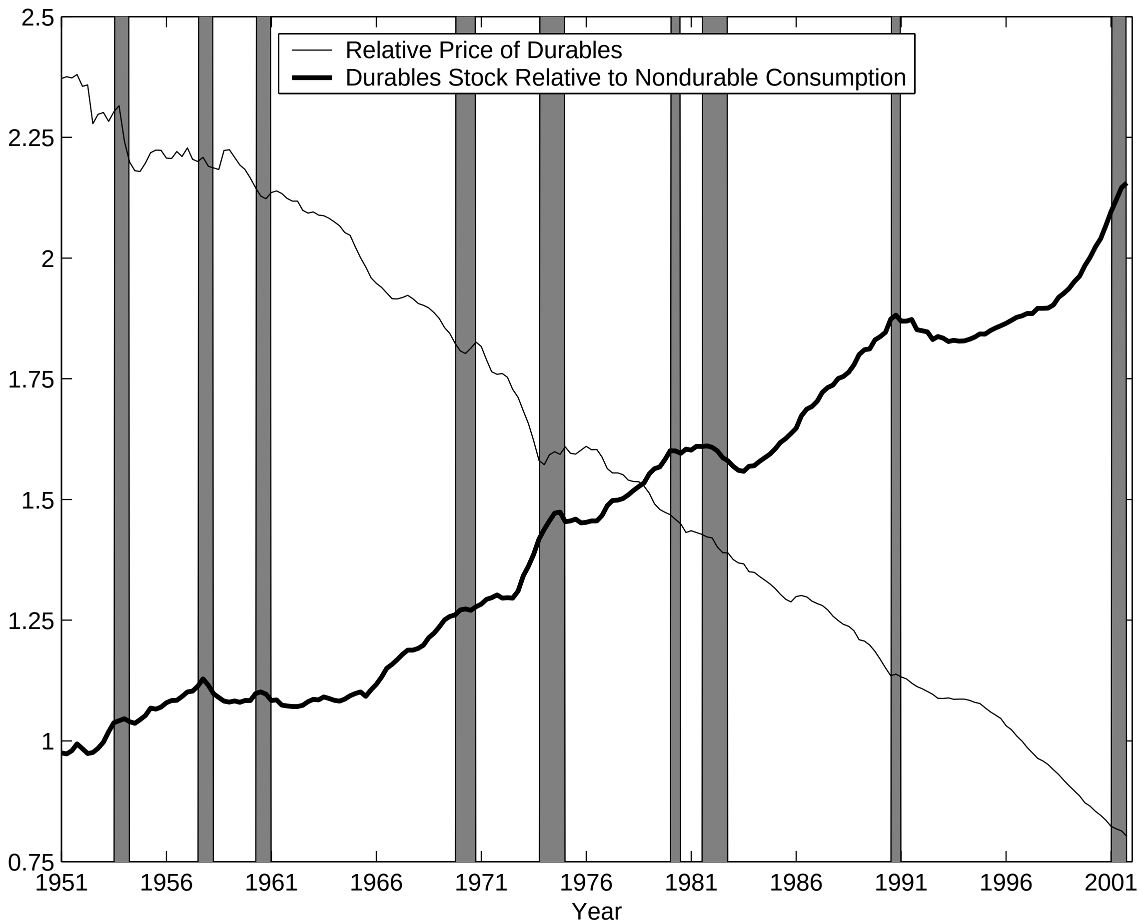

Figure 1.2: Price and Stock of Durables Relative to Nondurables

The figure is a time series plot of (1) the price of durables as a ratio of the price of nondurables and (2) the real stock of durables as a ratio of real nondurable consumption. The sample period is 1951:1–2001:4, and the shaded regions are NBER recessions.

Table 1.1: Descriptive Statistics

The table reports the mean, the standard deviation, and the first-order autocorrelation of the excess market return, the SMB return, the HML return, and nondurable and durable consumption growth. It also reports the correlation between these variables.

Variable	Mean (%)	Std Dev (%)	Autocorr	Correlation			
				Market	SMB	HML	Nondurables
Market	1.880	8.186	0.048				
SMB	0.508	5.580	-0.034	0.423			
HML	1.089	5.543	0.154	-0.386	-0.143		
Nondurables	0.513	0.542	0.282	0.281	0.130	0.004	
Durables	0.915	0.535	0.875	-0.110	-0.038	0.036	0.192

marginal utility (1.13), provided that the elasticity of substitution between the goods is high.

Table 1.1 reports descriptive statistics for nondurable and durable consumption growth, together with those for the three Fama-French factors. (Recall that the growth rate in the stock is the growth rate in the consumption of durable goods.) Nondurable consumption growth has mean 0.51% and standard deviation 0.54% per quarter. Durable consumption has mean 0.92% and standard deviation 0.54%. The correlation between them is 0.19. The Fama-French factors have low correlation with the two consumption-based factors, especially with durable consumption growth. Durable consumption growth is much more persistent than nondurable consumption growth. The first-order autocorrelations are 0.88 and 0.28, respectively.

1.5.3 Business-Cycle Properties

Figure 1.3(a) is a time series plot of the growth rates of nondurable and durable consumption in the postwar sample. Durable consumption growth is strongly pro-cyclical, peaking during booms and bottoming out during recessions. It is therefore a good indicator variable for the business cycle. Nondurable consumption growth is also pro-cyclical, but less so than durable consumption. It tends to fall sharply right at the onset of recessions. Figure 1.3(b) is a time series plot of nondurable consumption growth minus durable consumption growth. The growth rate of durable consumption generally exceeds that of nondurable consumption, except during and immediately after recessions. The series is strongly counter-cyclical, highest at business-cycle troughs and lowest at business-cycle peaks.

To examine the cyclical properties of nondurable consumption in further detail, Figure 1.4(a) shows the time series for nondurable consumption growth together with the growth rates of two of its components: (1) food and (2) housing. (At the end of 2001, food accounted for 16% and housing 17% of consumption expenditures on nondurables.) The figure illustrates the fact that the components of nondurable consumption share the time series properties of its aggregate: low volatility (compared to stock returns), low autocorrelation, and weak cyclicality. Although houses can be thought of as a "durable good", its service flows are more similar to that of "nondurable goods".

Implicit in studies of the CCAPM is the assumption that the various components of nondurable consumption are perfect substitutes. This appears to be a reasonable assumption for the purposes of empirical work since the various components share similar time series properties. Moreover, the gain from explicitly modeling non-separability between the various components of nondurable consumption appears to

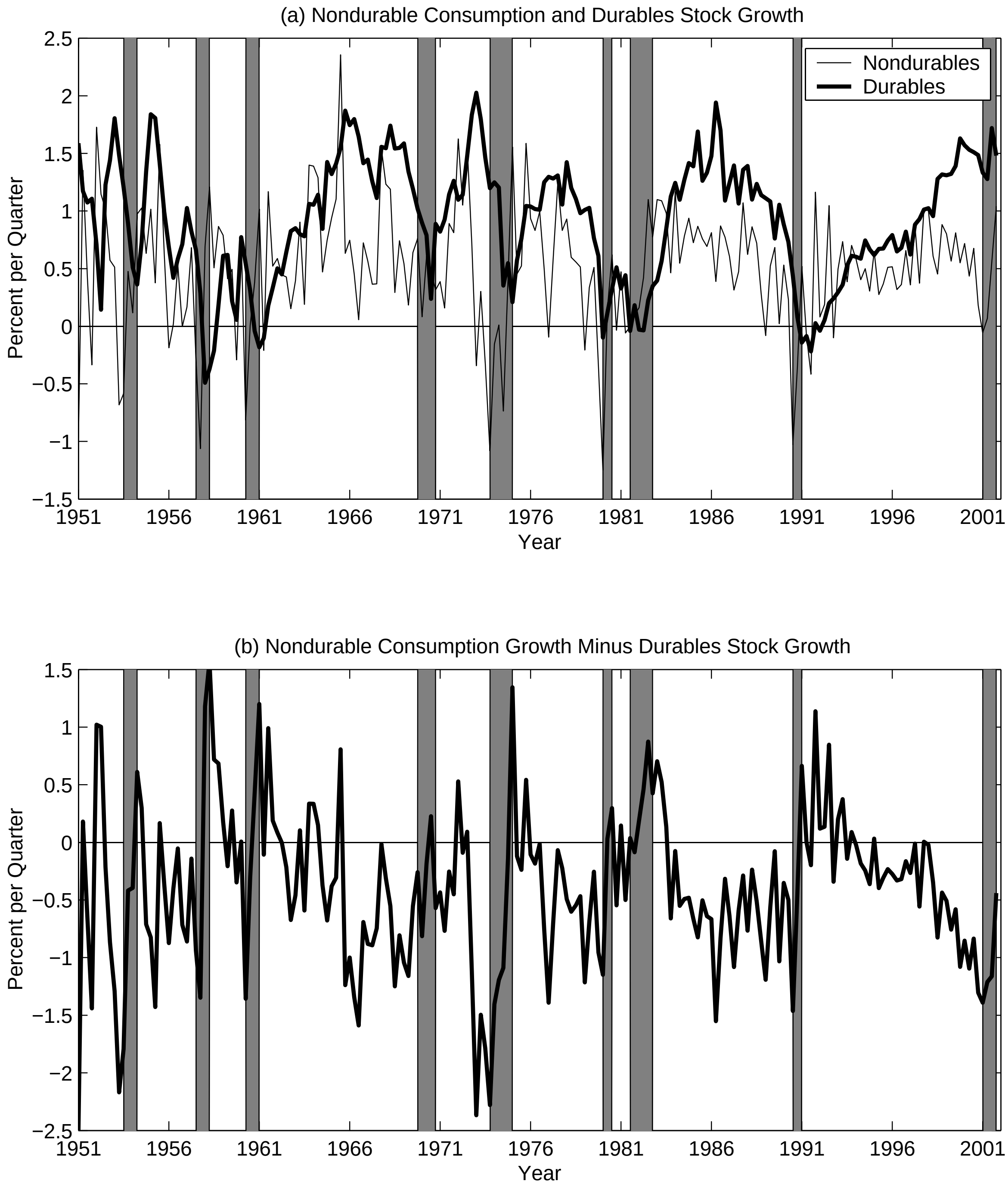

Figure 1.3: Nondurable and Durable Consumption Growth

The figure is a time series plot of (a) the real growth rates of nondurable consumption and the stock of durables and (b) the difference in the growth rates. See notes to Figure 1.2.

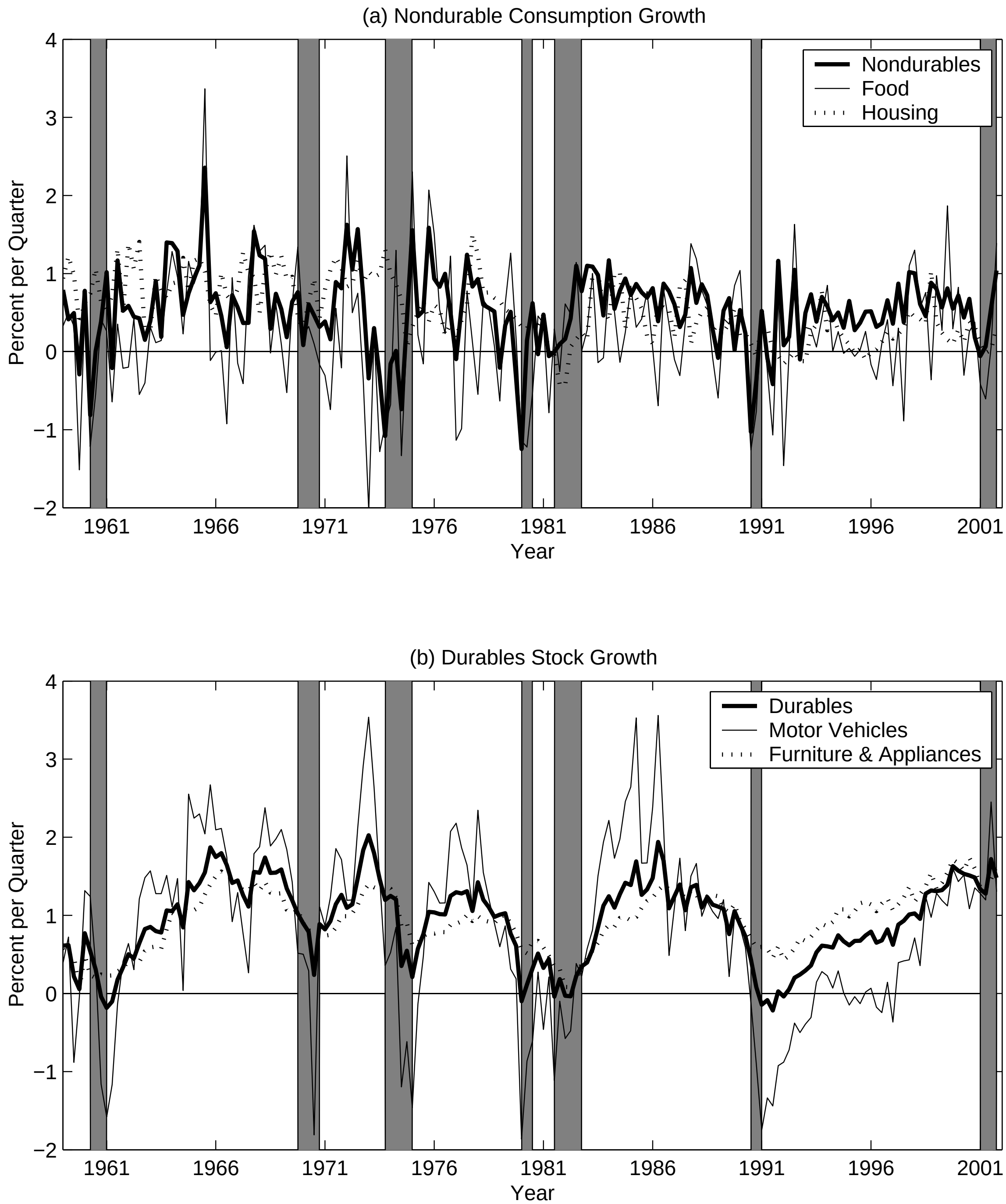

Figure 1.4: Components of Nondurable and Durable Consumption Growth

The figure is a time series plot of (a) the real growth rates of nondurable, food, and housing consumption and (b) the real growth rates of the stock of durables, motor vehicles, and furniture and appliances. The sample period is 1959:1–2001:4, and the shaded regions are NBER recessions.

be small, at least for the purposes of asset pricing. For instance, Piazessi, Schneider, and Tuzel (2003) find that a model that accounts for non-separability between housing and other nondurable goods cannot reconcile the size and value premia.[5]

Figure 1.4(b) is a time series plot of durable consumption growth together with the growth rates of two of its components: (1) motor vehicles and (2) furniture and appliances. (At the end of 2001, motor vehicles accounted for 30% and furniture and appliances 45% of the stock of consumer durables.) The figure illustrates the fact that the components of durable consumption share the time series properties of its aggregate: low volatility (compared to stock returns), high autocorrelation, and strong cyclicality. The consumption of motor vehicles is especially pro-cyclical with sharp falls during recessions. The strong cyclicality of durable consumption is consistent with that of luxury goods (Aït-Sahalia, Parker, and Yogo 2003).

1.6 Cross-Sectional Tests

In this section, I test the cross-sectional implications of the durable consumption model. The test assets are the 25 Fama-French portfolios (Section 1.6.1), portfolios sorted by book-to-market equity within industry (Section 1.6.2), and portfolios sorted by market and HML betas (Section 1.6.3). The empirical results focus on the linear two-factor model (1.23), rather than the nonlinear model (1.19) with SDF (1.14). The main advantage of the linear model is that it makes transparent the central economic finding, that small stocks and value stocks are pro-cyclical. It also makes the results readily comparable to the large literature on cross-sectional asset pricing, which has

[5]The test that the pricing errors are jointly zero for the 25 Fama-French portfolios rejects the model. However, Lustig and Van Nieuwerburgh (2002) argue that housing has an important role as collateral in risk-sharing markets. Using the housing-human wealth ratio as a conditioning variable, they find that the conditional CCAPM can explain the size and value premia. Their finding confirms that of Lettau and Ludvigson (2001), who use the consumption-wealth ratio as a conditioning variable.

focused on linear factor models. In Section 1.6.4, I estimate the nonlinear model to support the empirical findings for the linear model.

1.6.1 Fama-French Portfolios

Data

Fama and French (1993) construct 25 portfolios by independently sorting stocks into quintiles based on size (i.e. market equity) and book-to-market equity. Data on the Fama-French factors and portfolio returns were obtained from Professor Kenneth French's webpage. Excess returns are computed by subtracting the 90-day T-bill return, which is from the Center for Research in Security Prices (CRSP) Indices database. Because of the failures of the CAPM and the CCAPM in explaining their returns, the Fama-French portfolios have been the focus of recent work on cross-sectional asset pricing (e.g. Lettau and Ludvigson (2001), Campbell and Vuolteenaho (2002), and Parker and Julliard (2003)).

Test of Linear Factor Models

Table 1.2 reports estimates of the factor risk prices for the CAPM, the Fama-French three-factor model, the CCAPM, and the durable consumption model. Estimation is by two-step (efficient) GMM. Standard errors are heteroskedasticity and autocorrelation consistent (HAC), computed by the VARHAC procedure with automatic lag length selection by AIC (see den Haan and Levin (1997)).[6] The maximum lag length is set to three quarters to account for autocorrelation. The correction for autocorrelation is especially important in estimating the durable consumption model due to

[6] den Haan and Levin (2000) find that the VARHAC covariance matrix estimator performs better than the kernel-based estimators (e.g. Newey and West (1987) and Andrews (1991)) in various Monte Carlo setups.

Table 1.2: Estimation of Linear Factor Models with the Fama-French Portfolios

The table reports the estimated factor risk prices for the CAPM, the Fama-French three-factor model, the CCAPM, and the durable consumption model. It reports two estimates of the durable consumption model, with and without the intratemporal FOC as an additional moment restriction. The test assets are the 25 Fama-French portfolios sorted by size and book-to-market equity. Estimation is by two-step GMM. HAC standard errors in parentheses. The mean absolute pricing error (MAE) and R^2 are based on the first-stage estimate. The p-value for the J-test (test of overidentifying restrictions) in parentheses.

Factor Price	CAPM	Fama-French	CCAPM	Durable Model	
				No FOC	FOC
Market	2.659	4.319			
	(0.829)	(0.983)			
SMB		-0.621			
		(1.274)			
HML		6.225			
		(1.323)			
Nondurables			105.619	122.345	148.855
			(23.555)	(22.084)	(16.417)
Durables				197.139	203.264
				(39.342)	(40.498)
γ				319.484	352.119
				(46.222)	(48.000)
α (if $\rho = 0$)				0.619	0.579
				(0.062)	(0.048)
MAE (%)	0.654	0.257	0.329	0.198	0.192
R^2	-0.892	0.658	0.382	0.770	0.773
J-test	62.998	51.503	52.475	36.475	43.386
	(0.000)	(0.000)	(0.001)	(0.037)	(0.009)

the persistence of durable consumption growth.

The CAPM has a positive and significant risk price on the market return. The mean absolute pricing error from the first stage is 0.65% per quarter. Instead of reporting the mean squared pricing error, I report one minus its ratio to the variance of average portfolio returns, which is called the R^2, following Campbell and Vuolteenaho (2002). The R^2 for the CAPM is -89%, which suggests that the model fits the average T-bill return very poorly. The J-test, or the test of overidentifying restrictions, strongly rejects the model.

The Fama-French three-factor model is much more successful than the CAPM. The mean absolute pricing error is 0.26%, and the R^2 is 66%. The risk price for SMB is not significantly different from zero, while the risk price for HML is significantly positive. Hence, the improvement over the CAPM is mostly captured by the explanatory power of HML. Although the first-stage measures of fit are much better than the CAPM, the J-test rejects the model.

For the CCAPM, the risk price for nondurable consumption is positive and significantly different from zero. The large point estimate of 106, which is a consequence of the low volatility of nondurable consumption, is consistent with the literature on the equity premium puzzle. The mean absolute pricing error is 0.33%, and the R^2 is 38%. Although the CCAPM has better first-stage measures of fit than the CAPM, it falls short of the three-factor model. Moreover, the J-test strongly rejects the model.

In the last two columns of Table 1.2, I report two estimates of the durable consumption model. The first estimate is based only on the moment restrictions used to price the portfolios. The second estimate imposes an additional moment restriction corresponding to the intratemporal FOC (1.25). In other words, the second estimate forces the model to simultaneously explain the returns on the 25 Fama-French portfolios and the optimal consumption behavior implied by the FOC. In estimating equation (1.25), I set $a = 0.94$ since the depreciation rate is about 6% per quarter; the results are not sensitive to reasonable variations in a.

Without the intratemporal FOC, the risk price for nondurable consumption is comparable to that estimated for the CCAPM, with a point estimate of 122. The risk price for durable consumption is larger at 197 and statistically significant. Therefore, the CCAPM, which is a restriction that the risk price on durable consumption be equal to zero, is strongly rejected. Recall that the sum of the risk prices for nondurable and durable consumption is the risk aversion γ. The point estimate of γ is 319, which is a consequence of the low volatility of both nondurable and durable consumption. The model therefore does not resolve the equity premium puzzle. Assuming Cobb-Douglas intraperiod utility (i.e. $\rho = 0$), the point estimate of $\alpha = b_2/(b_1 + b_2 - 1)$ is 0.62. The mean absolute pricing error is 0.20%, and the R^2 is 77%. Although the J-test rejects at the 5% level, the rejection is solely due to the model's inability to price the small growth portfolio, as discussed below. The results are essentially the same when the intratemporal FOC is imposed.

Figure 1.5(d) provides a visual summary of the empirical success of the durable consumption model. On the vertical axis is the realized average excess return. On the horizontal axis is the return predicted by the model, based on the first-stage estimates. The points represent the 25 Fama-French portfolios, and the corresponding vertical distance to the diagonal line represents the pricing error. The pricing errors for the durable consumption model are much smaller than those for (a) the CAPM and (c) the CCAPM. It even outperforms (b) the Fama-French three-factor model.

Estimation Without the Small Growth Portfolio

Figure 1.5 reveals the small growth portfolio (i.e. the lowest quintile in both size and book-to-market equity) is an outlier for all the linear factor models. For the durable

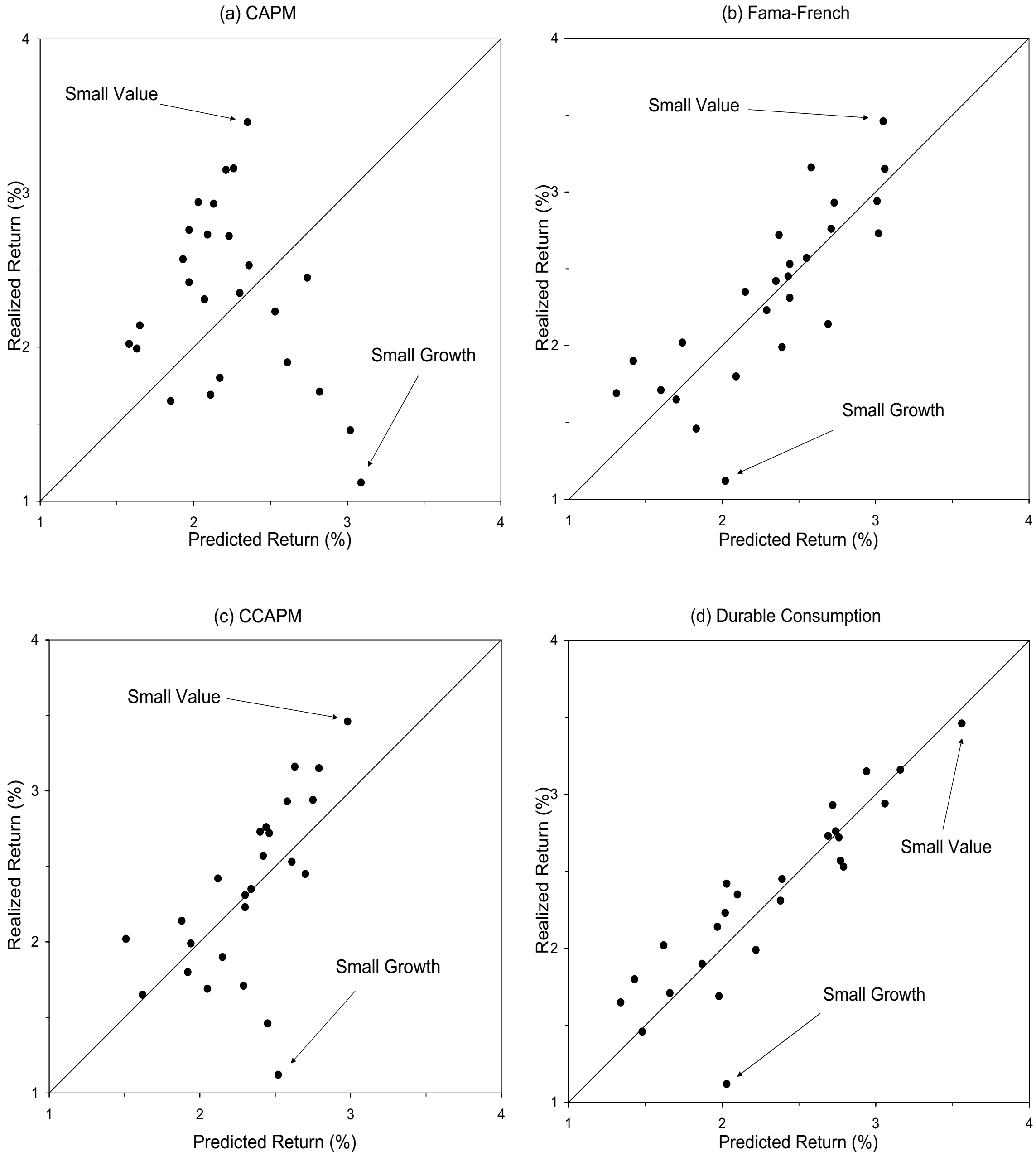

Figure 1.5: Realized vs. Predicted Returns for the Fama-French Portfolios.

The figure plots realized versus predicted excess returns (per quarter) for the 25 Fama-French portfolios sorted by size and book-to-market equity. The estimated models are (a) the CAPM, (b) the Fama-French three-factor model, (c) the CCAPM, and (d) the durable consumption model.

consumption model, its pricing error is nearly 1%. D'Avolio (2002) and Lamont and Thaler (2003) document limits to arbitrage, due to short-sale constraints, for the types of stocks that are generally characterized as small growth. It is perhaps unsurprising then that these frictionless equilibrium models have difficulty explaining the small growth portfolio.

In Table 1.3, I report estimates of the linear factor models using 24 of the Fama-French portfolios, excluding the small growth portfolio. The R^2 of the durable consumption model improves from 77% to 81%. In comparison, the R^2 of the Fama-French three-factor model improves from 66% to 74%. The J-test fails to reject the durable consumption model at the 5% level, both with and without the intratemporal FOC. The null hypothesis that the pricing errors are jointly zero is rejected for the other three models.

Consumption Betas

To better understand the success of the durable consumption model, Table 1.4 reports the nondurable and durable consumption betas implied by the first-stage GMM estimates. Panel A reports the average excess returns for the 25 Fama-French portfolios sorted by size and book-to-market equity. Reading down the columns of the panel, average returns decrease in size for a given book-to-market equity quintile. The only exception is for low book-to-market stocks, whose average returns roughly increase in size. Reading across the rows of the panel, average returns increase in book-to-market equity for a given size quintile. The table confirms the well-known size and value premia.

Panel B of the table reports the nondurable consumption betas. Reading down the columns of the panel, nondurable consumption beta decreases in size for a given book-to-market equity quintile. This pattern is broadly consistent with the size premium.

Table 1.3: Estimation of Linear Factor Models without the Small Growth Portfolio

The test assets are 24 of the Fama-French portfolios, excluding the small growth portfolio (i.e. smallest size and lowest book-to-market equity). See notes to Table 1.2.

Factor Price	CAPM	Fama-French	CCAPM	Durable Model	
				No FOC	FOC
Market	3.023	3.767			
	(0.781)	(1.004)			
SMB		-0.349			
		(1.292)			
HML		5.935			
		(1.337)			
Nondurables			138.188	158.887	160.937
			(26.764)	(23.256)	(17.003)
Durables				179.757	127.783
				(47.592)	(21.464)
γ				338.644	288.720
				(54.980)	(37.017)
α (if $\rho = 0$)				0.532	0.444
				(0.072)	(0.024)
MAE (%)	0.577	0.220	0.293	0.197	0.198
R^2	-0.693	0.740	0.485	0.805	0.805
J-test	51.952	43.025	38.217	32.291	12.600
	(0.001)	(0.003)	(0.024)	(0.073)	(0.960)

Table 1.4: Average Returns and Consumption Betas for the Fama-French Portfolios

Panel A reports average excess returns (per quarter) on the 25 Fama-French portfolios sorted by size and book-to-market equity. Panels B and C report nondurable and durable consumption betas, implied by the first-stage GMM estimate of the durable consumption model. The last row reports the difference between small and big stocks, and the last column reports the difference between high and low book-to-market stocks.

	Book-to-Market Equity					
Size	Low	2	3	4	High	High−Low
A. Average Excess Return (%)						
Small	1.121	2.448	2.531	3.160	3.464	2.343
2	1.458	2.225	2.716	2.929	3.150	1.692
3	1.707	2.345	2.313	2.756	2.937	1.230
4	1.896	1.797	2.417	2.568	2.725	0.829
Big	1.686	1.652	2.015	1.987	2.140	0.454
Small−Big	-0.565	0.796	0.516	1.173	1.324	
B. Nondurable Consumption Beta						
Small	6.425	6.635	6.386	6.309	7.149	0.724
2	6.164	5.621	5.940	6.209	6.726	0.561
3	5.709	5.693	5.601	5.883	6.660	0.951
4	5.302	4.692	5.105	5.863	5.780	0.477
Big	5.063	3.942	3.572	4.719	4.533	-0.530
Small−Big	1.362	2.693	2.814	1.590	2.616	
C. Durable Consumption Beta						
Small	-0.444	-0.030	0.675	1.253	1.396	1.840
2	-1.108	-0.044	0.869	0.668	0.710	1.818
3	-0.612	0.035	0.502	0.868	0.925	1.537
4	-0.083	-0.407	0.249	0.931	0.861	0.943
Big	0.204	-0.141	0.471	0.730	0.461	0.257
Small−Big	-0.649	0.111	0.204	0.523	0.935	

Reading across the rows of the panel, nondurable consumption beta also increases in book-to-market equity for a given size quintile. However, the variation in beta across book-to-market equity is relatively small compared to the variation across size. The difference in nondurable consumption beta between small and big stocks is at least 1.36 (for the lowest book-to-market quintile). On the other hand, the difference in beta between high and low book-to-market stocks is at most 0.95 (for size quintile 3). The relatively small variation in nondurable consumption beta across book-to-market equity explains why the CCAPM fails to explain the value premium.

Panel C of the table reports the durable consumption betas. Reading down the columns of the panel, durable consumption beta decreases in size for a given book-to-market equity quintile, with exception of low book-to-market stocks. This is consistent with the pattern in average returns across the size quintiles. Moreover, durable consumption beta increases in book-to-market equity for a given size quintile, explaining the value premium. The difference in durable consumption beta between high and low book-to-market stocks is in general larger than that difference between small and big stocks. For instance, the difference in beta between high and low book-to-market stocks is 1.54 for the median size quintile. On the other hand, the difference in beta between small and big stocks is only 0.20 for the median book-to-market equity quintile. Roughly speaking, durable consumption beta accounts for the variation in average returns across book-to-market equity (i.e. value premium), while nondurable consumption beta accounts for the variation in average returns across size (i.e. size premium).

1.6.2 Portfolios Sorted by Book-to-Market Equity within Industry

To examine the value premium in more detail, I now test the durable consumption model on portfolios sorted by book-to-market equity within industry. The question is whether value stocks, that is stocks with high book-to-market equity relative to other stocks in the same industry, have high consumption betas that account for their premia.

Portfolio Formation

The portfolios are formed using returns on ordinary common equity, traded in NYSE, AMEX, or Nasdaq, in the CRSP Monthly Stock database. In June of each year t, stocks are sorted into eight industries based on their two-digit SIC codes: (1) nondurables manufacturing, (2) durables manufacturing, (3) other manufacturing, (4) nondurables retail, (5) durables retail, (6) services, (7) finance, and (8) natural resource. Within each industry, stocks are then sorted into three levels of book-to-market equity using breakpoints of 30th and 70th percentiles, based on its value in December of $t-1$. Once the 24 portfolios are formed, their value-weighted returns are tracked from July of t through June of $t+1$.

The industry definitions are designed to create variation in book-to-market equity that is independent of nondurable and durable consumption; the corresponding SIC codes are in Table 1.5. The book equity data is a merge of historical data from Moody's Manuals (available from Professor French's webpage) and COMPUSTAT. I refer to Davis, Fama, and French (2000) for details on the computation of book equity.

Table 1.5: Industry Definitions

Industry	Two-Digit SIC Codes
Manufacturing:	
Nondurables	20–23, 26–28, 31
Durables	25, 36, 37, 39
Other	15–19, 24, 29, 30, 32–35, 38
Retail:	
Nondurables	51, 54, 56, 58, 59
Durables	50, 52, 53, 55, 57
Services	40–49, 70–99
Finance	60–69
Natural Resource	1–14

Test of Linear Factor Models

Table 1.6 reports estimates of linear factor models using the portfolios sorted by book-to-market equity within industry. For the durable consumption model without the intratemporal FOC, the point estimate of the risk price for durable consumption is 107, which is somewhat smaller than that estimated using the Fama-French portfolios. Since the risk price is significantly different from zero, the CCAPM is rejected. The R^2 for the model is 69%, compared to 58% for the Fama-French three-factor model. The J-test fails to reject the durable consumption model, while the three-factor model is rejected at the 10% level. When the intratemporal FOC is imposed, however, the J-test rejects the model. This is a rejection of the linear approximation to the FOC; the J-test fails to reject the nonlinear model, as shown below.

Consumption Betas

Panel A of Table 1.7 reports the average excess returns for the 24 portfolios. Reading across the rows of the panel, average returns increase in book-to-market equity for each industry. In all industries, the high book-to-market portfolio has higher average

Table 1.6: Estimation of Linear Factor Models with Portfolios Sorted by Book-to-Market Equity within Industry

The test assets are 24 portfolios sorted by book-to-market equity within industry. Portfolios are formed by first sorting stocks into 8 industries, then sorting into 3 levels of book-to-market equity (breakpoints of 30th and 70th percentiles) within each industry. See notes to Table 1.2.

Factor Price	CAPM	Fama-French	CCAPM	Durable Model	
				No FOC	FOC
Market	3.126	4.017			
	(0.802)	(1.027)			
SMB		0.505			
		(1.280)			
HML		5.149			
		(1.309)			
Nondurables			111.657	113.797	82.279
			(13.164)	(12.954)	(5.804)
Durables				107.148	130.188
				(17.947)	(12.264)
γ				220.945	212.466
				(21.524)	(16.669)
α (if $\rho = 0$)				0.487	0.616
				(0.052)	(0.017)
MAE (%)	0.624	0.354	0.424	0.313	0.340
R^2	-0.009	0.579	0.518	0.688	0.658
J-test	45.500	32.080	27.138	28.289	35.352
	(0.003)	(0.057)	(0.250)	(0.166)	(0.048)

Table 1.7: Average Returns and Consumption Betas for Portfolios Sorted by Book-to-Market Equity within Industry

Panel A reports average excess returns (per quarter) on 24 portfolios sorted by book-to-market equity within industry. Panels B and C report nondurable and durable consumption betas, implied by the first-stage GMM estimate of the durable consumption model. See notes to Table 1.6 for details on portfolio formation.

	A. Average Return (%)			B. Nondurable Beta			C. Durable Beta		
	Book-to-Market Equity								
Industry	Low	Med	High	Low	Med	High	Low	Med	High
Manufacturing:									
Nondurables	1.904	2.270	2.819	4.090	4.399	5.232	0.533	0.396	0.461
Durables	1.727	2.397	3.744	5.361	5.598	8.155	-1.183	-0.414	1.682
Other	1.516	1.894	2.664	4.830	3.741	4.983	-0.135	0.490	1.566
Retail:									
Nondurables	1.961	2.627	2.522	5.473	4.687	4.961	-0.679	-0.380	-0.106
Durables	2.259	2.052	3.480	5.413	5.723	5.944	-0.902	-1.398	-0.921
Services	1.670	1.298	2.182	4.106	3.193	5.432	-1.375	-0.673	0.807
Finance	1.537	2.584	3.104	4.517	5.040	4.470	-0.571	0.301	0.411
Natural Resource	0.277	1.627	2.928	1.634	3.471	4.733	-0.666	0.789	2.186

returns than the low book-to-market portfolio. Interestingly, the high book-to-market portfolios in the durables manufacturing and durables retail industries have the highest average returns.

Panel B reports the nondurable consumption betas. Reading across the rows of the panel, nondurable consumption beta increases in book-to-market equity for each industry, except for the nondurables retail and finance industries. Similarly, durable consumption beta (Panel C) increases in book-to-market equity, except for the nondurables manufacturing and durables retail industries. Table 1.7 makes clear the source of the value premia. In a given industry, high book-to-market stocks have returns that are more pro-cyclical than low book-to-market stocks. Value stocks therefore carry a high premium to compensate the investor for bearing business-cycle risk, measured by consumption growth.

1.6.3 Risk-Sorted Portfolios

This section examines whether the durable consumption model prices portfolios sorted by risk. Risk-sorted portfolios provide a tough test for asset pricing models by creating a large spread in the post-formation betas. I construct portfolios by sorting stocks based on their pre-formation market and HML betas. The sort works well in practice. Portfolios with high (low) pre-formation market betas have high (low) post-formation nondurable consumption betas, and portfolios with high (low) pre-formation HML betas have high (low) post-formation durable consumption betas.

The reason for using the market return and HML, rather than nondurable and durable consumption growth, in forming portfolios is that returns are much more noisy than consumption. Therefore, pre-formation consumption betas are too noisy and fails to create the desired spread in the post-formation betas. The results for portfolios sorted by nondurable and durable consumption betas are reported in Appendix A.2.

Portfolio Formation

The portfolios are formed using returns on ordinary common equity, traded in NYSE, AMEX, or Nasdaq, in the CRSP Monthly Stock database. In June of each year t, market and HML betas are computed for each stock using monthly returns from January of $t-5$ through December of $t-1$. Stocks with return data missing in any month are dropped from the sample. Then 25 portfolios are formed by independently sorting stocks into quintiles based on the market and HML betas. The value-weighted portfolio returns are then tracked from July of t through June of $t+1$.

Test of Linear Factor Models

Table 1.8 reports estimates of the durable consumption model using the portfolios sorted by market and HML betas. Without the intratemporal FOC, the point estimate of the risk price for nondurable consumption is 148. The estimate of the risk price for durable consumption is 83, which is significantly different from zero, implying a rejection of the CCAPM. The R^2 is 47%, and the J-test fails to reject the model. The results are similar when the intratemporal FOC is imposed, although the J-test rejects the model in this case. This is due to the limitations of the linear approximation to the FOC, as discussed below.

1.6.4 Estimation of the Nonlinear Model

The empirical work has so far focused on the linearized durable consumption model, which results from a log-linear approximation to the nonlinear SDF. In this section, I estimate the nonlinear model to check the accuracy of the approximation. The

Table 1.8: Estimation of Linear Factor Models with Portfolios Sorted by Market and HML Betas

The test assets are 25 portfolios formed by independently sorting stocks into quintiles based on pre-formation market and HML betas. See notes to Table 1.2.

Factor Price	CAPM	Fama-French	CCAPM	Durable Model	
				No FOC	FOC
Market	2.545	5.121			
	(0.647)	(1.012)			
SMB		-4.569			
		(1.859)			
HML		4.989			
		(1.459)			
Nondurables			143.673	147.880	136.495
			(21.288)	(22.017)	(15.394)
Durables				83.499	103.762
				(27.213)	(19.065)
γ				231.379	240.257
				(32.741)	(32.560)
α (if $\rho = 0$)				0.362	0.434
				(0.087)	(0.029)
MAE (%)	0.518	0.257	0.328	0.232	0.262
R^2	-1.783	0.305	-0.053	0.473	0.311
J-test	24.641	18.355	28.806	29.761	39.878
	(0.425)	(0.685)	(0.228)	(0.156)	(0.022)

estimation also allows for separate identification of the three preference parameters (γ, ρ, and α) that determine the risk prices for nondurable and durable consumption.

Table 1.9 reports two estimates of the durable consumption model. The first estimate is based only on the N moment restrictions (1.19), with the nonlinear SDF (1.14), used to price the portfolios. The second estimate imposes the unconditional expectation of the intratemporal FOC (1.15) as an additional moment restriction. In equation (1.15), I set $(1-\delta)\beta = 0.94$ since the depreciation rate is about 6% per quarter; the results are not sensitive to reasonable variations in this parameter. Estimation is by two-step (efficient) GMM. HAC standard errors are computed by the VARHAC procedure with automatic lag length selection by AIC. Although the errors are in theory a martingale difference sequence, the maximum lag length is set to one quarter to account for the possibility of time aggregation in consumption data (see Hall (1988)).

Panel A reports the estimates using the 25 Fama-French portfolios. Without the intratemporal FOC, the point estimate of γ is 543, which is somewhat larger than the point estimate of 319 for the linearized model (Table 1.2). The point estimate of ρ is 1, implying perfect substitutability between nondurables and durables, but the standard error is large. In particular, the Cobb-Douglas case (i.e. $\rho = 0$) is approximately two standard errors from the point estimate. A plot (not reported) reveals that the GMM objective function is flat in the direction of ρ in the region, roughly $[0, 1]$, where it is minimized. In other words, ρ is not identified well enough to distinguish between values corresponding to high elasticity of substitution, although low values $\rho < 0$ are easily rejected. The Wald test strongly rejects the CCAPM (i.e. additive separability), which corresponds to the linear restriction $\rho = 1 - \gamma$.

When the intratemporal FOC is included as an additional moment restriction, the estimates of γ and α are somewhat smaller, but the results are qualitatively similar.

Table 1.9: GMM Estimation of the Unconditional Euler Equation with Portfolio Returns

The table reports the estimated preference parameters for the durable consumption model. It reports two estimates of the model, with and without the intratemporal FOC as an additional moment restriction. The test assets are (A) 25 Fama-French portfolios sorted by size and book-to-market equity, (B) 24 portfolios sorted by book-to-market equity within industry, (C) 25 portfolios sorted by market and HML betas, and (D) all 74 portfolios. Estimation is by two-step GMM. HAC standard errors in parentheses. The p-values for the Wald test for additive separability ($\rho = 1 - \gamma$) and the J-test (test of overidentifying restrictions) in parentheses.

	A. Fama-French		B. Industry & BE/ME		C. Beta-Sorted		D. All Portfolios	
Parameter	No FOC	FOC	No FOC	FOC	No FOC	FOC	No FOC	FOC
γ	542.707	388.701	440.885	393.325	392.918	367.108	365.297	292.720
	(46.350)	(45.602)	(23.989)	(27.325)	(40.115)	(38.017)	(4.493)	(7.062)
ρ	1.000	0.934	-0.037	0.623	0.979	0.670	0.230	0.260
	(0.488)	(0.517)	(0.412)	(0.459)	(0.578)	(0.467)	(0.065)	(0.128)
α	0.683	0.586	0.696	0.611	0.633	0.612	0.686	0.614
	(0.019)	(0.053)	(0.025)	(0.047)	(0.039)	(0.045)	(0.004)	(0.012)
Test for $\rho = 1 - \gamma$	136.126	73.489	335.875	208.985	95.299	93.037	6531.084	1705.502
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
J-test	36.396	31.423	26.427	28.269	25.966	25.107	50.766	47.764
	(0.028)	(0.113)	(0.191)	(0.167)	(0.253)	(0.345)	(0.967)	(0.988)

Namely, high risk aversion and high elasticity of substitution between the goods are necessary to explain the size and value premia. These estimates that impose the FOC appear to better identify ρ and α, which are parameters that govern intratemporal substitution. The Wald test rejects additive separability. The J-test fails to reject the model at conventional significance levels.

Panel B reports estimates using the 24 portfolios sorted by book-to-market equity within industry, and Panel C reports estimates using the 25 portfolios sorted by market and HML betas. The parameter estimates are quite similar across the panels. A representative household model with high risk aversion (i.e. $\gamma \approx 400$), unit elasticity of substitution between nondurables and durables (i.e. $\rho = 0$), and a larger budget share for durables (i.e. $\alpha \approx 0.6$) appears to price the cross section of stock returns. The J-test fails to reject the model, even when the intratemporal FOC is imposed. This suggests that the rejections of the linear model (Tables 1.6 and 1.8) are a consequence of linearization error in the intertemporal FOC (1.25), rather than a failure of the FOC itself.

Panel D reports the results when the model is estimated on all 74 portfolios. When the intratemporal FOC is imposed, the estimate of γ is 293, and the estimate of α is 0.61. The estimate of ρ is 0.26, implying an elasticity of substitution of 1.35 between nondurables and durables. These estimates have much smaller standard errors than those for the individual sets of portfolios (Panels A–C). The J-test fails to reject the model. These results confirm the conclusion from the findings for the linear two-factor model, that the model successfully prices the cross section of stock returns.

1.7 Time Series Tests

I now test the time series implications of the durable consumption model. Section 1.7.2 tests the model by GMM using portfolio returns and instruments that predict returns. Section 1.7.3 ties these results to the predictability of stock returns.

1.7.1 Data

For the empirical work in this section, I focus on five portfolios that capture the common variation in returns across the 25 Fama-French portfolios. The first is the market portfolio, which is a value-weighted portfolio for NYSE and AMEX stocks from the CRSP Indices database. The other four are the small stock, the big stock, the high book-to-market, and the low book-to-market portfolios. These portfolios are based on six portfolios sorted by size (breakpoint at the median) and book-to-market equity (breakpoints at the 30th and 70th percentiles). The difference in returns between the small and big stock portfolios is the SMB return. The difference in returns between the high and low book-to-market portfolios is the HML return. See Fama and French (1993) for details on the construction of these portfolios; the data is available from Professor French's webpage. In computing excess returns, the 90-day T-bill return is used as the riskfree rate.

The time series tests of the durable consumption model require instruments that are informative about the state of the economy. In addition to a constant, I use five instruments in the tests: (1) nondurable consumption growth, (2) durable consumption growth, (3) the dividend-price ratio, (4) the value spread, and (5) the long-short yield spread.

The dividend-price ratio for the CRSP value-weighted portfolio is constructed as the sum of dividends over the past four quarters divided by the current price. The dividend-price ratio is related, by a present-value relationship, to the expectation of future returns and dividend growth and therefore predicts returns (Campbell and Shiller 1988a).

The value spread is the difference in book-to-market equity between the high and low book-to-market portfolios. The value spread is related, by a present-value relationship, to the expectation of future returns and profitability and therefore predicts HML returns (Cohen, Polk, and Vuolteenaho 2003). Following Cohen, Polk, and Vuolteenaho, the book-to-market equity in June of year t is the book equity in December of $t-1$ divided by the market equity in June of t. The book-to-market equity in the subsequent months from July of t through May of $t+1$ is the book equity in December of $t-1$ divided by that month's market equity.

Following Fama and French (1989), the long yield used in computing the yield spread is Moody's Seasoned Aaa Corporate Bond Yield. The short rate used is the 1-month T-bill rate from the CRSP Fama Risk Free Rates database. The yield spread "tends to be low near business-cycle peaks and high near troughs" (Fama and French 1989, p. 30), much like the difference in nondurable and durable consumption growth (Figure 1.3(b)).

1.7.2 Estimation of the Conditional Euler Equation

Excess Stock Returns

Panel A of Table 1.10 reports two estimates of the conditional Euler equation (1.9) for the durable consumption model, using excess stock returns and instruments. The first estimate is based on 30 moment restrictions, corresponding the product of five excess returns with six instruments. The second estimate imposes six additional moment restrictions, corresponding to the product of the intratemporal FOC (1.15)

Table 1.10: GMM Estimation of the Conditional Euler Equation with Stock Returns and Instruments

The test assets are the CRSP value-weighted portfolio, the small stock portfolio, the big stock portfolio, the high book-to-market portfolio, the low book-to-market portfolio, and the 90-day T-bill (only in Panel B). The instruments are second lags of nondurable and durable consumption growth, the log dividend-price ratio, the value spread, the yield spread, and a constant. See notes to Table 1.9.

Parameter	A. Without T-bill		B. With T-bill	
	No FOC	FOC	No FOC	FOC
γ	478.467	337.678	158.861	114.309
	(32.289)	(20.012)	(12.221)	(12.887)
ρ	1.000	-0.080	-3.053	-0.829
	(0.316)	(0.063)	(1.664)	(0.111)
α	0.651	0.661	0.491	0.165
	(0.017)	(0.007)	(0.129)	(0.006)
β			2.193	1.660
			(0.107)	(0.095)
Test for $\rho = 1 - \gamma$	216.778	283.505	166.341	76.010
	(0.000)	(0.000)	(0.000)	(0.000)
J-test	36.235	42.123	62.738	90.191
	(0.110)	(0.133)	(0.001)	(0.000)

with the instruments. The instruments are lagged twice to account for time aggregation in consumption data, but the results are similar using once lagged instruments. Estimation is by two-step GMM, as described in Section 1.6.4.

Including the moment restrictions for the intratemporal FOC, the estimate of γ is 338, implying high risk aversion. The estimate of ρ is -0.08, with the Cobb-Douglas case (i.e. $\rho = 0$) within two standard errors. The estimate of α is 0.66. These estimated preference parameters agree with those for the cross-sectional tests (Table 1.9). The Wald test strongly rejects the hypothesis of additive separability (i.e. $\rho = 1 - \gamma$), which is consistent with the well-known rejection of the CCAPM. The J-test fails to reject the durable consumption model at conventional significance levels.

The fact that the estimates in Tables 1.9 and 1.10 agree deserves emphasis since it has important asset pricing implications. On the one hand, the estimates in Table 1.9 are based on the *unconditional* Euler equation, using a large cross section of portfolio returns. A successful fit of the model implies that the variation in average returns *across stocks* can be explained by the SDF (i.e. the marginal rate of substitution in consumption). On the other hand, the estimates in Table 1.10 are based on the *conditional* Euler equation, using instruments that are informative about the state of the economy. A successful fit of the model implies that the variation in average stock returns *through time* can be explained by the SDF. I provide further evidence for the time variation in the equity premium below.

Riskfree Rate

In Panel B of Table 1.10, I repeat the estimation in Panel A with six additional moment restrictions, corresponding to the product of the conditional Euler equation (1.8) for the riskfree rate with the instruments. This allows the identification of the discount factor β in addition to the other preference parameters. I again report two sets of estimates, depending on whether the moment restrictions corresponding to the intratemporal FOC are included.

The estimate of β is greater than one, implying a negative rate of time preference. This is a consequence of the well-known riskfree rate puzzle. Since the EIS is the inverse of risk aversion under power utility (1.10), large risk aversion necessarily implies low EIS. However, since consumption grows over time, a negative rate of time preference is necessary to explain the low average riskfree rate. Although preferences with $\beta > 1$ may be counter-intuitive, it is not problematic in the sense that competitive equilibria can still exist in an infinite-horizon growth economy (Kocherlakota 1990).

What is more problematic is that the estimates of γ and α are much smaller than those reported in Panel A. The inability of the durable consumption model to simultaneously price stock returns and the riskfree rate can be best understood using the log-linear approximation to the riskfree rate (1.31). Recall that the risk price for durable consumption can be approximated as $b_2 \approx \alpha\gamma$. Since durable consumption growth is persistent (its first-order autocorrelation is 0.88), a large value of b_2 implies large persistent movements in the riskfree rate. A large risk price for durables, necessary for explaining the cross-sectional and time variation in expected stock returns, results in a "riskfree rate volatility puzzle". I will come back to this issue in Section 1.8, where I show that the puzzle can be resolved by preferences that separate the EIS from risk aversion.

1.7.3 Time Variation in Expected Stock Returns

Predictability of Returns

Stock returns can be predicted by various financial variables such as valuation ratios and asset returns (see the references in the introduction). In a factor pricing model (1.30), time variation in the equity premium must be explained by time variation in the quantity of risk, measured by the conditional covariance of the factors with returns. Therefore, the same variables that predict returns (in equation (1.34)) must predict the product of the innovation to returns with the factors (in equation (1.36)). I now document this connection between risk and return for the durable consumption model.

Using the instrumental variables methodology (Section 1.4.2), I estimate the model with excess log returns on the five portfolios: (1) market, (2) small stock, (3) big stock, (4) high book-to-market, and (5) low book-to-market. The instruments are the same as those used in the GMM estimation of the model: nondurable and durable consumption growth, the dividend-price ratio, the value spread, the yield spread, and a constant. I impose the risk prices implied by the estimated preference parameters, reported in the second column of Table 1.10. They are 115 and 222 for nondurables and durables, respectively.

Panel A of Table 1.11 reports estimates of regression model (1.34), corresponding to the conditional mean of stock returns. Coefficients that are significant at the 5% level (i.e. t-statistic greater than 1.645) are in bold. For all five portfolios, the coefficient on nondurable consumption growth is positive and significant, while the coefficient on durable consumption growth is negative and significant (with exception of the low book-to-market portfolio). This implies that expected stock returns are high when nondurable consumption growth is high and durable consumption growth is low. As shown in Figure 1.3(b), nondurable consumption growth is high (low) relative to durable consumption growth at business-cycle troughs (peaks). The coefficients therefore imply a counter-cyclical equity premium.

The dividend-price ratio and the yield spread predict returns on the market portfolio, consistent with the findings reported in the literature (e.g. Campbell and Shiller (1988b) and Fama and French (1989)). Since the yield spread is counter-cyclical, its positive coefficient implies that the equity premium is counter-cyclical. The value spread reliably predicts returns on all five portfolios.

Panel B of Table 1.11 reports estimates of regression model (1.35), corresponding to the conditional variance of stock returns. The squared innovation to returns is less predictable than returns. Moreover, the coefficients are much smaller in magnitude than those for returns (Panel A), which implies the conditional variance has a relatively small contribution in the movements in expected returns (i.e. left side of equation (1.32)).

Table 1.11: Conditional Mean and Variance of Stock Returns

The table reports the conditional mean and variance of excess returns on the CRSP value-weighted portfolio, the small stock portfolio, the big stock portfolio, the high book-to-market portfolio, and the low book-to-market portfolio. The conditional mean (variance) is estimated from a regression of returns (squared innovation to returns) onto the instruments. The instruments are lags of nondurable and durable consumption growth, the log dividend-price ratio, the value spread, the yield spread, and a constant. Estimation is by two-step GMM. HAC standard errors in parentheses. Coefficients significant at the 5% level are in bold.

Instrument	Market	Size		BE/ME	
		Small	Big	High	Low
A. Conditional Mean					
Nondurables	**1.347**	**2.131**	**1.493**	**2.269**	**1.612**
	(0.601)	(1.059)	(0.625)	(0.876)	(0.883)
Durables	**-1.459**	**-1.661**	**-1.652**	**-1.841**	-1.395
	(0.666)	(0.997)	(0.656)	(0.751)	(0.966)
Dividend-Price	**0.024**	0.027	0.021	**0.029**	0.020
	(0.013)	(0.019)	(0.013)	(0.014)	(0.018)
Value Spread	**0.036**	**0.050**	**0.034**	**0.040**	**0.051**
	(0.015)	(0.026)	(0.016)	(0.020)	(0.024)
Yield Spread	**0.531**	0.643	**0.567**	0.567	0.571
	(0.285)	(0.449)	(0.290)	(0.361)	(0.406)
B. Conditional Variance					
Nondurables	**0.193**	**-0.702**	0.035	**-0.471**	-0.081
	(0.076)	(0.187)	(0.069)	(0.131)	(0.148)
Durables	**-0.217**	0.181	**-0.210**	-0.024	-0.060
	(0.089)	(0.169)	(0.080)	(0.106)	(0.156)
Dividend-Price	0.002	-0.002	0.002	0.002	0.000
	(0.002)	(0.004)	(0.002)	(0.003)	(0.003)
Value Spread	**0.005**	**0.011**	**0.006**	**0.008**	**0.013**
	(0.002)	(0.006)	(0.002)	(0.004)	(0.005)
Yield Spread	-0.056	-0.027	-0.038	-0.052	-0.013
	(0.049)	(0.084)	(0.045)	(0.056)	(0.080)

Table 1.12: Conditional Covariance of Stock Returns with Consumption Growth

The table reports the conditional covariance of excess returns with (A) nondurable and (B) durable consumption growth. The conditional covariance, reported in percent, is estimated from a regression of consumption growth times the innovation to returns onto the instruments. See notes to Table 1.11.

Instrument	Market	Size		BE/ME	
		Small	Big	High	Low
		A. Nondurable Consumption			
Nondurables	-0.482	-0.217	-0.363	0.074	-0.656
	(0.310)	(0.567)	(0.317)	(0.481)	(0.483)
Durables	-0.333	-0.423	**-0.510**	**-0.740**	-0.266
	(0.312)	(0.537)	(0.310)	(0.422)	(0.481)
Dividend-Price	**0.018**	**0.034**	**0.022**	**0.036**	**0.023**
	(0.007)	(0.013)	(0.008)	(0.011)	(0.011)
Value Spread	**0.033**	**0.047**	**0.038**	**0.047**	**0.047**
	(0.012)	(0.022)	(0.012)	(0.018)	(0.019)
Yield Spread	-0.110	-0.089	-0.137	-0.159	-0.082
	(0.129)	(0.190)	(0.130)	(0.161)	(0.180)
		B. Durable Consumption			
Nondurables	**0.899**	**0.912**	**0.867**	**0.876**	**1.046**
	(0.244)	(0.511)	(0.253)	(0.405)	(0.406)
Durables	**-0.532**	-0.487	**-0.526**	-0.449	-0.503
	(0.312)	(0.514)	(0.304)	(0.371)	(0.475)
Dividend-Price	0.002	-0.006	-0.002	-0.006	-0.003
	(0.006)	(0.010)	(0.006)	(0.008)	(0.009)
Value Spread	0.000	0.001	-0.003	-0.004	0.002
	(0.008)	(0.014)	(0.008)	(0.010)	(0.013)
Yield Spread	**0.283**	**0.329**	**0.317**	**0.326**	**0.296**
	(0.112)	(0.186)	(0.114)	(0.141)	(0.166)

Panel A of Table 1.12 reports estimates of regression model (1.36), where the factor is nondurable consumption growth. The dividend-price ratio and the value spread reliably predict the product of the innovation to returns with nondurable consumption growth. This implies that the conditional covariance of returns with nondurable consumption growth is high when the dividend-price ratio and the value spread are high.

Panel B reports estimates of regression model (1.36), where the factor is durable consumption growth. Nondurable consumption growth predicts the product of the innovation to returns with durable consumption growth positively, while durable consumption growth predicts it negatively. The yield spread predicts the product positively. This implies that the conditional covariance of returns with durable consumption growth is high when (1) nondurable consumption growth is high relative to durable consumption growth or (2) the yield spread is high. In other words, the conditional covariance of returns with durable consumption growth is counter-cyclical.

To summarize, Tables 1.11–1.12 have uncovered some interesting facts about the predictability of stock returns. On the one hand, the dividend-price ratio and the value spread predict returns because they predict *nondurable consumption risk*, that is the product of the innovation to returns with nondurable consumption growth. On the other hand, nondurable and durable consumption growth and the yield spread predict returns because they predict *durable consumption risk*, that is the product of the innovation to returns with durable consumption growth. This is consistent with the implications of the conditional factor model (1.32); time variation in expected returns must be accounted for by time variation in the conditional covariance of returns with either nondurable or durable consumption growth.

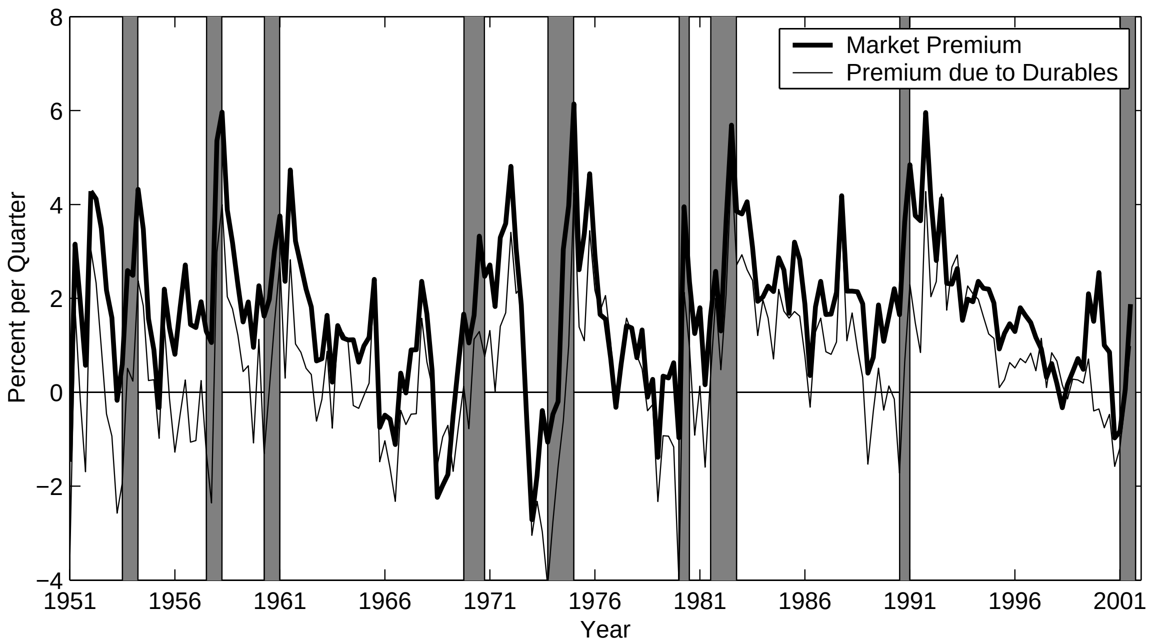

Figure 1.6: Time Variation in the Market Premium

The figure is a time series plot of expected excess returns on the CRSP value-weighted portfolio. The sample period is 1951:1–2001:3, and the shaded regions are NBER recessions.

Variance Decomposition of Returns

Figure 1.6 is a time series plot of the market premium (i.e. expected excess returns on the market portfolio), implied by the estimates in Tables 1.11–1.12. The dark line represents the total market premium, $\mathbf{E}_{t-1}[r_{it}] + \mathrm{Var}_{t-1}(r_{it})/2$, and the light line represents the part due to durables, $b_2\mathrm{Cov}_{t-1}(\Delta d_t, r_{it})$. The difference, of course, is the premium due to nondurables, $b_1\mathrm{Cov}_{t-1}(\Delta c_t, r_{it})$. The plot reveals two interesting facts. First, the two lines tend to overlap, which implies that most of the time variation in the equity premium is driven by the time variation in durable consumption risk. Second, the equity premium is strongly counter-cyclical, highest at business-cycle troughs and lowest at business-cycle peaks. Similar plots for the premium on the other four portfolios are reported in Figures 1.7–1.8.

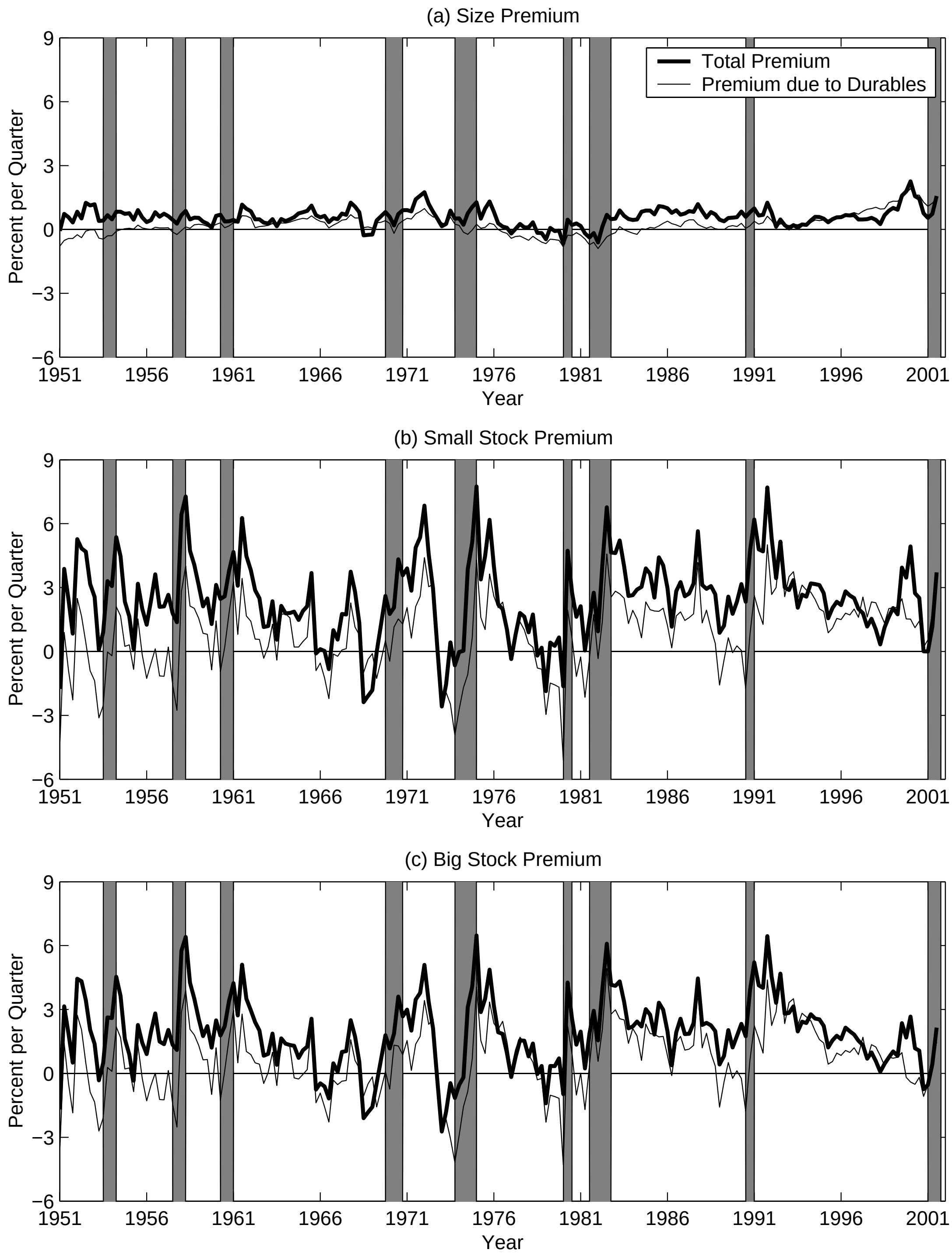

Figure 1.7: Time Variation in the Size Premium

The figure is a time series plot of expected excess returns on (b) the small stock portfolio and (c) the big stock portfolio. The size premium (a) is (b) minus (c). See notes to Figure 1.6.

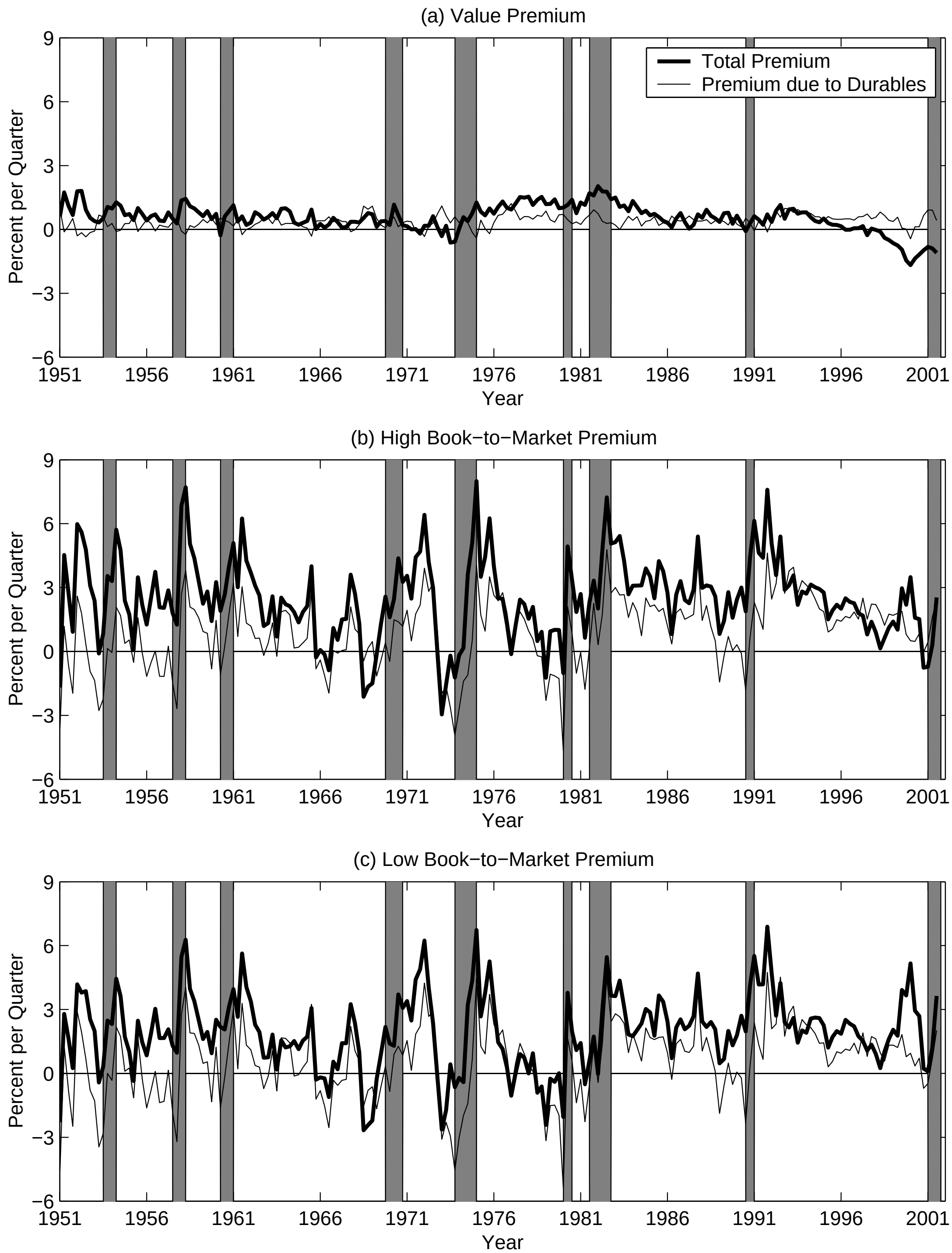

Figure 1.8: Time Variation in the Value Premium

The figure is a time series plot of expected excess returns on (b) the high book-to-market portfolio and (c) the low book-to-market portfolio. The value premium (a) is (b) minus (c). See notes to Figure 1.6.

Table 1.13: Variance Decomposition of Expected Stock Returns

The table reports the mean, the standard deviation, and the first-order autocorrelation of expected excess returns (per quarter) on the CRSP value-weighted portfolio, the small stock portfolio, the big stock portfolio, the high book-to-market portfolio, and the low book-to-market portfolio. It also reports a decomposition of the variance into the part due to nondurables premium, durables premium, and the covariance between the two premia.

Return	Mean (%)	Std Dev (%)	Autocorr	% of Variance due to		
				Nondurables	Durables	Covariance
Market	1.712	1.584	0.659	33.154	97.539	-30.693
Small	2.490	1.891	0.623	47.415	85.011	-32.427
Big	1.901	1.660	0.661	39.602	94.816	-34.418
Small−Big	0.589	0.429	0.695	79.108	126.971	-106.079
High BE/ME	2.502	1.908	0.610	53.546	75.730	-29.276
Low BE/ME	1.948	1.775	0.632	44.657	96.767	-41.425
High−Low	0.554	0.628	0.825	128.087	24.691	-52.779

The plot of the market premium resembles the plot of the difference between nondurable and durable consumption growth (Figure 1.3(b)). During a recession, durable consumption falls sharply relative to nondurable consumption, causing the marginal utility of consumption to rise sharply. This causes the market premium to rise sharply at the business-cycle trough. As durable consumption rises relative to nondurable consumption during the subsequent boom, marginal utility falls gradually, and so does the market premium. Time variation in the market premium simply reflects time variation in risk, measured by the marginal utility of consumption.

Table 1.13 reports the mean, the standard deviation, and the first-order autocorrelation of expected excess returns on the five portfolios. It also reports a variance decomposition of expected returns into the fraction due to nondurables premium, durables premium, and two times the covariance between the two premia. A large fraction of the variation in expected returns is due to variation in the durables premium. For instance, the nondurables premium only accounts for 33% of the variance in the market premium, while the durables premium accounts for 98%. (-31% is accounted for by the covariance between the two premia.) This explains why the CCAPM fails to explain the time variation in expected returns; it misses an important component of the cyclical variation in expected returns by ignoring the durables premium.

1.8 Riskfree Rate Puzzle

As noted in Section 1.7.2, the durable consumption model runs into a riskfree rate volatility puzzle. To assess the magnitude of the problem, I compute the implied riskfree rate using equation (1.31) and plot its time series in Figure 1.9(a). The risk prices for nondurables and durables are the same as those used to generate the implied market premium in Figure 1.6. I also use the same instruments to model the conditional moments of consumption growth. The rate of time preference $\overline{r} = 0$.

The expected riskfree rate has a mean 224% and fluctuates in the range of -200% to 500% per quarter! Most of the variation in the riskfree rate is due to intertemporal substitution (i.e. predictable movements in the first moment of consumption growth) rather than precautionary savings (i.e. predictable movements in the second moments). The large volatility results from a combination of the large risk price for durables and the high persistence of durable consumption growth. The expected riskfree rate is essentially a magnified version of durable consumption growth, shown in Figure 1.3(a).

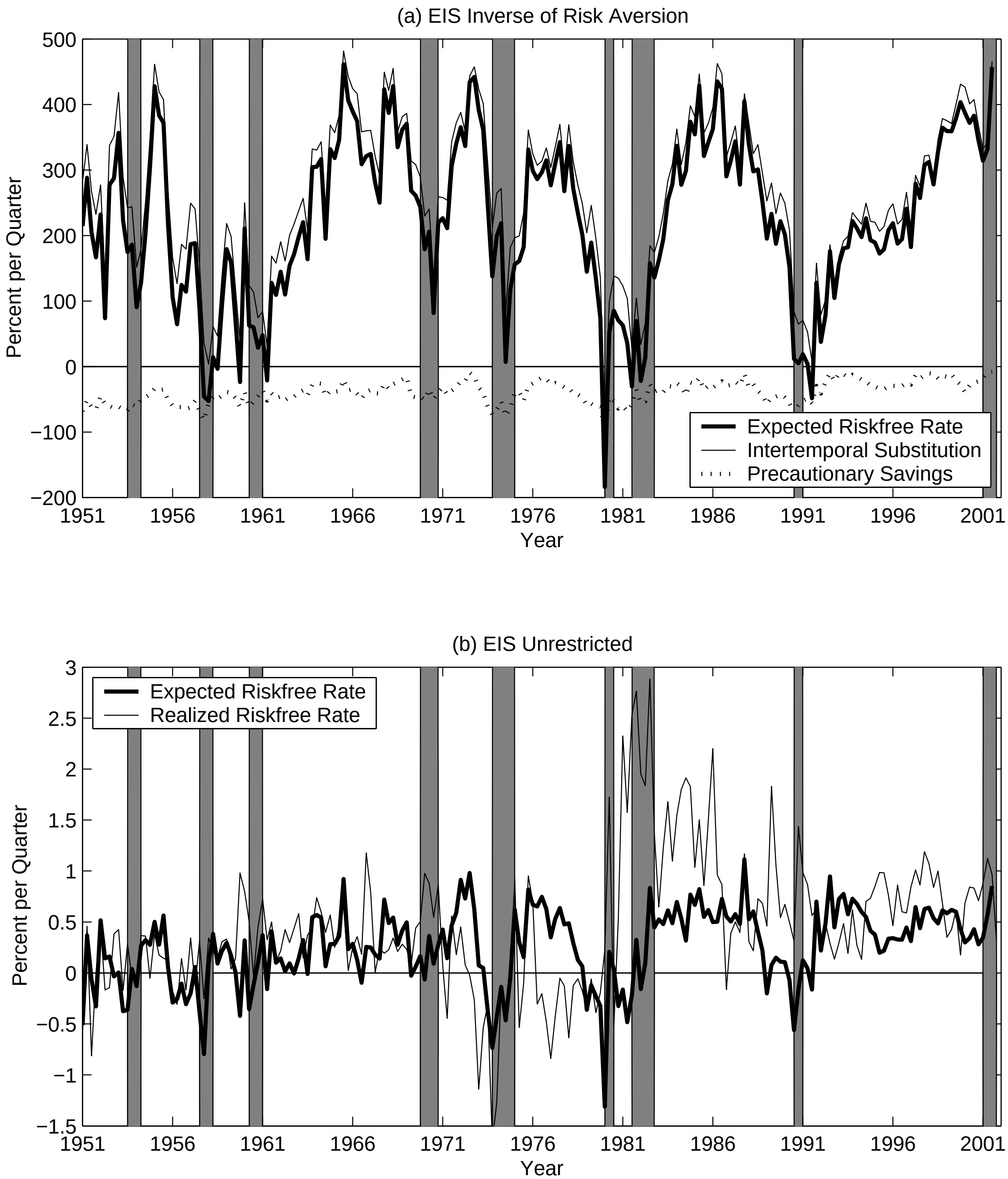

Figure 1.9: Expected Riskfree Rate

The figure is a time series plot of the expected riskfree rate implied by the durable consumption model when (a) the EIS is restricted to be the inverse of risk aversion and (b) the EIS is unrestricted. See notes to Figure 1.6.

1.8.1 OCE Preferences

In order to resolve the riskfree rate puzzle, I introduce preferences that allow for separation of the EIS from risk aversion. This allows me to retain the large risk price for durables, necessary for explaining expected stock returns, while getting rid of the large implied volatility in the riskfree rate. The derivations of the equations in this section are contained in Appendix A.3.

Household preferences are a generalization of ordinal certainty equivalent (OCE) preferences (Selden 1978) to the two good case. OCE preferences have been used in related empirical work by Hall (1985) and Attanasio and Weber (1989). The household's problem is the same as in Section 1.2, except his objective function (1.2) is now

$$\sum_{t=0}^{\infty} \beta^t \frac{\mathbf{E}_0[v_t^{1-\gamma}]^{\frac{1-\sigma}{1-\gamma}}}{1-\sigma}, \tag{1.41}$$

where $v_t = v(C_t, D_t)$. The parameter $\gamma > 0$ governs risk aversion, that is preferences over uncertain future utility flow. The parameter $\sigma > 0$ governs intertemporal substitution, that is the willingness to substitute the certainty equivalent of utility flow over time. In the special case $\gamma = \sigma$, the EIS $1/\sigma$ is the inverse of risk aversion γ, and the objective function reduces to (1.2).

The household's FOC results in an Euler equation

$$\mathbf{E}_{t-1}\left[\left(\frac{v_t}{v_{t-1}}\right)^{1-\gamma}\right]^{\frac{\gamma-\sigma}{1-\gamma}} \mathbf{E}_{t-1}[M_t R_{it}] = 1. \tag{1.42}$$

When $\gamma = \sigma$, this reduces to the Euler equation (1.8) for the durable consumption model. A nice property of OCE preferences is that equation (1.9) still holds. In other words, the equation that prices *excess* returns does not change, although *gross* returns are now priced by equation (1.42). Intuitively, the gross return on an asset is determined by both intertemporal substitution and risk aversion. In comparing the return of one asset relative to another, the part due to intertemporal substitution cancels, leaving only the part due to risk aversion.

Suppose the intraperiod utility is Cobb-Douglas (i.e. $\rho = 0$). Define the functions $b_1(x) = x + \alpha(1 - x)$ and $b_2(x) = -\alpha(1 - x)$. By a second-order log-linear approximation of equation (1.42) for the riskfree rate,

$$\begin{aligned} r_{0t} &= \bar{r} + b_1(\sigma)\mathbf{E}_{t-1}[\Delta c_t] + b_2(\sigma)\mathbf{E}_{t-1}[\Delta d_t] \\ &\quad -\frac{b_1(\gamma) + b_1(\sigma)(b_1(\gamma) - 1)}{2}\text{Var}_{t-1}(\Delta c_t) - \frac{b_2(\gamma)b_2(\sigma)}{2}\text{Var}_{t-1}(\Delta d_t) \\ &\quad -b_1(\sigma)b_2(\gamma)\text{Cov}_{t-1}(\Delta c_t, \Delta d_t). \end{aligned} \tag{1.43}$$

Note that the part due to intertemporal substitution now depends on σ, rather than γ. There are two special cases of interest. When $\gamma = \sigma$, this equation reduces to the riskfree rate under the durable consumption model (i.e. equation (1.31)). When $\sigma = 1$, $b_2(\sigma) = 0$ and

$$r_{0t} = \bar{r} + \mathbf{E}_{t-1}[\Delta c_t] + \frac{1}{2}\text{Var}_{t-1}(\Delta c_t) - \text{Cov}_{t-1}(\Delta c_t, b_1(\gamma)\Delta c_t + b_2(\gamma)\Delta d_t). \tag{1.44}$$

Note that the riskfree rate does not depend on $\mathbf{E}_{t-1}[\Delta d_t]$ in this case. An EIS close to one should therefore get rid of the persistent variation in the riskfree rate caused by durable consumption growth.

A problem with OCE preferences is dynamic inconsistency. Because the certainty equivalent of future utility depends on today's expectations, today's consumption plan will not be carried out when expectations are updated tomorrow. In an economy with a single nondurable consumption good, Epstein and Zin (1989) remedied this problem with recursive utility. It is not known whether Epstein-Zin utility can be extended to the case with two consumption goods, one of which is durable. A durable consumption good prevents a clean separation of the intratemporal optimization problem from the intertemporal problem.

Leaving these issues aside, OCE preferences are attractive because excess returns can be priced with the same equation as the durable consumption model. Therefore, all the empirical results in Sections 1.6–1.7 for excess stock returns continue to hold under OCE preferences.

1.8.2 Time Variation in the Riskfree Rate

Figure 1.9(b) is a time series plot of the riskfree rate using equation (1.43). The rate of time preference $\overline{r} = 0$ as before, and $\sigma = 1.14$ (EIS equal to 0.88) is chosen to minimize the squared difference between the left and right sides of equation (1.43).

The expected riskfree rate generated by the model is now reasonable, resolving the riskfree rate volatility puzzle. Its mean is 0.26% per quarter, which is somewhat lower than 0.47% for the realized rate (i.e. 90-day T-bill return minus inflation in the price index for nondurable goods). The difference is primarily due to an unexpectedly high real interest rate in the 1980's. It is interesting that the expected riskfree rate, generated using only consumption data, tracks some of the variation in the realized rate.

1.9 Conclusion

The findings of this essay suggest that there is much empirical content in the theoretical paradigm of consumption-based asset pricing. The central insight of the CCAPM is that the marginal utility of consumption is the relevant measure of risk for an investor. This essay has shown the marginal utility of consumption, when suitably modeled, can explain the tradeoff between risk and return reflected in the size premium, the value premium, and the time-varying equity premium.

The central ingredient is a non-separable utility function in nondurable and durable consumption, where the elasticity of substitution between the goods is high relative to the additively separable case. Small stocks and value stocks deliver low returns when marginal utility rises, that is during recessions when durable consumption falls. These stocks must therefore have high expected returns to reward the investor for bearing risk. In addition, stocks deliver unexpectedly low returns when marginal utility rises sharply, that is at business-cycle troughs when durable consumption falls sharply relative to nondurable consumption. The equity premium must therefore be high during recessions to reward the investor for bearing risk.

The mechanism through which the durable consumption model generates a counter-cyclical equity premium is similar to that of the external habit-formation model (Campbell and Cochrane 1999). In the Campbell-Cochrane model, the surplus consumption ratio is strongly pro-cyclical and magnifies the counter-cyclicality of marginal utility relative to the canonical CCAPM. In the durable consumption model, the ratio of durable to nondurable consumption is strongly pro-cyclical and magnifies the counter-cyclicality of marginal utility.

Although the durable consumption model can explain both the cross section of expected stock returns and the time variation in the equity premium, it requires rather high risk aversion to do so because of the low volatility of both nondurable and durable consumption. The riskfree rate volatility puzzle caused by high risk aversion can be resolved through preferences that separate the EIS from risk aversion. However, one may still "reject" the model on the grounds that high risk aversion is *a priori* unreasonable. The risk aversion implied by the Campbell-Cochrane model is also high, and in that model, the riskfree rate volatility puzzle is avoided by having intertemporal substitution exactly offset precautionary savings. I agree with the view that "high risk aversion is inescapable (or at least has not yet been escaped) in the class of identical-agent models that are consistent with the equity premium facts..." (Campbell and Cochrane 1999, p. 243)

Regardless of whether one believes in the representative household model, this essay has uncovered some intriguing facts about stock returns and the business cycle, which should guide future research.

1. Small stocks and value stocks have higher nondurable and durable consumption betas than big stocks and growth stocks. The returns on small stocks and value stocks are more pro-cyclical than those on big stocks and growth stocks.

2. The expected stock return is high (low) when nondurable consumption growth is high (low) relative to durable consumption growth. The equity premium is strongly counter-cyclical.

3. The conditional covariance of stock returns with durable consumption growth is high (low) when nondurable consumption growth is high (low) relative to durable consumption growth. Stock returns tend to be unexpectedly low (high) during recessions (booms).

Chapter 2

Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak[1]

2.1 Introduction

The elasticity of intertemporal substitution (EIS) in consumption is a parameter of central importance in macroeconomics and finance. In a basic model of the effects of monetary policy, the EIS is the parameter that relates current and expected future real interest rates to the current level of aggregate demand in the "intertemporal IS relation" (Woodford 2003, Chapter 4). In the consumption and portfolio choice problem of an infinitely lived investor with Epstein and Zin (1989) preferences, the EIS is the key parameter in the optimal consumption rule (Campbell and Viceira 1999).

[1]This chapter is forthcoming in *The Review of Economics and Statistics* 86:3 (August 2004) and is published here with permission from the MIT Press.

To estimate the EIS, denoted by ψ, one typically uses the regression equation

$$\Delta c_{t+1} = \tau_i + \psi r_{i,t+1} + \xi_{i,t+1}, \tag{2.1}$$

where Δc_{t+1} is the consumption growth at time $t+1$, $r_{i,t+1}$ is the real return on asset i at $t+1$, and τ_i is a constant. The error $\xi_{i,t+1}$, which is linear in the innovation to consumption growth and asset return, is correlated with the regressor $r_{i,t+1}$. However, given a vector of instruments Z_t uncorrelated with the error, ψ can be identified by the moment restriction

$$\mathbf{E}[Z_t \xi_{i,t+1}] = 0. \tag{2.2}$$

Z_t typically consists of economic variables known at time t, such as lagged consumption growth and asset return. Equation (2.1) can be estimated by two stage least squares (TSLS) if the error is homoskedastic, or by linear generalized method of moments (GMM) if the error is heteroskedastic.

Regression equation (2.1) can be written in the reversed form as

$$r_{i,t+1} = \mu_i + \frac{1}{\psi} \Delta c_{t+1} + \eta_{i,t+1}, \tag{2.3}$$

where μ_i is a constant and $\eta_{i,t+1}$ is the error. The inverse of the EIS, which is also the coefficient of relative risk aversion under power utility, is then identified by the moment restriction

$$\mathbf{E}[Z_t \eta_{i,t+1}] = 0. \tag{2.4}$$

Moment restrictions (2.2) and (2.4) are equivalent up to a linear transformation.

Using equation (2.1) or (2.3), numerous papers have estimated the EIS with US data (e.g. Hansen and Singleton (1983), Hall (1988), and Campbell and Mankiw

(1989)) and international data (e.g. Campbell (2003)). The general empirical finding is that the EIS estimated by equation (2.1) is small (Hall 1988), while its inverse estimated by equation (2.3) is also small (Hansen and Singleton 1983). For instance, Campbell (2003, Table 9) reports a 95% confidence interval of [-0.14,0.28] for ψ, using quarterly US data (1947–1998) on nondurable consumption and T-bill returns. On the other hand, he reports a 95% confidence interval of [-0.73,2.14] for $1/\psi$.

Therefore, one rejects the null hypothesis $\psi = 1$ using equation (2.1), which instruments for T-bill return, but fails to reject $\psi = 1$ using equation (2.3), which instruments for consumption growth. Whether $\psi < 1$ is of economic interest because it has important implications on the relative magnitudes of income and substitution effects in the intertemporal consumption decision of an investor facing time-varying expected returns. Campbell and Viceira (1999) show that when the EIS is less (greater) than one, the investor's optimal consumption-wealth ratio is increasing (decreasing) in expected returns.

Although equations (2.1) and (2.3) correspond to the same moment restriction up to a linear transformation, GMM is not invariant to such transformations. Therefore, the choice of normalization for the moment restriction can affect point estimates and confidence intervals. According to conventional first-order asymptotic theory, the choice of normalization should be negligible in large samples, leading to the same (at least approximately) inference of the EIS. In practice, however, equations (2.1) and (2.3) give very different (even contradictory) confidence intervals for the EIS as discussed above.

The leading explanation for this apparent failure of first-order asymptotics is weak instruments. In order for a vector of instruments Z_t to be valid, it must not only be exogenous but *relevant*, that is correlated with the endogenous variable $r_{i,t+1}$ in equation (2.1) or Δc_{t+1} in equation (2.3). As Neely, Roy, and Whiteman (2001) and

Campbell (2003) note, weak instruments is a problem in estimating the EIS because both consumption growth and asset returns are notoriously difficult to predict. Weak instruments can cause estimators to be severely biased and the finite-sample distribution of test statistics to depart sharply from the limiting distribution, leading to large size distortions in hypothesis tests (see Nelson and Startz (1990), Staiger and Stock (1997), or Stock, Wright, and Yogo (2002) for a recent survey).

The purpose of this essay is to estimate and make valid inference of the EIS for the eleven developed countries in Campbell's (2003) dataset, carefully accounting for problems caused by weak instruments. The idea that weak instruments is a problem in estimating the EIS is not new, and the paper that is closest to this is Neely, Roy, and Whiteman (2001). Showing that weak instruments may account for the discrepancy between small values of ψ estimated by equation (2.1) and small values of $1/\psi$ estimated by equation (2.3), Neely, Roy, and Whiteman (2001, p. 403) conclude that "prior beliefs grounded in economic theory seem to be necessary to settle the consumption CAPM debate over small versus large risk aversion" because of identification failure.[2]

Compared to Neely, Roy, and Whiteman (2001), this essay goes a step further to estimate the EIS despite near identification failure. I am able to make progress on the EIS debate due to recent methods that have been developed to handle weak instruments. Stock and Yogo (2003) have developed a pretest, based on the first-stage F-statistic, to formally test whether given instruments are weak. Some instrumental variables (IV) estimators, such as the limited information maximum likelihood (LIML) estimator, provide more reliable point estimates and inferences with weak instruments, compared to TSLS (Hausman, Hahn, and Kuersteiner 2001, Stock and Yogo 2003). Kleibergen (2002) and Moreira (2001, 2003) have developed pivotal

[2]Neely, Roy, and Whiteman assume power utility, so the risk aversion is the inverse of the EIS.

statistics to test coefficients in the structural equation, which result in tests with correct size regardless of the strength of identification. Using these methods, I conclude that the EIS is small across the eleven developed countries, which agrees with Hall's (1988) finding for the US.

The rest of the essay is organized as follows. Section 2.2 reviews the assumptions necessary to derive regression equations (2.1) and (2.3) from the Euler equation for Epstein and Zin (1989) preferences. Section 2.3 outlines the relevant econometric methods when instruments are weak. Section 2.4 applies these econometric methods to data from eleven developed countries and discusses the empirical findings. Section 2.5 concludes.

2.2 Linearized Euler Equation

Let δ be the subjective discount factor, γ be the coefficient of relative risk aversion, and define $\theta = (1-\gamma)/(1-1/\psi)$. The Epstein and Zin (1989, 1991) objective function is defined recursively by

$$U_t = [(1-\delta)C_t^{(1-\gamma)/\theta} + \delta(\mathbf{E}_t U_{t+1}^{1-\gamma})^{1/\theta}]^{\theta/(1-\gamma)}, \tag{2.5}$$

where C_t is consumption at time t. In the special case $\gamma = 1/\psi$, (2.5) reduces to the familiar time separable power utility model with period utility $U(C_t) = C_t^{1-\gamma}/(1-\gamma)$. The representative household maximizes objective function (2.5) subject to the intertemporal budget constraint

$$W_{t+1} = (1 + R_{w,t+1})(W_t - C_t), \tag{2.6}$$

where W_{t+1} is the household's wealth and $1 + R_{w,t+1}$ is the gross real return on the portfolio of all invested wealth at $t + 1$. Epstein and Zin (1991) show that (2.5) and (2.6) together imply an Euler equation of the form

$$\mathbf{E}_t \left[\left(\delta \left(\frac{C_{t+1}}{C_t} \right)^{-1/\psi} \right)^{\theta} \left(\frac{1}{1 + R_{w,t+1}} \right)^{1-\theta} (1 + R_{i,t+1}) \right] = 1, \tag{2.7}$$

where $1 + R_{i,t+1}$ is the gross real return on asset i.

2.2.1 Conditional Homoskedasticity

Let lowercase letters denote the log of the corresponding uppercase variables (e.g. $r_{i,t+1} = \log(1 + R_{i,t+1})$). Assuming that asset returns and consumption are homoskedastic and jointly lognormal conditional on information at time t, the Euler equation (2.7) can be linearized as

$$\mathbf{E}_t r_{i,t+1} = \mu_i + \frac{1}{\psi} \mathbf{E}_t \Delta c_{t+1}, \tag{2.8}$$

$$\mu_f = -\log \delta + \frac{\theta - 1}{2} \mathrm{Var}(r_{w,t+1} - \mathbf{E}_t r_{w,t+1}) - \frac{\theta}{2\psi^2} \mathrm{Var}(\Delta c_{t+1} - \mathbf{E}_t \Delta c_{t+1}), \tag{2.9}$$

$$\begin{aligned} \mu_i = {} & \mu_f - \frac{1}{2} \mathrm{Var}(r_{i,t+1} - \mathbf{E}_t r_{i,t+1}) + \frac{\theta}{\psi} \mathrm{Cov}(r_{i,t+1} - \mathbf{E}_t r_{i,t+1}, \Delta c_{t+1} - \mathbf{E}_t \Delta c_{t+1}) \\ & + (1 - \theta) \mathrm{Cov}(r_{i,t+1} - \mathbf{E}_t r_{i,t+1}, r_{w,t+1} - \mathbf{E}_t r_{w,t+1}). \end{aligned} \tag{2.10}$$

(See Campbell (2003) or Campbell and Viceira (2002, Chapter 2) for a textbook treatment.) Without the assumption of lognormality, equation (2.8) holds as a second-order loglinear approximation of (2.7). For a conditionally riskfree asset, equation (2.8) reduces to

$$r_{f,t+1} = \mu_f + \frac{1}{\psi} \mathbf{E}_t \Delta c_{t+1}. \tag{2.11}$$

Regression equation (2.3) is obtained from equation (2.8) by setting

$$\eta_{i,t+1} = r_{i,t+1} - \mathbf{E}_t r_{i,t+1} - \frac{1}{\psi}(\Delta c_{t+1} - \mathbf{E}_t \Delta c_{t+1}).$$

The error $\eta_{i,t+1}$ is conditionally homoskedastic by the same assumption used to linearize the Euler equation. It is straightforward to show that $\eta_{i,t+1}$ is serially uncorrelated and satisfies the moment restriction (2.4). The efficient two-step GMM estimator is TSLS in this case. In order for the instruments to be relevant (i.e. not weak), they must be correlated with consumption growth Δc_{t+1}.

Regression equation (2.1) is obtained by rearranging (2.3), which implies that

$$\xi_{i,t+1} = \Delta c_{t+1} - \mathbf{E}_t \Delta c_{t+1} - \psi(r_{i,t+1} - \mathbf{E}_t r_{i,t+1}).$$

Since moment restriction (2.2) is satisfied, ψ can be estimated by TSLS. In this normalization, the instruments are weak if they are weakly correlated with asset return $r_{i,t+1}$.

2.2.2 Conditional Heteroskedasticity

If asset returns and consumption are conditionally heteroskedastic, Euler equation (2.7) can still be linearized as equation (2.8). The only difference is that the variance and covariance terms that appear in the intercept μ_i must be replaced by conditional variances and covariances. In this section, I show that the EIS can still be identified by the same moment restrictions.

To simplify the notation, consider the linearized Euler equation for the riskfree asset (2.11) in the special case of power utility (i.e. $\gamma = 1/\psi$ and $\theta = 1$),

$$r_{f,t+1} = \mu_{f,t} + \gamma \mathbf{E}_t \Delta c_{t+1}, \tag{2.12}$$

$$\mu_{f,t} = -\log\delta - \frac{\gamma^2}{2}\mathrm{Var}_t(\Delta c_{t+1} - \mathbf{E}_t \Delta c_{t+1}). \tag{2.13}$$

The intercept $\mu_{f,t}$ is now subscripted by t to account for conditional heteroskedasticity in consumption, which represents precautionary savings. As long as the vector of instruments Z_t is uncorrelated with the innovation to the conditional variance of consumption, that is $\mathbf{E}[Z_t(\mu_{f,t} - \mu_f)] = 0$, the inverse of the EIS (the coefficient of relative risk aversion in this case) is identified by moment restriction (2.4). In this case, TSLS is consistent but is no longer the efficient two-step GMM estimator.

This suggests that even if instruments Z_t are correlated with the conditional variance and covariance terms that appear in $\mu_{i,t}$, a vector of twice lagged instruments Z_{t-1} satisfies the moment restriction $\mathbf{E}[Z_{t-1}\eta_{i,t+1}] = 0$, where

$$\eta_{i,t+1} = \mu_{i,t} - \mu_i + r_{i,t+1} - \mathbf{E}_t r_{i,t+1} - \frac{1}{\psi}(\Delta c_{t+1} - \mathbf{E}_t \Delta c_{t+1}).$$

In other words, the inverse of the EIS can still be identified by regression equation (2.3) although inference must now account for conditional heteroskedasticity in the error $\eta_{i,t+1}$. A similar point has been made by Attanasio and Low (2000) in the context of estimating the linearized Euler equation on household data. They argue that the coefficient of relative risk aversion can be identified with sufficiently long time series data in response to Carroll's (2001) criticism that it cannot be estimated consistently on a cross section of households.

2.2.3 Estimation of the Nonlinear Euler Equation

This essay focuses on estimation of the EIS based on the linearized Euler equation. In this section, I briefly compare this approach to estimation based on the nonlinear Euler equation.

Given a vector of instruments, the preference parameters δ, γ, and ψ can be estimated by GMM through the nonlinear Euler equation (2.7). This is the approach taken by Hansen and Singleton (1982) for the power utility case and by Epstein and Zin (1991) for Epstein-Zin preferences. As noted by Epstein and Zin (1991), the difficulty with this approach is that it requires knowledge of returns on the wealth portfolio, which includes returns on human capital. Hence, Roll's (1977) critique on the testability of CAPM applies. In contrast, the EIS can be estimated from the linearized Euler equation (2.8) without knowledge of returns on the wealth portfolio.

Aside from this practical advantage, the reason for focusing on the linearized Euler equation is that much more is known about weak instruments in the linear IV regression model. Many of the recent econometric methods that handle weak instruments (e.g. Stock and Yogo (2003), Kleibergen (2002), and Moreira (2003)) apply to the linear IV model with conditional homoskedasticity. I therefore impose the assumption of conditional homoskedasticity for most of the empirical work in Section 2.4, although I also check that the results are robust to heteroskedasticity.

The main disadvantage of the linearized Euler equation is that the discount factor δ cannot be identified since it enters additively in the intercept (2.10), along with unknown second moments of innovations to consumption and asset returns (Attanasio and Low 2000). Nevertheless, the study of weak instruments in the linear model is interesting because of the large existing literature that uses this methodology, starting with Hansen and Singleton (1983) and Hall (1988). For those interested in the nonlinear model, I refer to a related study by Stock and Wright (2000). They develop GMM asymptotic theory under weak identification and apply it to estimation of preference parameters through the nonlinear Euler equation.

2.3 Econometric Methods for Weak Instruments

Following the notation in Staiger and Stock (1997), the linear IV regression model is

$$y = Y\beta + X\gamma + u, \tag{2.14}$$

$$Y = Z\Pi + X\Phi + V, \tag{2.15}$$

where (2.14) is the structural equation of interest and (2.15) is the reduced form for the n endogenous regressors. y is a $T \times 1$ vector of T observations, Y is a $T \times n$ matrix of endogenous regressors, X is a $T \times K_1$ matrix of included exogenous regressors, and Z is a $T \times K_2$ matrix of instruments excluded from the structural equation. All matrices have full rank, and the order condition $K_2 \geq n$ is satisfied. u and V are $T \times 1$ vector and $T \times n$ matrix of errors, respectively, whose rows are assumed to be serially uncorrelated with mean zero and covariance matrix

$$\mathbf{E}\left[\begin{pmatrix} u_t \\ V_t \end{pmatrix} \begin{pmatrix} u_t & V_t' \end{pmatrix}\right] = \Sigma = \begin{bmatrix} \sigma_{uu} & \Sigma_{Vu}' \\ \Sigma_{Vu} & \Sigma_{VV} \end{bmatrix}. \tag{2.16}$$

Let $\overline{Z} = [X, Z]$, where $\overline{Z}_t'$ denotes its tth row. Then the identifying assumption is $\mathbf{E}[\overline{Z}_t(u_t, V_t')] = 0$.

The reduced form for y is

$$y = Z\Pi\beta + X(\Phi\beta + \gamma) + v, \tag{2.17}$$

where $v = u + V\beta$. The rows of $\overline{V} = [v, V]$ are serially uncorrelated with mean zero and covariance matrix

$$\mathbf{E}\left[\begin{pmatrix} v_t \\ V_t \end{pmatrix} \begin{pmatrix} v_t & V_t' \end{pmatrix}\right] = \Omega = \Sigma + \begin{bmatrix} \beta'\Sigma_{VV}\beta + 2\beta'\Sigma_{Vu} & \beta'\Sigma_{VV} \\ \Sigma_{VV}\beta & 0 \end{bmatrix}. \tag{2.18}$$

2.3.1 k-Class Estimators

Let $\overline{Y} = [y, Y]$ and $\overline{X} = [Y, X]$. Define the matrices $P_X = X(X'X)^{-1}X'$ and $M_X = I - P_X$. (Analogous notation is used for projection onto matrices other than X.) Let the superscript $\perp$ denote the residual from the projection onto X (e.g. $Y^\perp = M_X Y$). The k-class estimator of β is

$$\widehat{\beta}(k) = [Y^{\perp\prime}(I - kM_{Z^\perp})Y^\perp]^{-1}[Y^{\perp\prime}(I - kM_{Z^\perp})y^\perp]. \tag{2.19}$$

The three special cases of interest in this essay are

1. TSLS with $k = 1$;
2. LIML with $k = \widehat{k}_{LIML}$, where $\widehat{k}_{LIML}$ is the smallest root of the determinantal equation $|\overline{Y}'M_X\overline{Y} - k\overline{Y}'M_{\overline{Z}}\overline{Y}| = 0$;
3. Fuller-k (Fuller 1977) with $k = k_{LIML} - 1/(T - K_1 - K_2)$.

The Wald statistic for testing the null hypothesis $\beta = \beta_0$ is

$$W(k) = \frac{(\widehat{\beta}(k) - \beta_0)'[Y^{\perp\prime}(I - kM_{Z^\perp})Y^\perp](\widehat{\beta}(k) - \beta_0)}{n\widehat{\sigma}_{uu}(k)}, \tag{2.20}$$

where $\widehat{\sigma}_{uu}(k) = \widehat{u}(k)'\widehat{u}(k)/(T - K_1 - n)$ and $\widehat{u}(k) = y^\perp - Y^\perp\widehat{\beta}(k)$.

Under conventional first-order asymptotics, the three k-class estimators and the corresponding Wald statistics have the same asymptotic distribution (see Amemiya

(1985, pp. 236–238)). However, first-order asymptotics is a poor approximation in finite samples when instruments are weak (Nelson and Startz 1990). Staiger and Stock (1997) develop an alternative asymptotic framework, "weak-instrument asymptotics," which accurately approximates the sampling distribution of estimators and test statistics even when instruments are weak.

Under weak-instrument asymptotics, the three estimators and the corresponding Wald statistics have nonstandard limiting distributions that differ from one another. Both TSLS and Fuller-k are biased, but the bias of Fuller-k is less severe for given population parameters. Similarly, the size distortion of the LIML Wald test is less severe than that of the TSLS Wald test (Stock and Yogo 2003). Hence, Fuller-k and LIML can be thought of as estimators that are more robust to weak instruments than TSLS (see Stock, Wright, and Yogo (2002, Section 6)).

2.3.2 Test for Weak Instruments

Suppose there is only one endogenous regressor in the structural equation (i.e. $n = 1$). Then the key population parameter that measures the relevance of the instruments is the concentration parameter,

$$\mu^2 = \frac{\Pi' Z^{\perp\prime} Z^{\perp} \Pi}{\Sigma_{VV}}. \tag{2.21}$$

Following the discussion in Rothenberg (1984, Section 6), μ^2 can be thought of as the "sample size" in simultaneous equations models. When μ^2 is large, the TSLS estimator is approximately unbiased, and the distribution of the its t-statistic is approximately standard normal. When μ^2 is small, the TSLS estimator can be badly biased, and the distribution of the its t-statistic can be highly skewed (see Stock, Wright, and Yogo (2002, Figure 1)).

This suggests that one can test whether instruments are weak by testing whether μ^2 is sufficiently small to cause bias or size distortion. To test the null hypothesis that instruments are weak, Stock and Yogo (2003) propose using the first-stage F-statistic,

$$F = \frac{\widehat{\Pi}' Z^{\perp\prime} Z^{\perp} \widehat{\Pi}}{K_2 \widehat{\Sigma}_{VV}}, \tag{2.22}$$

where $\widehat{\Pi} = [Z^{\perp\prime} Z^{\perp}]^{-1} Z^{\perp\prime} Y^{\perp}$ and $\widehat{\Sigma}_{VV} = Y' M_{\overline{Z}} Y/(T - K_1 - K_2)$. Note that the F-statistic is the sample analog of the concentration parameter (2.21), scaled by K_2. The null hypotheses that I consider in this essay are:

1. The bias of TSLS as a fraction of OLS bias is greater than 10%. (10.27)

2. The actual size of the TSLS t-test at 5% significance can be greater than 10%. (24.58)

3. The bias of Fuller-k as a fraction of OLS bias is greater than 10%. (6.37)

4. The actual size of the LIML t-test at 5% significance can be greater than 10%. (5.44)

The numbers in parentheses are the critical values of the test at 5% significance when $K_2 = 4$, taken from Stock and Yogo (2003, Tables 1–4). For instance, to assure that TSLS relative bias is no greater than 10%, the F-statistic must be greater than 10.27. That the critical value for TSLS is greater than the critical value for Fuller-k is a reflection of the fact that the latter is more robust to weak instruments. Likewise, LIML is less prone to size distortion than TSLS for the same level of instrument relevance, which results in a lower critical value.

2.3.3 Similar Tests

The pretest described in the last section can detect weak instruments, protecting the researcher from biased estimates and misleading inferences. However, a researcher may be interested in making valid inference of the structural parameter β despite having weak instruments. In this section, I outline methods fully robust to weak instruments that accomplish this task.

Moreira (2001, 2003) has characterized the family of similar tests in the IV regression model when instruments are *fixed* (i.e. Z is nonrandom), the reduced-form errors $\overline{V}_t$ are independently and identically distributed *normal*, and the reduced-form covariance matrix Ω is *known*. Under these assumptions, he showed that there is a pair of independent sufficient statistics, $\mathcal{S}$ and $\mathcal{T}$, for the unknown parameters β and Π. Under the null hypothesis $\beta = \beta_0$, $\mathcal{S}$ is pivotal (i.e. its distribution does not depend on Π), and $\mathcal{T}$ is sufficient for nuisance parameter Π. Hence, any nonpivotal statistic $\phi(\mathcal{S}, \mathcal{T}, \beta_0)$ becomes a pivotal statistic conditional on $\mathcal{T} = \tau$. Let $c(\tau, \beta_0, \alpha)$ be the upper α-quantile of the null distribution of $\phi(\mathcal{S}, \tau, \beta_0)$. The test that rejects the null if $\phi(\mathcal{S}, \tau, \beta_0) > c(\tau, \beta_0, \alpha)$ is similar at the level α (see Moreira (2003, Theorem 1)).

In the general IV regression model (i.e. stochastic regressors, non-Gaussian errors, and unknown Ω), Moreira's exact finite-sample results hold asymptotically under weak-instrument asymptotics (Moreira 2003, Theorem 2). This result is not surprising since weak-instrument asymptotics corresponds to the finite-sample distribution theory for the simultaneous equations model with fixed regressors, Gaussian errors, and known reduced-form covariance matrix. Consequently, the family of similar tests forms a basis for fully robust inference in the presence of weak instruments.[3]

To characterize these tests, define the vectors $a_0 = (\beta_0, 1)'$ and $b_0 = (1, -\beta_0)'$ and

[3]For simplicity, I use the terminology "similar test" instead of "asymptotically similar test," hopefully without confusion.

the statistics

$$\begin{aligned}\mathcal{S} &= \frac{(Z^{\perp\prime}Z^{\perp})^{-1/2}Z^{\perp\prime}\overline{Y}^{\perp}b_0}{(b_0'\widehat{\Omega}b_0)^{1/2}}, &(2.23)\\ \mathcal{T} &= \frac{(Z^{\perp\prime}Z^{\perp})^{-1/2}Z^{\perp\prime}\overline{Y}^{\perp}\widehat{\Omega}^{-1}a_0}{(a_0'\widehat{\Omega}^{-1}a_0)^{1/2}}, &(2.24)\end{aligned}$$

where $\widehat{\Omega} = \overline{Y}' M_{\overline{Z}}\overline{Y}/(T - K_1 - K_2)$ is a consistent estimator of Ω. In this essay, I consider three Gaussian similar tests:

1. The Anderson-Rubin (AR) test (Anderson and Rubin 1949) based on the statistic

$$AR(\beta_0) = \frac{\mathcal{S}'\mathcal{S}}{K_2}, \tag{2.25}$$

which is asymptotically distributed $\chi^2_{K_2}/K_2$ under the null (Staiger and Stock 1997, Theorem 5).

2. The Lagrange multiplier (LM) test (Kleibergen 2002) based on the statistic

$$LM(\beta_0) = \frac{(\mathcal{S}'\mathcal{T})^2}{\mathcal{T}'\mathcal{T}}, \tag{2.26}$$

which is asymptotically distributed χ^2_1 under the null (Kleibergen 2002, Theorem 1).

3. The conditional likelihood ratio (LR) test (Moreira 2003) based on the statistic

$$LR(\beta_0) = \frac{1}{2}(\mathcal{S}'\mathcal{S} - \mathcal{T}'\mathcal{T} + \sqrt{(\mathcal{S}'\mathcal{S} + \mathcal{T}'\mathcal{T})^2 - 4[(\mathcal{S}'\mathcal{S})(\mathcal{T}'\mathcal{T}) - (\mathcal{S}'\mathcal{T})^2]}), \tag{2.27}$$

whose critical values can be computed as a function of K_2 and $\mathcal{T}'\mathcal{T}$ by Monte Carlo simulation.

It is well known that the AR test is invariant to linear transformations of the GMM moment restriction. In the context of moment restrictions (2.2) and (2.4), the AR test rejects the null hypothesis $\psi = \psi_0$ based on (2.2) if and only if it rejects $1/\psi = 1/\psi_0$ based on (2.4). It can be readily verified that the LM and the conditional LR tests share this invariance property. In contrast, two-step GMM is not invariant to linear transformations of the moment restriction, which results in contradictory inference about the EIS depending on whether one uses moment restriction (2.2) or (2.4).

These similar tests can be inverted to construct confidence regions for β. For instance, one can construct a $(1-\alpha)100\%$ confidence region based on the AR test as

$$\{\beta_0 \in \mathcal{B} | AR(\beta_0) < \chi^2_{K_2,\alpha}/K_2\},$$

where $\mathcal{B}$ is the parameter space for β and $\chi^2_{K_2,\alpha}$ is the upper α-quantile of the $\chi^2_{K_2}$ distribution. If the parameter space is unrestricted, $\mathcal{B}$ is the set of reals; if β is restricted to be positive, which may be the case for the EIS, $\mathcal{B}$ is the set of positive reals. In general, these confidence regions can consist of disjoint intervals. By taking the minimum and maximum values of β in the confidence region, one obtains a confidence interval that has coverage of at least $(1-\alpha)100\%$.

2.3.4 Power of Similar Tests

Because more powerful tests lead to tighter confidence intervals, one would like to use the most powerful similar test to construct confidence intervals that are robust to weak instruments. Unfortunately, there is no uniformly most powerful test (Moreira 2001), so the three similar tests have relative power advantages in different regions of the parameter space. In this section, I discuss their power properties.

A natural way to evaluate the power of similar tests is to consider their asymptotic power under weak-instrument asymptotics. Asymptotically, the power functions depend on the scaled concentration parameter μ^2/K_2 and the degree of endogeneity $\rho = \Sigma_{Vu}/(\Sigma_{VV}\sigma_{uu})^{1/2}$. In an earlier version of this essay, I reported the power functions, which have since been published in Stock, Wright, and Yogo (2002, Figure 2). To avoid redundancy, I refer to their figures in the following discussion.[4] Appendix A.4 details the asymptotic results used to plot the power functions, which were omitted from Stock, Wright, and Yogo (2002) to save space.

Stock, Wright, and Yogo plot the power functions for two levels of instrument relevance $\mu^2/K_2 = 1, 5$ and two levels of endogeneity $\rho = 0.5, 0.99$. When instruments are very weak (i.e. $\mu^2/K_2 = 1$), all three similar tests have poor power in the sense that the power is far less than one even at distant alternatives. As a consequence, confidence intervals based on similar tests can be unbounded when instruments are very weak. When $\rho = 0.5$, the AR and conditional LR tests have better power than the LM test. When $\rho = 0.99$, the LM and conditional LR tests have better power than the AR test. When instruments are moderately weak (i.e. $\mu^2/K_2 = 5$), the similar tests have much better power; their power approaches one for distant alternatives. Among the three similar tests, the conditional LR test has the best power properties; its power function comes close to the power envelope for similar tests at both values of ρ. This suggests that the conditional LR test should usually result in confidence intervals that are tightest among the three similar tests.

[4]Figure 2 in Stock, Wright, and Yogo (2002) is for $K_2 = 5$ instruments, whereas $K_2 = 4$ in the empirical application of this essay. However, this is not a substantive difference since the power functions for $K_2 = 4$ and $K_2 = 5$ are essentially the same.

2.3.5 Heteroskedasticity and Simultaneous Estimation

In this section, I discuss econometric methods for the more general GMM setting. This generalization allows for conditional heteroskedasticity and simultaneous estimation. For instance, if the linearized Euler equation (2.8) holds for both the interest rate and stock return, GMM allows for simultaneous estimation of the EIS using both moment restrictions. The cost of going to the more general setup is that methods designed to handle weak instruments are much less developed. Generalizations of the test for weak instruments or similar tests to GMM are topics of ongoing research.

To be concrete with notation, suppose there are two $K_2 \times 1$ vectors of linear moment restrictions, which is the relevant case for this essay. Let y_1, y_2, Y_1, and Y_2 be $T \times 1$ vectors of T observations on jointly endogenous variables. As before, Z is a $T \times K_2$ matrix of instruments, and superscript $\perp$ denotes the residual from projection onto the included exogenous regressors X. To simplify notation, let X include a column of ones so that the residuals from the projection have mean zero. Define the $2K_2 \times 1$ vector

$$\phi_t(\beta) = \begin{bmatrix} Z_t^{\perp}(y_{1t}^{\perp} - Y_{1t}^{\perp}\beta) \\ Z_t^{\perp}(y_{2t}^{\perp} - Y_{2t}^{\perp}\beta) \end{bmatrix}. \tag{2.28}$$

The moment restriction is $\mathbf{E}[\phi_t(\beta_0)] = 0$.

Define the heteroskedasticity robust weighting matrix

$$V(\beta) = T^{-1} \sum_{t=1}^{T} \phi_t(\beta)\phi_t(\beta)' \tag{2.29}$$

and the objective function

$$S(\beta, \overline{\beta}) = \left[T^{-1/2} \sum_{t=1}^{T} \phi_t(\beta)\right]' V(\overline{\beta})^{-1} \left[T^{-1/2} \sum_{t=1}^{T} \phi_t(\beta)\right]. \tag{2.30}$$

The efficient two-step GMM estimator $\widehat{\beta}_2$ minimizes $S(\beta, \widehat{\beta}_1)$ for some consistent first-step estimator $\widehat{\beta}_1$. For example, the first-step estimator can be obtained by minimizing (2.30) using the weighting matrix $V(\overline{\beta}) = I_2 \otimes (Z^{\perp\prime} Z^{\perp})$. Since the objective function is quadratic in this case, the two-step estimator has the closed form

$$\widehat{\beta}_2 = \left[\begin{bmatrix} Z^{\perp\prime} Y_1^{\perp} \\ Z^{\perp\prime} Y_2^{\perp} \end{bmatrix}' V(\widehat{\beta}_1)^{-1} \begin{bmatrix} Z^{\perp\prime} Y_1^{\perp} \\ Z^{\perp\prime} Y_2^{\perp} \end{bmatrix}\right]^{-1} \left[\begin{bmatrix} Z^{\perp\prime} Y_1^{\perp} \\ Z^{\perp\prime} Y_2^{\perp} \end{bmatrix}' V(\widehat{\beta}_1)^{-1} \begin{bmatrix} Z^{\perp\prime} y_1^{\perp} \\ Z^{\perp\prime} y_2^{\perp} \end{bmatrix}\right]. \tag{2.31}$$

An alternative to two-step GMM is the continuous updating estimator (CUE), which minimizes the objective function $S(\beta, \beta)$. In a Monte Carlo experiment designed to simulate estimation of the linearized Euler equation, Hansen, Heaton, and Yaron (1996) find that CUE is less biased and its confidence intervals have better coverage rates than two-step GMM. The intuition for this result is that CUE is a generalization of LIML to GMM, just as two-step GMM is a generalization of TSLS.

In the GMM setting, the analog of weak instruments is weak identification (Stock and Wright 2000), which, loosely speaking, occurs if $\mathbf{E}[\phi_t(\beta)] \approx 0$ even when $\beta \neq \beta_0$. Confidence intervals for β with the correct coverage can be constructed by inverting the objective function (2.30) evaluated at β_0. By Stock and Wright (2000, Theorem 2), $S(\beta_0, \beta_0)$ is asymptotically distributed $\chi^2_{2K_2}$. This test, which I refer to as the S-test, is a generalization of the AR test to GMM.

2.4 Empirical Results

2.4.1 Data

The dataset that I use is from Campbell (2003). It consists of quarterly data on equity markets at an aggregate level and macroeconomic variables for eleven developed countries: Australia (AUL), Canada (CAN), France (FR), Germany (GER), Italy (ITA), Japan (JAP), Netherlands (NTH), Sweden (SWD), Switzerland (SWT), the United Kingdom (UK), and the United States (USA). In addition, a longer time series is available at annual frequency for Sweden, the UK, and the US. The primary sources of international data are Morgan Stanley Capital International and International Financial Statistics of the International Monetary Fund. The sample periods vary by country and frequency, which are reported in Table 2.1. With exception of the US, quarterly data is only available starting in 1970. For the quarterly US series, I report the results for both the full sample, which starts in 1947, and a truncated sample that starts in 1970. For a full description of the dataset, see Campbell (2003) and the accompanying data appendix Campbell (1998).

For each country, I estimate the EIS using two asset returns: the real interest rate, denoted by r_f, and the real aggregate stock return, denoted by r_e. The real stock return is constructed as log of the gross stock return deflated by the consumer price index. The real interest rate is constructed in the same way, using an available proxy for the short-term interest rate. Real consumption growth is the first difference in log real consumption per capita. For all quarterly series except for the US, the consumption measure is total consumption rather than nondurables and services due to data availability. The timing convention used for consumption is “beginning of the period,” following Campbell (2003). In other words, I assume that the consumption data for a given time period is the flow measured at the beginning of the period rather than at the end.

2.4.2 Test for Weak Instruments

The coefficients of interest are the EIS ψ, estimated by equation (2.1), and its inverse $1/\psi$, estimated by equation (2.3). The instruments that I use for the endogenous regressor Δc_{t+1} in equation (2.3) and $r_{i,t+1}$ in equation (2.1) are the nominal interest rate, inflation, consumption growth, and log dividend-price ratio.[5] All instruments are lagged twice to avoid problems with time aggregation in consumption data (Hall 1988). As discussed in Section 2.2, this also assures that instruments are exogenous even if consumption or asset returns are conditionally heteroskedastic.

Assuming that the error is conditionally homoskedastic, equations (2.1) and (2.3) can be estimated by TSLS. In Table 2.1, I report the first-stage F-statistic for each of the possible endogenous regressors (consumption growth, interest rate, and stock return), which is the relevant statistic to test for weak instruments. Next to the F-statistic, I report the p-values of the test. A p-value less than 0.05 means that the test would reject the null hypothesis of weak instruments at the 5% significance level. As explained in Section 2.3, the p-value depends on the type of estimator (TSLS, Fuller-k, or LIML) used for estimation or inference.

At the quarterly frequency, consumption growth and stock return both have low predictability as evidenced by low F-statistics, so the test fails to reject the null of weak instruments. Hence, a researcher should suspect that the TSLS estimator is biased and the TSLS t-test is size distorted. In fact, instruments are so weak in this case that estimation or inference based on Fuller-k or LIML are also suspect. On the other hand, the interest rate appears to be more predictable for all countries.

[5] Campbell (2003) uses the real interest rate, instead of the nominal interest rate and inflation, for a total of three instruments. For many countries, the nominal interest rate appears to contain important information about future real asset returns.

Table 2.1: Test for Weak Instruments

The table reports the first-stage F-statistic from a regression of the endogenous variable onto the instruments. The endogenous variables are consumption growth (Δc), real interest rate (r_f), and real stock return (r_e). The instruments are twice lagged nominal interest rate, inflation, consumption growth, and log dividend-price ratio. The table also reports the p-value of the test for weak instruments. The null hypotheses are: (1) TSLS relative bias is greater than 10%, (2) the size of 5% TSLS t-test can be greater than 10%, (3) Fuller-k relative bias is greater than 10%, and (4) the size of 5% LIML t-test can be greater than 10%.

				p-value			
Country	Sample Period	Variable	F	TSLS Bias	TSLS Size	Fuller-k	LIML
USA	1947.3–1998.4	Δc	2.93	0.93	1.00	0.53	0.37
		r_f	15.53	0.00	0.66	0.00	0.00
		r_e	2.88	0.93	1.00	0.54	0.39
AUL	1970.3–1998.4	Δc	1.79	0.99	1.00	0.81	0.69
		r_f	21.81	0.00	0.14	0.00	0.00
		r_e	1.82	0.99	1.00	0.80	0.68
CAN	1970.3–1999.1	Δc	3.03	0.92	1.00	0.50	0.35
		r_f	15.37	0.00	0.67	0.00	0.00
		r_e	2.51	0.96	1.00	0.64	0.48
FR	1970.3–1998.3	Δc	0.17	1.00	1.00	1.00	1.00
		r_f	38.43	0.00	0.00	0.00	0.00
		r_e	3.09	0.91	1.00	0.49	0.34
GER	1979.1–1998.3	Δc	0.83	1.00	1.00	0.97	0.93
		r_f	17.66	0.00	0.45	0.00	0.00
		r_e	0.69	1.00	1.00	0.98	0.95
ITA	1971.4–1998.1	Δc	0.73	1.00	1.00	0.98	0.95
		r_f	19.01	0.00	0.33	0.00	0.00
		r_e	1.10	1.00	1.00	0.94	0.88
JAP	1970.3–1998.4	Δc	1.18	1.00	1.00	0.93	0.86
		r_f	8.64	0.14	0.99	0.01	0.00
		r_e	3.49	0.87	1.00	0.40	0.25
NTH	1977.3–1998.4	Δc	0.89	1.00	1.00	0.96	0.92
		r_f	12.05	0.01	0.91	0.00	0.00
		r_e	0.73	1.00	1.00	0.98	0.95

(continued on the next page)

				p-value			
Country	Sample period	Variable	F	TSLS Bias	TSLS Size	Fuller-k	LIML
SWD	1970.3–1999.2	Δc	0.48	1.00	1.00	0.99	0.98
		r_f	17.08	0.00	0.51	0.00	0.00
		r_e	2.24	0.97	1.00	0.70	0.56
SWT	1976.2–1998.4	Δc	0.97	1.00	1.00	0.95	0.90
		r_f	8.55	0.14	0.99	0.01	0.00
		r_e	0.11	1.00	1.00	1.00	1.00
UK	1970.3–1999.1	Δc	2.52	0.96	1.00	0.63	0.48
		r_f	17.04	0.00	0.51	0.00	0.00
		r_e	2.62	0.95	1.00	0.61	0.45
USA	1970.3–1998.4	Δc	3.53	0.86	1.00	0.39	0.25
		r_f	11.92	0.02	0.92	0.00	0.00
		r_e	2.16	0.97	1.00	0.72	0.58
SWD	1921–1994	Δc	1.02	1.00	1.00	0.95	0.89
		r_f	5.50	0.55	1.00	0.10	0.05
		r_e	1.67	0.99	1.00	0.84	0.72
UK	1921–1994	Δc	1.93	0.98	1.00	0.78	0.65
		r_f	4.87	0.66	1.00	0.16	0.08
		r_e	4.18	0.77	1.00	0.26	0.15
USA	1891–1995	Δc	1.55	0.99	1.00	0.86	0.76
		r_f	2.87	0.93	1.00	0.54	0.39
		r_e	1.00	1.00	1.00	0.95	0.90

The F-statistic is large enough that the Fuller-k estimator is approximately unbiased. In addition, the LIML t-test leads to approximately correct inference, although the TSLS t-test may be size distorted.

For the annual series, none of the regressors appear to be sufficiently predictable to avoid problems with weak instruments. The only possible exceptions are the UK and the US, where the instruments are somewhat relevant in predicting the interest rate. The LIML t-test should lead to approximately correct inference since the test for weak instruments rejects at the 10% level.

2.4.3 Estimates of the EIS Using the Interest Rate

In the first three columns of Table 2.2, I report the point estimate and standard error of $1/\psi$ with the interest rate as the dependent variable in equation (2.3). I report results using TSLS, Fuller-k, and LIML. The first fact to note is that the three estimators give very different results. Under conventional first-order asymptotics, the three estimators have the same asymptotic distribution. Therefore, the fact that the three estimators give very different results is indirect evidence for weak instruments. In general, the magnitude of both the coefficient and standard error increases from TSLS to Fuller-k and from Fuller-k to LIML. The 95% confidence intervals for $1/\psi$ based on these estimators include rather large values of the EIS. In particular, one cannot reject the null hypothesis $\psi = 1$, except for Canada and Switzerland.

In the last three columns of Table 2.2, I report estimates of the EIS using equation (2.1) with the interest rate as the endogenous regressor. In contrast to inference based on equation (2.3), which requires that the instruments predict consumption growth, weak instruments is not a problem because the interest rate is sufficiently predictable, as documented in Table 2.1. Consequently, the three estimators give very

Table 2.2: Estimates of the EIS Using the Interest Rate

The inverse of the EIS is estimated from $r_{f,t+1} = \mu_f + (1/\psi)\Delta c_{t+1} + \eta_{f,t+1}$, and the EIS is estimated from $\Delta c_{t+1} = \tau_f + \psi r_{f,t+1} + \xi_{f,t+1}$. The instruments are twice lagged nominal interest rate, inflation, consumption growth, and log dividend-price ratio. Standard error in parentheses.

		$1/\psi$			ψ		
Country	Sample Period	TSLS	Fuller-k	LIML	TSLS	Fuller-k	LIML
USA	1947.3–1998.4	0.68	3.30	34.11	0.06	0.03	0.03
		(0.48)	(3.20)	(112.50)	(0.09)	(0.10)	(0.10)
AUL	1970.3–1998.4	0.50	2.37	30.03	0.05	0.04	0.03
		(0.48)	(2.45)	(107.71)	(0.11)	(0.12)	(0.12)
CAN	1970.3–1999.1	-1.04	-2.40	-2.98	-0.30	-0.33	-0.34
		(0.39)	(1.13)	(1.54)	(0.16)	(0.17)	(0.17)
FR	1970.3–1998.3	-3.12	-1.83	-12.38	-0.08	-0.08	-0.08
		(3.75)	(1.72)	(29.61)	(0.19)	(0.19)	(0.19)
GER	1979.1–1998.3	-1.05	-1.38	-2.29	-0.42	-0.43	-0.44
		(0.62)	(0.90)	(1.87)	(0.35)	(0.35)	(0.36)
ITA	1971.4–1998.1	-3.34	-5.82	-14.81	-0.07	-0.07	-0.07
		(1.98)	(4.47)	(18.55)	(0.08)	(0.08)	(0.08)
JAP	1970.3–1998.4	-0.18	-0.86	-21.56	-0.04	-0.04	-0.05
		(0.43)	(1.23)	(106.53)	(0.21)	(0.23)	(0.23)
NTH	1977.3–1998.4	-0.53	-1.41	-6.94	-0.15	-0.15	-0.14
		(0.41)	(1.33)	(13.96)	(0.28)	(0.29)	(0.29)
SWD	1970.3–1999.2	-0.10	-0.21	-399.86	0.00	0.00	0.00
		(1.10)	(1.54)	(16075.06)	(0.10)	(0.10)	(0.10)
SWT	1976.2–1998.4	-1.56	-1.51	-2.00	-0.49	-0.49	-0.50
		(0.83)	(0.79)	(1.18)	(0.29)	(0.29)	(0.29)
UK	1970.3–1999.1	1.06	3.76	6.21	0.17	0.16	0.16
		(0.45)	(2.42)	(5.17)	(0.13)	(0.13)	(0.13)
USA	1970.3–1998.4	0.53	2.19	47.66	0.06	0.02	0.02
		(0.50)	(2.60)	(249.47)	(0.09)	(0.11)	(0.11)
SWD	1921–1994	1.17	3.30	17.77	0.06	0.06	0.06
		(1.13)	(3.34)	(38.67)	(0.11)	(0.12)	(0.12)
UK	1921–1994	2.40	2.99	3.52	0.26	0.27	0.28
		(1.01)	(1.33)	(1.65)	(0.12)	(0.13)	(0.13)
USA	1891–1995	-0.38	-1.17	-39.71	-0.03	-0.03	-0.03
		(1.12)	(2.90)	(257.54)	(0.11)	(0.15)	(0.16)

similar coefficients and standard errors. The point estimates of ψ are small, although sometimes negative. The 95% confidence intervals based on these estimators reject large values of the EIS, in particular one.

To summarize the results in Table 2.2, one would conclude the EIS is small and significantly less than one while its inverse is not significantly different from one. The hypothesis $\psi = 1$ is of economic interest because with Epstein-Zin preferences, an investor's optimal consumption choice is a constant fraction of wealth when the EIS is equal to one. Moreover, in the special case of power utility where the EIS is equal to the inverse of risk aversion, $\gamma = 1/\psi = 1$ leads to myopic portfolio choice (see Campbell and Viceira (2002, Chapter 2)). This apparent empirical puzzle, emphasized by Neely, Roy, and Whiteman (2001), can be accounted for by weak instruments. Regression equation (2.3) leads to biased estimates and confidence intervals with poor coverage because the instruments cannot predict consumption growth adequately to identify $1/\psi$. On the other hand, estimation by equation (2.1) leads to valid inference since instruments are not weak for the interest rate.

The sensitivity of inference to the particular normalization of the moment restriction is an unattractive property of k-class estimators.[6] In contrast, confidence intervals based on the similar tests (AR, LM, and conditional LR) are invariant to this normalization. Moreover, since these methods are fully robust to weak instruments, there is no need for a pretest to make sure that the instruments are relevant.

In Table 2.3, I report the 95% confidence intervals for the EIS constructed from the similar tests. Of the three similar tests, the conditional LR test tends to give the tightest confidence intervals, consistent with the fact that it has the best power properties. Focusing on the quarterly series and the conditional LR confidence interval, the EIS is less than 0.5 across all eleven countries. For the annual series, the EIS is

[6] The point estimate of LIML is invariant to normalization, but its confidence interval is not.

Table 2.3: Weak Instrument Robust Confidence Intervals for the EIS Using the Interest Rate

The table reports 95% confidence intervals for the EIS, constructed from AR, LM, and conditional LR tests. ∅ indicates an empty confidence interval. The instruments are twice lagged nominal interest rate, inflation, consumption growth, and log dividend-price ratio.

Country	Sample Period	AR	LM	Cond LR
USA	1947.3–1998.4	∅	[-0.21,0.23]	[-0.19,0.22]
AUL	1970.3–1998.4	[-0.16,0.21]	[-0.22,13.74]	[-0.22,0.27]
CAN	1970.3–1999.1	[-0.54,-0.14]	[-0.73,14.15]	[-0.71,0.00]
FR	1970.3–1998.3	[-0.68,0.53]	[-0.47,0.31]	[-0.48,0.33]
GER	1979.1–1998.3	[-1.57,0.54]	[-1.21,0.26]	[-1.23,0.28]
ITA	1971.4–1998.1	[-0.29,0.18]	[-0.24,0.11]	[-0.24,0.12]
JAP	1970.3–1998.4	[-0.60,0.49]	[-∞,∞]	[-0.56,0.45]
NTH	1977.3–1998.4	[-0.91,0.64]	[-∞,∞]	[-0.76,0.48]
SWD	1970.3–1999.2	[-0.30,0.29]	[-∞,∞]	[-0.22,0.21]
SWT	1976.2–1998.4	[-1.69,0.37]	[-1.19,0.07]	[-1.22,0.09]
UK	1970.3–1999.1	[0.04,0.28]	[-∞,∞]	[-0.12,0.43]
USA	1970.3–1998.4	∅	[-∞,∞]	[-0.23,0.23]
SWD	1921–1994	[-0.30,0.40]	[-∞,∞]	[-0.25,0.35]
UK	1921–1994	[-0.05,0.88]	[0.01,0.70]	[0.01,0.70]
USA	1891–1995	[-0.49,0.46]	[-∞,∞]	[-∞,∞]

similarly small for Sweden and the UK. There appears to be identification failure for the annual US series as evidenced by the uninformative confidence intervals $[-\infty, \infty]$ for both the LM and the conditional LR tests, although the AR test gives small estimates of the EIS. In summary, the weak instrument robust confidence intervals indicate that the EIS is small and not significantly different from zero for the eleven developed countries.

2.4.4 Estimates of the EIS Using the Stock Return

In Tables 2.4 and 2.5, I report the same set of results as Tables 2.2 and 2.3 using the stock return instead of the interest rate. As demonstrated in Table 2.1, both consumption growth and stock return are difficult to predict, so either normalization of the moment restriction, (2.2) or (2.4), runs into problems with weak instruments. This is evidenced by the fact that most of the weak instrument robust confidence intervals in Table 2.5 are uninformative. For few of the countries (Canada, France, and Japan at quarterly frequency and the UK at annual frequency), the confidence intervals are informative. Note that these are precisely the series for which the first-stage F-statistics for stock return were relatively large in Table 2.1, ranging from 2.51 to 4.18. The confidence intervals indicate that the EIS is small, agreeing with the results for the interest rate in Table 2.3.

In Tables 2.3 and 2.5, I have reported the unconstrained confidence intervals. Constrained confidence intervals that restrict the EIS to be nonnegative can usually be obtained by truncating the unconstrained confidence intervals at zero. Although the truncated confidence interval has the correct coverage rate provided that $\psi \geq 0$, it may be conservative when identification is sufficiently weak so that the confidence region is disjoint. In other words, the truncated confidence interval may not coincide with the actual constrained confidence interval. In Table 2.5, this occurs only for the quarterly (1970.3-1998.4) US series, where the actual constrained confidence intervals are $[0.05, \infty]$ and $[0.02, \infty]$ for the AR and conditional LR tests, respectively, since these tests reject $\psi = 0$.

Table 2.4: Estimates of the EIS Using the Stock Return

The inverse of the EIS is estimated from $r_{e,t+1} = \mu_e + (1/\psi)\Delta c_{t+1} + \eta_{e,t+1}$, and the EIS is estimated from $\Delta c_{t+1} = \tau_e + \psi r_{e,t+1} + \xi_{e,t+1}$. The instruments are twice lagged nominal interest rate, inflation, consumption growth, and log dividend-price ratio. Standard error in parentheses.

		$1/\psi$			ψ		
Country	Sample Period	TSLS	Fuller-k	LIML	TSLS	Fuller-k	LIML
USA	1947.3–1998.4	-1.33	-10.04	-14.18	-0.01	-0.05	-0.07
		(4.48)	(9.63)	(12.68)	(0.02)	(0.05)	(0.06)
AUL	1970.3–1998.4	6.63	9.19	11.24	0.05	0.07	0.09
		(4.55)	(6.15)	(7.57)	(0.04)	(0.05)	(0.06)
CAN	1970.3–1999.1	7.86	7.46	7.99	0.12	0.11	0.13
		(3.12)	(2.96)	(3.17)	(0.05)	(0.05)	(0.05)
FR	1970.3–1998.3	-19.92	-11.10	-50.48	-0.02	-0.02	-0.02
		(24.66)	(11.32)	(97.01)	(0.04)	(0.04)	(0.04)
GER	1979.1–1998.3	-1.63	-2.83	-5.83	-0.03	-0.05	-0.17
		(4.91)	(6.52)	(10.51)	(0.07)	(0.10)	(0.31)
ITA	1971.4–1998.1	3.25	8.48	40.70	0.01	0.01	0.02
		(9.94)	(16.68)	(82.89)	(0.03)	(0.04)	(0.05)
JAP	1970.3–1998.4	10.16	12.75	17.20	0.05	0.05	0.06
		(5.73)	(7.52)	(11.13)	(0.03)	(0.04)	(0.04)
NTH	1977.3–1998.4	1.29	2.18	4.20	0.03	0.07	0.24
		(3.76)	(5.15)	(8.06)	(0.08)	(0.13)	(0.46)
SWD	1970.3–1999.2	-8.57	-13.35	-64.89	-0.01	-0.01	-0.02
		(11.19)	(16.60)	(127.62)	(0.03)	(0.03)	(0.03)
SWT	1976.2–1998.4	-0.35	-0.40	-0.29	-0.05	-0.03	-3.45
		(4.10)	(3.82)	(4.35)	(0.19)	(0.12)	(51.80)
UK	1970.3–1999.1	-0.68	-5.54	-9.24	-0.01	-0.07	-0.11
		(2.76)	(6.51)	(10.10)	(0.04)	(0.08)	(0.12)
USA	1970.3–1998.4	6.92	7.77	8.05	0.03	0.08	0.12
		(4.86)	(6.61)	(7.11)	(0.02)	(0.05)	(0.11)
SWD	1921–1994	-1.75	-5.37	-12.94	-0.03	-0.05	-0.08
		(3.57)	(6.79)	(16.63)	(0.06)	(0.08)	(0.10)
UK	1921–1994	5.28	13.95	29.64	0.04	0.03	0.03
		(3.05)	(10.62)	(33.96)	(0.03)	(0.04)	(0.04)
USA	1891–1995	0.47	-1.06	-2.47	0.02	-0.08	-0.41
		(2.27)	(3.33)	(4.46)	(0.08)	(0.18)	(0.73)

Table 2.5: Weak Instrument Robust Confidence Intervals for the EIS Using the Stock Return

See notes to Table 2.3.

Country	Sample Period	AR	LM	Cond LR
USA	1947.3–1998.4	[-0.21,-0.02]	[-∞,∞]	[-∞,∞]
AUL	1970.3–1998.4	[-∞,∞]	[-∞,∞]	[-∞,∞]
CAN	1970.3–1999.1	[0.02,4.03]	[0.05,0.35]	[0.04,0.41]
FR	1970.3–1998.3	[-0.28,0.20]	[-∞,∞]	[-0.16,0.11]
GER	1979.1–1998.3	[-∞,∞]	[-∞,∞]	[-∞,∞]
ITA	1971.4–1998.1	[-∞,∞]	[-∞,∞]	[-∞,∞]
JAP	1970.3–1998.4	[-0.05,0.32]	[-1.01,0.20]	[-0.02,0.21]
NTH	1977.3–1998.4	[-∞,∞]	[-∞,∞]	[-∞,∞]
SWD	1970.3–1999.2	[-∞,∞]	[-∞,∞]	[-∞,∞]
SWT	1976.2–1998.4	[-∞,∞]	[-∞,∞]	[-∞,∞]
UK	1970.3–1999.1	[-0.51,-0.02]	[-∞,∞]	[-∞,∞]
USA	1970.3–1998.4	[-∞,∞]	[-∞,∞]	[-∞,∞]
SWD	1921–1994	[-∞,∞]	[-∞,∞]	[-∞,∞]
UK	1921–1994	[-0.04,0.10]	[-∞,∞]	[-0.10,0.14]
USA	1891–1995	[-∞,∞]	[-∞,∞]	[-∞,∞]

2.4.5 Heteroskedasticity and Simultaneous Estimation

In the first two columns of Table 2.6, I report GMM and CUE estimates of the EIS using moment restriction (2.2) for the interest rate. These estimates are heteroskedasticity robust versions of the estimates by TSLS and LIML, reported in Table 2.2. Comparing GMM and TSLS, the point estimates and standard errors are quite similar, so heteroskedasticity does not appear to be a dominant feature of the data. Likewise, the LIML and CUE estimates are quite similar. In the third column, I report weak instrument robust confidence intervals computed by inverting the S-test. These are heteroskedasticity robust versions of the AR confidence intervals reported in Table 2.3. In general, these confidence intervals are comparable to those reported in Table 2.3, indicating that the EIS is small and less than one. The only exceptions are the US, which contains only negative values in the confidence interval, and Germany, whose confidence interval cannot exclude one.

The last three columns of Table 2.6 report the same set of results using the moment restriction (2.2) for both the interest rate and the stock return. With four instruments and two assets, the moment restriction has dimension eight. Note that GMM gives point estimates and standard errors that are very small. This appears to be a consequence of weak identification from the low correlation between the instruments and the stock return. The CUE, which is more robust to weak instruments than GMM, gives estimates that are similar to those obtained using just the interest rate. Likewise, the weak instrument robust confidence intervals based on the S-test are similar to those obtained using just the interest rate. The confidence intervals are actually slightly wider for many of the series. This is not surprising in light of Table 2.5, which showed that the moment restriction implied by the stock return contains little information that is useful for identifying the EIS.

Table 2.6: Heteroskedasticity Robust Estimates of the EIS

The table reports the EIS estimated by two-step GMM and CUE with standard error in parentheses. The 95% confidence interval is constructed from the S-test. $\emptyset$ indicates an empty confidence interval. The instruments are twice lagged nominal interest rate, inflation, consumption growth, and log dividend-price ratio.

		Interest Rate			Interest Rate & Stock Return		
Country	Sample Period	GMM	CUE	95% CI	GMM	CUE	95% CI
USA	1947.3–1998.4	0.05	-0.11	[-0.14,-0.08]	0.00	-0.36	$\emptyset$
		(0.09)	(0.11)		(0.00)	(0.09)	
AUL	1970.3–1998.4	0.09	0.08	[-0.17,0.30]	0.01	0.12	[-0.12,0.35]
		(0.12)	(0.12)		(0.00)	(0.04)	
CAN	1970.3–1999.1	-0.34	-0.33	[-0.77,0.11]	0.02	-0.19	[-0.71,0.34]
		(0.17)	(0.17)		(0.01)	(0.06)	
FR	1970.3–1998.3	-0.12	-0.12	[-0.57,0.36]	0.00	-0.16	[-0.53,0.22]
		(0.14)	(0.14)		(0.00)	(0.05)	
GER	1979.1–1998.3	-0.44	-0.48	[-1.95,1.63]	0.00	-0.55	[-2.06,1.90]
		(0.43)	(0.43)		(0.00)	(0.20)	
ITA	1971.4–1998.1	-0.08	-0.07	[-0.34,0.20]	0.00	-0.09	[-0.46,0.32]
		(0.08)	(0.08)		(0.00)	(0.05)	
JAP	1970.3–1998.4	-0.18	-0.21	[-0.93,0.39]	0.00	-0.25	[-0.81,0.24]
		(0.21)	(0.21)		(0.00)	(0.08)	
NTH	1977.3–1998.4	-0.25	-0.28	[-0.57,0.09]	0.00	-0.25	[-0.59,0.33]
		(0.20)	(0.20)		(0.00)	(0.09)	
SWD	1970.3–1999.2	0.01	0.00	[-0.28,0.28]	0.00	-0.02	[-0.38,0.33]
		(0.09)	(0.09)		(0.00)	(0.01)	
SWT	1976.2–1998.4	-0.39	-0.41	[-1.42,0.50]	-0.22	-0.44	[-2.38,1.13]
		(0.25)	(0.25)		(0.21)	(0.24)	
UK	1970.3–1999.1	0.22	0.28	[-0.45,0.51]	0.00	0.17	[-1.04,0.65]
		(0.12)	(0.12)		(0.00)	(0.07)	
USA	1970.3–1998.4	0.02	-0.09	[-0.14,-0.02]	0.01	-0.05	$\emptyset$
		(0.08)	(0.09)		(0.00)	(0.02)	
SWD	1921–1994	0.00	-0.09	[-0.51,0.51]	0.00	-0.07	[-0.61,0.74]
		(0.10)	(0.10)		(0.00)	(0.04)	
UK	1921–1994	0.25	0.27	[0.01,0.64]	0.03	0.39	[-0.05,0.82]
		(0.09)	(0.09)		(0.01)	(0.08)	
USA	1891–1995	-0.02	-0.01	[-0.21,0.15]	0.00	0.00	$\emptyset$
		(0.06)	(0.06)		(0.00)	(0.00)	

For the US, the model is entirely rejected as indicated by empty confidence intervals. One potential explanation of this result is that the Euler equation for stock return does not hold for the representative consumer because of limited participation in asset markets. Vissing-Jørgensen (2002) has found some evidence for this theory using the Consumer Expenditure Survey.

2.4.6 Implications for the Equity Premium Puzzle

In the power utility model, the coefficient of relative risk aversion is equal to the inverse of the EIS. In that case, the small estimates of the EIS reported in this essay is evidence for large values of risk aversion. For instance, the confidence intervals in Table 2.3 indicate that the EIS is in the range [0,0.5] across the eleven countries, which implies that risk aversion is in the range $[2, \infty]$. While this is consistent with evidence from the large literature on the equity premium puzzle (Mehra and Prescott 1985), I hesitate to draw conclusions about risk aversion based on the estimates of the EIS.

2.5 Conclusion

The econometric lesson to take away from this essay is that weak instruments are relevant in practice and that conventional t-tests can lead to misleading inference. There are now various methods available for handling weak instruments, from the simple pretest based on the first-stage F-statistic to fully robust confidence intervals based on similar tests. These methods are not necessarily a “cure” for weak instruments since the resulting confidence intervals are often uninformative when identification is poor, but they prevent the researcher from making erroneous inferences.

The economic lesson to take away from this essay is that the EIS is small and not significantly different from zero. In particular, the EIS appears to be less than one, which implies that an investor's optimal consumption-wealth ratio is increasing in expected returns. In my preferred estimates, reported in Table 2.3, the upper end of the 95% confidence interval for the EIS is never greater than 0.5 across eleven developed countries. For the US, the value is about 0.2, which remarkably agrees with Hall (1988, p. 350): "My overall conclusion... is that the evidence points in the direction of a low value for the intertemporal elasticity. The value may even be zero and is probably not above .2."

Chapter 3

Efficient Tests of Stock Return Predictability[1]

3.1 Introduction

Numerous studies in the last two decades have asked whether stock returns can be predicted by financial variables such as the dividend-price ratio, the earnings-price ratio, and various measures of the interest rate.[2] The econometric method used in a typical study is an OLS regression of stock returns onto the lag of the financial variable. The main finding of such regressions is that the t-statistic is typically greater than two and sometimes greater than three. Using conventional critical values for the t-test, one would conclude that there is strong evidence for the predictability of returns.

This statistical inference of course relies on first-order asymptotic distribution

[1]This chapter is coauthored by John Y. Campbell.

[2]See, for example, Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988b), Fama and French (1988, 1989), and Hodrick (1992). The focus of these papers, as well as this one, is classical hypothesis testing. Other approaches include out-of-sample forecasting (Goyal and Welch 2003) and Bayesian inference (Kothari and Shanken 1997, Stambaugh 1999).

theory, which implies that the t-statistic is approximately standard normal in large samples. However, both simulation and analytical studies have shown that the large-sample theory provides a poor approximation to the actual finite-sample distribution of test statistics when the predictor variable is persistent and its innovations are highly correlated with returns (see Mankiw and Shapiro (1986a), Elliott and Stock (1994), and Stambaugh (1999)).

To be concrete, suppose the log dividend-price ratio is used to predict returns. Even if we were to know on prior grounds that the dividend-price ratio is stationary, a time series plot (or more formally a unit root test) shows that it is highly persistent, much like a nonstationary process. Since first-order asymptotics fails when the regressor is nonstationary, it provides a poor approximation in finite samples when the regressor is persistent. Elliott and Stock (1994, Table 1) provide Monte Carlo evidence which suggests that the size distortion of the one-sided t-test is approximately 20 percentage points for plausible parameter values and sample size in the dividend-price ratio regression.[3] They propose an alternative asymptotic framework in which the regressor is modeled as having a local-to-unit root, which provides an accurate approximation to the finite-sample distribution.

These econometric problems have led some recent papers to reexamine (and even cast serious doubt on) the evidence for predictability using tests that have the correct size even if the predictor variable is highly persistent or contains a unit root. Torous, Valkanov, and Yan (2001) develop a test procedure, extending the work of Richardson and Stock (1989) and Cavanagh, Elliott, and Stock (1995), and find evidence for predictability at short horizons but not at long horizons. By testing the stationarity of long-horizon returns, Lanne (2002) concludes that stock returns cannot

[3]We report their result for the one-sided t-test at the 10% level when the sample size is 100, the regressor follows an AR(1) with an autoregressive coefficient of 0.975, and the correlation between the innovations to the dependent variable and the regressor is -0.9.

be predicted by a highly persistent predictor variable. Building on the finite-sample theory of Stambaugh (1999), Lewellen (2003) finds some evidence for predictability with valuation ratios.

A difficulty with understanding the rather large literature on predictability is the sheer variety of test procedures that have been proposed, which have led to different conclusions about the predictability of returns. The first contribution of this essay is to provide an understanding of the various test procedures and their empirical implications within the unifying framework of statistical optimality theory. When the degree of persistence of the predictor variable is known, there is a uniformly most powerful (UMP) test conditional on an ancillary statistic. Although the degree of persistence is not known in practice, this provides a useful benchmark for thinking about the relative power advantages of the various test procedures. In particular, Lewellen's (2003) test is UMP when the predictor variable contains a unit root.

Based on the infeasible UMP test, our second contribution is to propose a new test procedure that is computationally simple (i.e. can be implemented with standard regression output) and has good power (i.e. more powerful than the Bonferroni t-test of Cavanagh, Elliott, and Stock (1995)). Our test is asymptotically valid under fairly general assumptions about the dynamics of the predictor variable (i.e. a finite-order autoregression with the largest root less than, equal to, or even greater than one) and the distribution of the innovations (i.e. even heteroskedastic). The intuition for our approach is as follows. A regression of stock returns onto a lagged financial variable has low power because stock returns are extremely noisy. If we can eliminate some of this noise, we can increase the power of the test. When the innovations to returns and the predictor variable are correlated, we can subtract off the part of the innovation to the predictor variable that is correlated with returns to obtain a less noisy dependent variable for our regression. Of course, this procedure requires us to measure the innovation to the predictor variable. When the predictor variable is highly persistent, it is possible to do so in a way that retains power advantages over the conventional regression.

Although tests derived under local-to-unity asymptotics (e.g. Cavanagh, Elliott, and Stock (1995) or the one proposed in this essay) always lead to correct inference, they can be somewhat more difficult to implement than the conventional t-test. A researcher may therefore be interested in knowing when the conventional t-test leads to correct inference. Our third contribution is to develop a simple pretest based on the confidence interval for the largest autoregressive root of the predictor variable. If the confidence interval indicates that the predictor variable is sufficiently stationary, for a given level of correlation between the innovations to returns and the predictor variable, one can proceed with inference based on the t-test with conventional critical values.

Our final contribution is empirical. We apply our methods to annual, quarterly, and monthly US data, looking first at the dividend-price and the smoothed earnings-price ratios. Using the pretest, we find that these valuation ratios are sufficiently persistent for the conventional t-test to be misleading. Using our test that is robust to the persistence problem, we find that the earnings-price ratio reliably predicts returns at all frequencies in the full sample since 1926. The dividend-price ratio also predicts returns at annual frequency, but we cannot reject the null hypothesis at quarterly and monthly frequencies.

In the sub-sample since 1952, we find that the dividend-price ratio predicts returns at all frequencies if its largest autoregressive root is less than or equal to one. However, since statistical tests do not reject an explosive root for the dividend-price ratio, we have evidence for return predictability only if we are willing to rule out an explosive root based on prior knowledge. This reconciles the "contradictory" findings by Torous, Valkanov, and Yan (2001, Table 3), who report that the dividend-price ratio does not predict monthly returns in the postwar sample, and Lewellen (2003, Table 2), who reports strong evidence for predictability.

Finally, we consider the short-term nominal interest rate and the long-short yield spread as predictor variables in the period since 1952. Our pretest indicates that the conventional t-test is valid for these interest rate variables since their innovations have low correlation with returns. Using either the conventional t-test or our more generally valid test procedure, we find strong evidence that these variables predict returns.

The rest of the essay is organized as follows. In Section 3.2, we derive the UMP test of predictability when the degree of persistence of the predictor variable is known. We then compare its power to that of the conventional t-test under first-order asymptotics. Although first-order asymptotics is not applicable for persistent predictor variables, the calculations provide intuition for the UMP test in a familiar framework. In Section 3.3, we briefly review local-to-unity asymptotics in the context of predictive regressions (in order to provide a self-contained treatment), then compare the power of various tests of predictability. We find that a feasible version of the UMP test has good power. We also introduce the pretest for determining when the conventional t-test leads to correct inference. In Section 3.4, we apply our test procedure to US equity data and reexamine the empirical evidence for predictability. We reinterpret previous empirical studies within our unifying framework. Section 3.5 concludes. Appendix A.6 contains tables necessary for implementing the econometric methods in this essay.

3.2 Predictive Regressions

3.2.1 The Regression Model

Let r_t denote the excess stock return in period t, and let x_{t-1} denote a variable observed at $t-1$ which may have the ability to predict r_t. For instance, x_{t-1} may be the log dividend-price ratio at $t-1$. The regression model that we consider is

$$r_t = \beta x_{t-1} + u_t, \tag{3.1}$$

$$x_t = \rho x_{t-1} + e_t, \tag{3.2}$$

for $t = 1, \ldots, T$ with $x_0 = 0$. β is the unknown coefficient of interest. We say that the variable x_{t-1} has the ability to predict returns if $\beta \neq 0$. For simplicity, we assume that both r_t and x_t have mean zero, so the usual intercept terms do not appear in equations (3.1) and (3.2). In addition, we assume that $(u_t, e_t)'$ is independently and identically distributed (i.i.d.) normal with mean zero and covariance matrix

$$\overline{\Sigma} = \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix}. \tag{3.3}$$

We further assume that the correlation $\delta \leq 0$ between the innovations is known. (The negativity of δ is without loss of generality since the sign of β is unrestricted; redefining the predictor variable as $-x_t$ flips the signs of both β and δ.) We will later relax these assumptions to a more realistic model. For now, this simple model captures the essence of the problem.

In equation (3.2), ρ is the unknown degree of persistence in the variable x_t. If $|\rho| < 1$ and fixed, x_t is integrated of order zero, denoted as I(0). If $\rho = 1$, x_t

is integrated of order one, denoted as I(1). Since β and ρ are the only unknown parameters in the model, (two times) the joint log likelihood is given by

$$L(\beta,\rho) = -\frac{1}{1-\delta^2}\sum_{t=1}^{T}[(r_t-\beta x_{t-1})^2-2\delta(r_t-\beta x_{t-1})(x_t-\rho x_{t-1})+(x_t-\rho x_{t-1})^2], \quad (3.4)$$

up to an additive constant.

The focus of this essay is the null hypothesis $\beta = \beta_0$. We consider two types of alternative hypotheses. The first is the simple alternative $\beta = \beta_1$, and the second is the one-sided (composite) alternative $\beta > \beta_0$. The hypothesis testing problem is complicated by the fact that ρ is an unknown nuisance parameter.

3.2.2 Likelihood Ratio Test

One way to test the hypothesis of interest in the presence of the nuisance parameter ρ is through the likelihood ratio test (LRT). Let $\widehat{\beta}$ be the OLS estimator of β, and let

$$t(\beta_0) = \frac{\widehat{\beta}-\beta_0}{(\sum_{t=1}^{T} x_{t-1}^2)^{-1/2}} \quad (3.5)$$

be the associated t-statistic. The LRT rejects the null if

$$\max_{\beta,\rho} L(\beta,\rho) - \max_{\rho} L(\beta_0,\rho) = t(\beta_0)^2 > C, \quad (3.6)$$

for some constant C. (With a slight abuse of notation, we use C to denote a generic constant throughout the essay.) In other words, the LRT corresponds to the t-test.

Note that we would obtain the same test (3.6) starting from the marginal likelihood $L(\beta) = -\sum_{t=1}^{T}(r_t - \beta x_{t-1})^2$. The LRT can thus be interpreted as a test that ignores information contained in equation (3.2) of the regression model. Although the LRT is not derived from statistical optimality theory, it has desirable large-sample properties when x_t is I(0) (see Cox and Hinkley (1974, Chapter 9)). For instance, the t-statistic is asymptotically pivotal, that is, its asymptotic distribution does not depend on the nuisance parameter ρ. The t-test is therefore a solution to the hypothesis testing problem when x_t is I(0) and ρ is unknown, provided that the large-sample approximation is sufficiently accurate.

3.2.3 Optimal Test When ρ is Known

The problem with the nuisance parameter can also be resolved if ρ were known *a priori*. Since β is then the only unknown parameter in the likelihood function (3.4), the Neyman-Pearson Lemma implies that the most powerful test against the simple alternative rejects the null if

$$\begin{aligned}(1-\delta^2)(L(\beta_1,\rho)-L(\beta_0,\rho)) &= 2(\beta_1-\beta_0)\sum_{t=1}^{T}x_{t-1}[r_t-\delta(x_t-\rho x_{t-1})] \\ &\quad -(\beta_1^2-\beta_0^2)\sum_{t=1}^{T}x_{t-1}^2 > C. \end{aligned} \tag{3.7}$$

Since the optimal test statistic is a weighted sum of two minimal sufficient statistics with the weights depending on the alternative β_1, there is no UMP test.

However, the second statistic $\sum_{t=1}^{T}x_{t-1}^2$ is ancillary (i.e. its distribution does not depend on β). Hence, it is natural to restrict ourselves to tests that condition on the ancillary statistic. The optimal *conditional* test rejects the null if

$$Q(\beta_0,\rho)=\frac{\sum_{t=1}^{T}x_{t-1}[r_t-\beta_0 x_{t-1}-\delta(x_t-\rho x_{t-1})]}{(1-\delta^2)^{1/2}(\sum_{t=1}^{T}x_{t-1}^2)^{1/2}} > C. \tag{3.8}$$

Since the test takes the same form for all $\beta_1 > \beta_0$, it is UMP against one-sided alternatives when ρ is known. For simplicity, we will refer to this (infeasible) test as the Q-test.

When $\beta_0 = 0$, $Q(\beta_0, \rho)$ is the t-statistic that results from regressing $r_t - \delta(x_t - \rho x_{t-1})$ onto x_{t-1}. It collapses to the conventional t-statistic (3.5) when $\delta = 0$. Since $e_t = x_t - \rho x_{t-1}$, knowledge of ρ allows us to subtract off the part of innovation to returns that is correlated with the innovation to the predictor variable, resulting in a more powerful test. If we let $\widehat{\rho}$ denote the OLS estimator of ρ, then the Q-statistic can be written as

$$Q(\beta_0, \rho) = \frac{(\widehat{\beta} - \beta_0) - \delta(\widehat{\rho} - \rho)}{(1 - \delta^2)^{1/2} (\sum_{t=1}^{T} x_{t-1}^2)^{-1/2}}. \tag{3.9}$$

Drawing on the work of Stambaugh (1999), Lewellen (2003) motivates the statistic by interpreting the term $\delta(\widehat{\rho} - \rho)$ as the "finite-sample bias" of the OLS estimator. Assuming that $\rho \leq 1$, Lewellen tests the predictability of returns using the statistic $Q(\beta_0, 1)$.

3.2.4 Power under First-Order Asymptotics

We now derive the power functions of the t-test and the Q-test under first-order asymptotics to illustrate the power gains that would result from incorporating knowledge of the persistence parameter ρ. When x_t is I(0), the OLS estimator $\widehat{\beta}$ is $\sqrt{T}$-consistent. Hence, any reasonable test, such as the conventional t-test, rejects alternatives that are a fixed distance from the null with probability one as the sample size becomes arbitrarily large. In practice, however, we have a finite sample and are interested in the relative efficiency of test procedures. A natural way to evaluate the power of tests in finite samples is to consider their ability to reject local alternatives.[4]

[4]See Lehmann (1999, Chapter 3) for a textbook treatment of local alternatives and relative efficiency.

Formally, we consider a sequence of alternatives of the form $\beta = \beta_0 + b/\sqrt{T}$ for some fixed constant b.

Let $\Phi(z)$ denote one minus the standard normal cumulative distribution function, and let z_α denote the upper α-quantile of that distribution. Under first-order asymptotics, the probability that the t-test rejects a local alternative b is

$$\pi_t(b) = \Phi(z_\alpha - |b|\sigma_x), \tag{3.10}$$

where $\sigma_x^2 = \mathbf{E}[x_t^2] = 1/(1-\rho^2)$. Similarly, the power function of the Q-test is

$$\pi_Q(b) = \Phi\left(z_\alpha - \frac{|b|\sigma_x}{(1-\delta^2)^{1/2}}\right). \tag{3.11}$$

Since the Q-test is UMP against one-sided alternatives conditional on the ancillary statistic $T^{-1}\sum_{t=1}^{T} x_{t-1}^2$, $\pi_Q(b)$ is the power envelope for conditional tests when ρ is known. Moreover, it is asymptotically the power envelope for all (including unconditional) tests since the ancillary statistic has a degenerate asymptotic distribution.

Figure 3.1 shows the power functions for various combinations of ρ (0.99 and 0.75) and δ (-0.95 and -0.75). These values are chosen to correspond to the relevant region of the parameter space when the predictor variable is a valuation ratio (i.e. the log dividend-price ratio or the log earnings-price ratio). As expected, the power of the Q-test dominates that of the t-test. A comparison of (3.10) and (3.11) shows that the power gain arises from $\delta^2 \neq 0$ and is increasing in the degree of correlation. Intuitively, when ρ is known, the innovation $e_t = x_t - \rho x_{t-1}$ is known as well. Then by subtracting off the portion of the innovation u_t that is correlated with e_t (i.e. δe_t), the Q-test is able to gain power through the reduction in noise. When the predictor variable is a valuation ratio, the power gain from using the Q-test is especially large since the

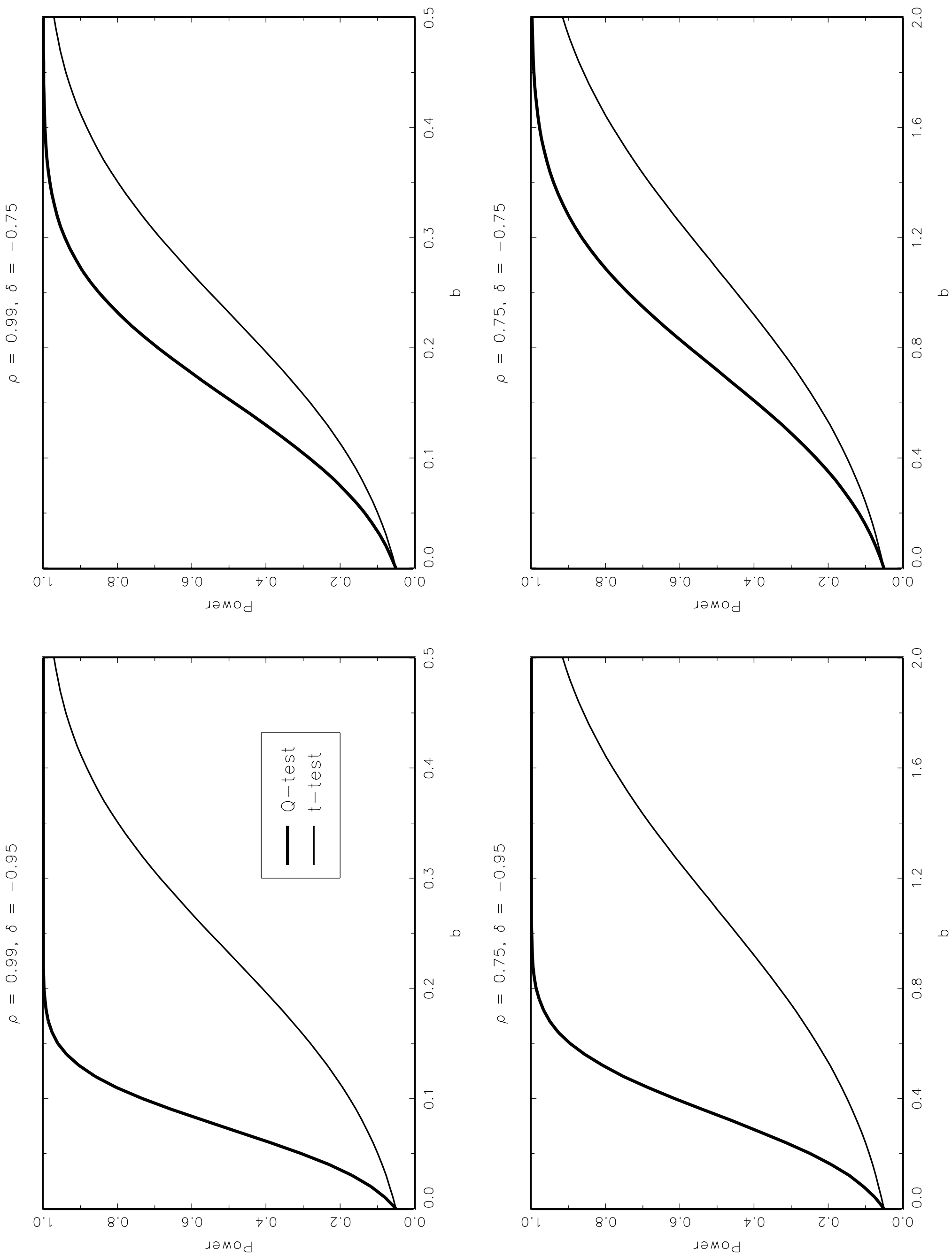

Figure 3.1: Local Asymptotic Power under First-Order Asymptotics

This figure plots the power of the Q-test and the t-test when the predictor variable is an AR(1). The null hypothesis is $\beta = \beta_0$ against the local alternatives $b = \sqrt{T}(\beta - \beta_0) > 0$. $\rho = 0.99, 0.75$ is the autoregressive root of the predictor variable, and $\delta = -0.95, -0.75$ is the correlation between the innovations to returns and the predictor variable.

innovations to returns and the valuation ratio are highly correlated through the stock price. In practice, the Q-test is infeasible because ρ is unknown, so unfortunately, the large power gains over the t-test cannot be realized.

3.2.5 Generalizing the Regression Model

The regression model (3.1)–(3.2) in which the Q-test (3.8) is UMP, conditional on an ancillary statistic, is restrictive. In this section, we show that a generalization of the Q-test is UMP invariant, conditional on an ancillary statistic, in a more realistic and empirically relevant model.

Consider the regression model

$$r_t = \alpha + \beta x_{t-1} + u_t, \tag{3.12}$$

$$x_t = \gamma + \rho x_{t-1} + e_t, \tag{3.13}$$

for $t = 1, \ldots, T$. The model now includes intercept terms assumed away in the simplified model (3.1)–(3.2), account for the mean of returns and the predictor variable. Their magnitudes depend on the units in which the variables are measured. For instance, there is an arbitrary scaling factor involved in computing the dividend-price ratio, which results in an arbitrary constant shifting the level of the log dividend-price ratio. Since we do not want inference to depend on the units in which the variables are measured, it is natural to restrict ourselves to tests that are invariant to translations in α and γ (see Lehmann (1986, Chapter 6)). That is, we only consider test statistics whose values do not change with additive shifts in r_t and x_t.

Now suppose ρ is known and the following distributional assumptions hold.

Assumption 1 (Normality). $w_t = (u_t, e_t)'$ *is independently distributed* $\mathbf{N}(0, \Sigma)$, *where*

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{ue} \\ \sigma_{ue} & \sigma_e^2 \end{bmatrix}$$

is known. x_0 *is fixed and known.*

Let $x_{t-1}^{\mu} = x_{t-1} - T^{-1}\sum_{t=1}^{T} x_{t-1}$ be the de-meaned predictor variable. By the same argument as that in Section 3.2.3, the test based on the statistic

$$Q(\beta_0, \rho) = \frac{\sum_{t=1}^{T} x_{t-1}^{\mu}[r_t - \beta_0 x_{t-1} - \frac{\sigma_{ue}}{\sigma_e^2}(x_t - \rho x_{t-1})]}{\sigma_u (1-\delta^2)^{1/2} (\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}. \tag{3.14}$$

is UMP conditional on the ancillary statistic $\sum_{t=1}^{T} x_{t-1}^{\mu 2}$. In other words, the statistic (3.14) is a generalization of (3.8) to the model (3.12)–(3.13). The asymptotic power function for the Q-test is given by (3.11), when the variances are normalized as $\sigma_u^2 = \sigma_e^2 = 1$. Moreover, it corresponds to the power envelope for tests that use knowledge of the autoregressive root ρ.

3.3 Inference with a Persistent Regressor

Figure 3.2 is a time series plot of the log dividend-price ratio for the CRSP NYSE/AMEX value-weighted index and the log smoothed earnings-price ratio for the S&P 500 index at quarterly frequency. Following Campbell and Shiller (1988b), earnings are smoothed by taking a backwards moving average over ten years. Both valuation ratios are persistent and even appear to be nonstationary, especially toward the end of the sample period. The 95% confidence intervals for ρ are $[0.957, 1.007]$ and $[0.939, 1.000]$ for the dividend-price ratio and the earnings-price ratio, respectively.

The persistence of financial variables typically used to predict returns has important implications for inference about predictability. Even if the predictor variable is

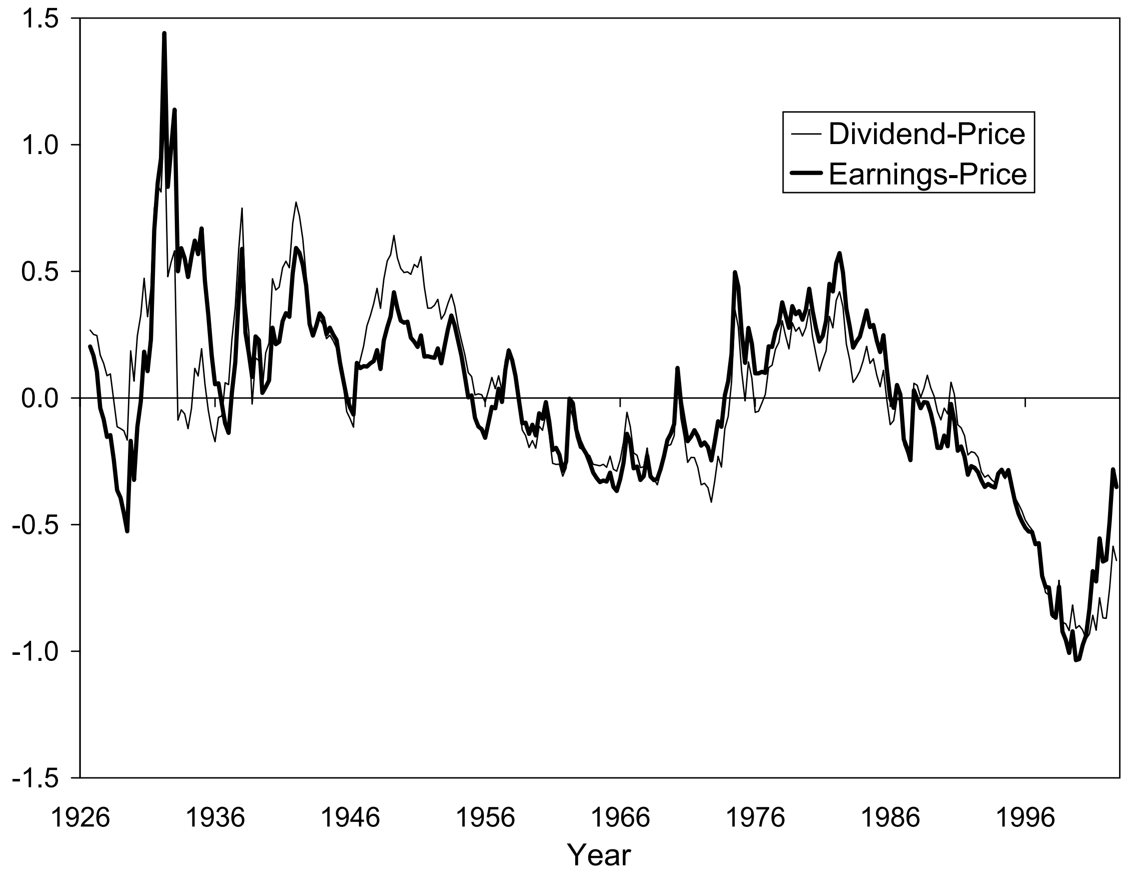

Figure 3.2: Time Series Plot of the Valuation Ratios

This figure plots the log dividend-price ratio for the CRSP value-weighted index and the log earnings-price ratio for the S&P 500. Earnings are smoothed by taking a ten-year moving average. The sample period is 1926:4–2002:4.

I(0), first-order asymptotics can be a poor approximation in finite samples when ρ is close to one because of the discontinuity in the asymptotic distribution at $\rho = 1$. Inference based on first-order asymptotics may therefore be invalid due to size distortions. The solution is to base inference on more accurate approximations to the actual (unknown) sampling distribution of test statistics. There are two main approaches that have been used in the literature.

The first approach is the exact finite-sample theory under the assumption of normality (i.e. Assumption 1). This is the approach taken by Evans and Savin (1981, 1984) for autoregression and Stambaugh (1999) for predictive regressions. The second approach is local-to-unity asymptotics, which has been applied successfully to approximate the finite-sample behavior of persistent time series in the unit root testing literature. (See Stock (1994) for a survey and references.) Local-to-unity asymptotics has been applied to the present context of predictive regressions by Elliott and Stock (1994), who derived the asymptotic distribution of the t-statistic. This has been extended to long-horizon t-tests by Torous, Valkanov, and Yan (2001). The advantage of local-to-unity asymptotics, over the exact Gaussian theory, is that it allows for a wide variety of distributions for the innovations, including short-run dynamics and heteroskedasticity. Hence, this is the approach that we follow in this essay.

3.3.1 Local-to-Unity Asymptotics

The AR(1) model for the predictor variable (3.13) is restrictive since it does not allow for short-run dynamics. Let L be the lag operator. We generalize the model as

$$x_t = \gamma + \rho x_{t-1} + v_t, \tag{3.15}$$

$$b(L)v_t = e_t, \tag{3.16}$$

where $b(L) = \sum_{i=0}^{p-1} b_i L^i$ with $b_0 = 1$ and $b(1) \neq 0$. All the roots of $b(L)$ are assumed to be fixed and less than one in absolute value. Equations (3.15) and (3.16) together imply that

$$\Delta x_t = \tau + \theta x_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta x_{t-i} + e_t. \tag{3.17}$$

Hence, the dynamics of the predictor variable is captured by an AR(p), which is written here in the augmented Dickey-Fuller form.

To apply the asymptotic theory, we assume that the sequence of innovations satisfies the following fairly weak distributional assumptions.

Assumption 2 (Martingale Difference Sequence). *Let $\mathcal{F}_t = \{w_s | s \leq t\}$ be the filtration generated by the process $w_t = (u_t, e_t)'$. Then*

1. $\mathbf{E}[w_t | \mathcal{F}_{t-1}] = 0$,

2. $\mathbf{E}[w_t w_t'] = \Sigma$,

3. $\sup_t \mathbf{E}[u_t^4] < \infty$, $\sup_t \mathbf{E}[e_t^4] < \infty$, *and* $\mathbf{E}[x_0^2] < \infty$.

In other words, w_t is a martingale difference sequence with finite fourth moments. The assumption allows the sequence of innovations to be conditionally heteroskedastic as long as it is covariance stationary (i.e. unconditionally homoskedastic). Assumption 1 is a special case when the innovations are i.i.d. normal and the covariance matrix Σ is known.

Local-to-unity asymptotics is an asymptotic framework where the largest autoregressive root is modeled as $\rho = 1 + c/T$ with c a fixed constant. Within this framework, the asymptotic distribution theory is not discontinuous when x_t is I(1) (i.e. $c = 0$). This device also allows x_t to be stationary but nearly integrated (i.e. $c < 0$) or even explosive (i.e. $c > 0$). For the rest of the essay, we assume that the true processes for excess returns and the predictor variable are (3.12) and (3.15), respectively, where $c = T(\rho - 1)$ is fixed as T becomes arbitrarily large. Without loss of generality, we continue to assume that $\delta = \sigma_{ue}/(\sigma_u \sigma_e) \leq 0$.

An important feature of the nearly integrated case is that sample moments (e.g. mean and variance) of the process x_t are not well defined. However, when appropriately scaled, these objects converge to functionals of a diffusion process. Let $(W_u(s), W_e(s))'$ be a two-dimensional Weiner process with covariance matrix (3.3). Let $J_c(s)$ be the diffusion process defined by the stochastic differential equation $dJ_c(s) = cJ_c(s)ds + dW_e(s)$ with initial condition $J_c(0) = 0$. Let $J_c^\mu(s) = J_c(s) - \int J_c(r)dr$, where integration is over $[0, 1]$ unless otherwise noted. Let $\Rightarrow$ denote weak convergence in the space $D[0, 1]$ of cadlag functions (see Billingsley (1999, Chapter 3)). Collecting results from Phillips (1987, Lemma 1) and Cavanagh, Elliott, and Stock (1995), we have the following useful lemma.

Lemma 1 (Weak Convergence). *Suppose Assumption 2 holds. The following limits hold jointly.*

1. $T^{-3/2} \sum_{t=1}^T x_t^\mu \Rightarrow \omega \int J_c^\mu(s)ds$,

2. $T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2} \Rightarrow \omega^2 \int J_c^\mu(s)^2 ds$,

3. $T^{-1} \sum_{t=1}^T x_{t-1}^\mu v_t \Rightarrow \omega^2 \int J_c^\mu(s)dW_e(s) + \frac{1}{2}(\omega^2 - \sigma_v^2)$,

4. $T^{-1} \sum_{t=1}^T x_{t-1}^\mu u_t \Rightarrow \sigma_u \omega \int J_c^\mu(s)dW_u(s)$,

where $\omega = \sigma_e/b(1)$ *and* $\sigma_v^2 = \mathbf{E}[v_t^2]$.

Under first-order asymptotics, the t-statistic from regression (3.12) has a standard normal asymptotic distribution. Under local-to-unity asymptotics, the t-statistic has the null distribution

$$t(\beta_0) \Rightarrow \delta \frac{\tau_c}{\kappa_c} + (1 - \delta^2)^{1/2} Z, \tag{3.18}$$

where $\kappa_c = (\int J_c^\mu(s)^2 ds)^{1/2}$, $\tau_c = \int J_c^\mu(s) dW_e(s)$, and Z is a standard normal random variable independent of $(W_e(s), J_c(s))$ (see Elliott and Stock (1994)). Note that the t-statistic is no longer asymptotically pivotal. That is, its asymptotic distribution depends on an unknown nuisance parameter c through the random variable τ_c/κ_c, which makes the test infeasible.

When the predictor variable is an AR(1) (i.e. model (3.13)), the Q-statistic (3.14) has a standard normal asymptotic distribution under the null. Under the more general model (3.15) which allows for higher-order autocorrelation, the statistic (3.14) is not asymptotically pivotal. However, a suitably modified statistic

$$Q(\beta_0, \rho) = \frac{\sum_{t=1}^T x_{t-1}^\mu [r_t - \beta_0 x_{t-1} - \frac{\sigma_{ue}}{\sigma_e \omega}(x_t - \rho x_{t-1})] + \frac{T}{2}\frac{\sigma_{ue}}{\sigma_e \omega}(\omega^2 - \sigma_v^2)}{\sigma_u (1-\delta^2)^{1/2} (\sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}} \tag{3.19}$$

has a standard normal asymptotic distribution (see Appendix A.5 for details). In the absence of short-run dynamics (i.e. $b(1) = 1$ so $\omega^2 = \sigma_v^2 = \sigma_e^2$), the Q-statistic reduces to (3.14). The correction term involving $(\omega^2 - \sigma_v^2)$ is analogous to the correction of the Dickey-Fuller (1981) test by Phillips and Perron (1988).

Although the Q-statistic is asymptotically pivotal, the test is infeasible since it requires knowledge of ρ (or equivalently c) to compute the test statistic. Even if ρ were known, the statistic (3.19) also requires knowledge of the nuisance parameters Σ, ω^2, and σ_v^2. However, a feasible version of the statistic that replaces these nuisance parameters with consistent estimators has the same asymptotic distribution. Therefore, there is no loss of generality in assuming knowledge of these parameters for the purposes of asymptotic theory.

3.3.2 Relation to First-Order Asymptotics and a Simple Pretest

In this section, we first discuss the relationship between first-order and local-to-unity asymptotics. We then develop a simple pretest to determine whether inference based on first-order asymptotics is reliable.

In general, the asymptotic distribution of the t-statistic (3.18) is nonstandard because of its dependence on τ_c/κ_c. However, the t-statistic is standard normal in the special case $\delta = 0$. The t-statistic should therefore be approximately normal when $\delta \approx 0$. Likewise, the t-statistic should be approximately normal when $c \ll 0$ because first-order asymptotics is a satisfactory approximation when the predictor variable is stationary. Formally, Phillips (1987, Theorem 2) shows that $\tau_c/\kappa_c \Rightarrow \widetilde{Z}$ as $c \to -\infty$, where $\widetilde{Z}$ is a standard normal random variable independent of Z.

Figure 3.3 is a plot of the actual size of the nominal 5% one-sided t-test as a function of c and δ. In other words, we plot

$$p(c, \delta; \alpha) = \Pr\left(\delta \frac{\tau_c}{\kappa_c} + (1 - \delta^2)^{1/2} Z > z_\alpha\right), \tag{3.20}$$

where $\alpha = 0.05$. The t-test that uses conventional critical values has approximately the correct size when δ is small in absolute value or c is large in absolute value.[5] The size distortion of the t-test peaks when $\delta = -1$ and $c \approx 1$. The size distortion arises from the fact that the distribution of τ_c/κ_c is skewed to the left, which causes the distribution of the t-statistic to be skewed to the right when $\delta < 0$. This causes a right-tailed t-test that uses conventional critical values to over-reject, and a left-tailed test to under-reject. When the predictor variable is a valuation ratio (e.g.

[5] The fact that the t-statistic is approximately normal for $c \gg 0$ corresponds to asymptotic results for explosive AR(1) with Gaussian errors. See Phillips (1987) for a discussion.

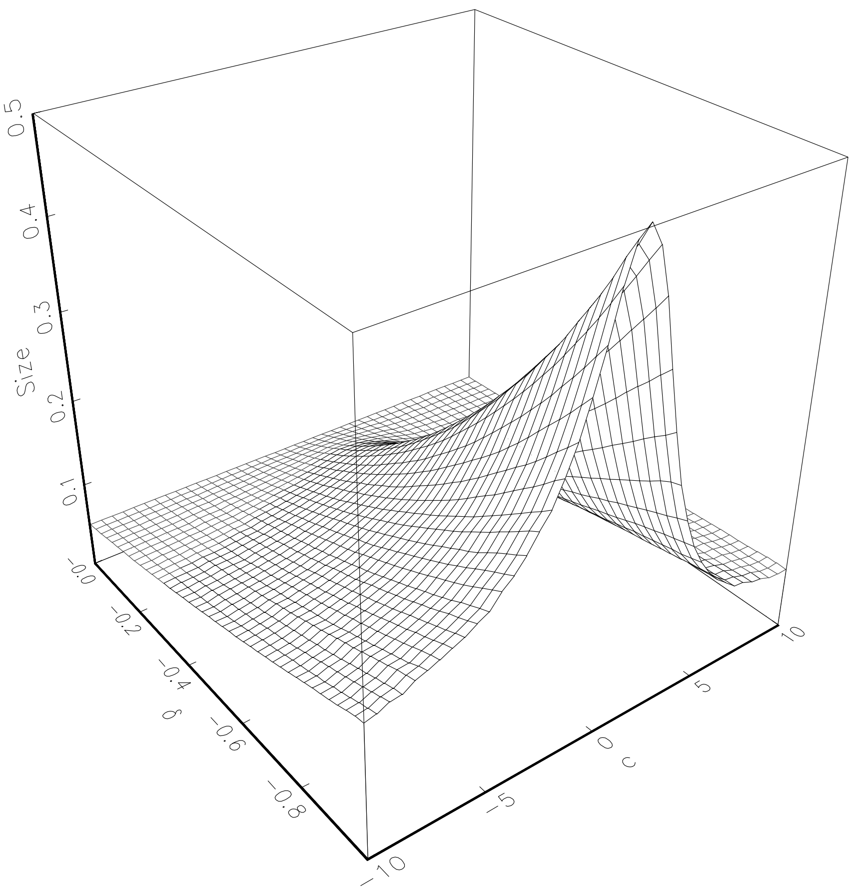

Figure 3.3: Asymptotic Size of the One-Sided t-test at 5% Significance

This figure plots the actual size of the nominal 5% t-test when the largest autoregressive root of the predictor variable is $\rho = 1 + c/T$. The null hypothesis is $\beta = \beta_0$ against the one-sided alternative $\beta > \beta_0$. δ is the correlation between the innovations to returns and the predictor variable. The dark shade indicates regions where the size is greater than 7.5%.

the dividend-price ratio), $\delta \approx -1$ and the hypothesis of interest is $\beta = 0$ against the alternative $\beta > 0$. Thus we may worry that the evidence for predictability is a consequence of size distortion.

In Table 3.1, we tabulate the values of $c \in (c_{\min}, c_{\max})$ for which the size of the right-tailed t-test exceeds 7.5%, for selected values of δ. For instance, when $\delta = -0.95$, the nominal 5% t-test has actual size greater than 7.5% if $c \in (-79.318, 8.326)$. The table can be used to construct a pretest to determine whether inference based on the conventional t-test is sufficiently reliable. Suppose a researcher is willing to tolerate an actual size of up to 7.5% for a nominal 5% test of predictability. Let $\Theta = \{c, \delta | p(c, \delta; 0.05) > 0.075\}$. Then the goal is to test

$$H_0: \quad \{c, \delta\} \in \Theta$$
$$H_1: \quad \{c, \delta\} \notin \Theta.$$

To test this hypothesis, we first construct a $100(1 - \alpha_1)\%$ confidence interval for c, denoted as $C_c(\alpha_1)$. We then estimate δ using the residuals from regressions (3.12) and (3.17). We reject null hypothesis if $C_c(\alpha_1) \bigcap (c_{\min}, c_{\max}) = \emptyset$, where $(c_{\min}, c_{\max})$ is taken from Table 3.1 using the estimated correlation $\widehat{\delta}$. That is, we reject the null if the confidence interval for c lies strictly below (or above) the region of the parameter space $(c_{\min}, c_{\max})$ where size distortion is large. As emphasized by Elliott and Stock (1994), the rejection of the unit root hypothesis $c = 0$ is not sufficient to assure that the size distortion is acceptably small. Asymptotically, this pretest has size α_1.

Since there is no UMP test for an autoregressive unit root (see Elliott, Rothenberg, and Stock (1996)), there is no uniformly most accurate confidence interval for c. However, as discussed in Elliott and Stock (2001), a relatively accurate confidence interval can be constructed by using a relatively efficient unit root test. In our em-

Table 3.1: Parameters Leading to Size Distortion of the One-Sided t-test

This table reports the regions of the parameter space where the actual size of the nominal 5% t-test is greater than 7.5%. The null hypothesis is $\beta = \beta_0$ against the alternative $\beta > \beta_0$. For a given δ, the size of the t-test is greater than 7.5% if $c \in (c_{\min}, c_{\max})$. Size is less than 7.5% for all c if $\delta \leq -0.125$.

δ	$c_{\min}$	$c_{\max}$	δ	$c_{\min}$	$c_{\max}$
-1.000	-83.088	8.537	-0.550	-28.527	6.301
-0.975	-81.259	8.516	-0.525	-27.255	6.175
-0.950	-79.318	8.326	-0.500	-25.942	6.028
-0.925	-76.404	8.173	-0.475	-23.013	5.868
-0.900	-69.788	7.977	-0.450	-19.515	5.646
-0.875	-68.460	7.930	-0.425	-17.701	5.435
-0.850	-63.277	7.856	-0.400	-14.809	5.277
-0.825	-59.563	7.766	-0.375	-13.436	5.111
-0.800	-58.806	7.683	-0.350	-11.884	4.898
-0.775	-57.618	7.585	-0.325	-10.457	4.682
-0.750	-51.399	7.514	-0.300	-8.630	4.412
-0.725	-50.764	7.406	-0.275	-6.824	4.184
-0.700	-42.267	7.131	-0.250	-5.395	3.934
-0.675	-41.515	6.929	-0.225	-4.431	3.656
-0.650	-40.720	6.820	-0.200	-3.248	3.306
-0.625	-36.148	6.697	-0.175	-1.952	2.800
-0.600	-33.899	6.557	-0.150	-0.614	2.136
-0.575	-31.478	6.419	-0.125	—	—

pirical application, we therefore construct the confidence interval for c by applying Stock's (1991) method of confidence belts to the DF-GLS test (Elliott, Rothenberg, and Stock 1996). A lookup table for the confidence interval for c, given the value of the DF-GLS statistic, is provided in Appendix A.6.

3.3.3 Feasible Tests of Predictability

As discussed in Section 3.3.1, both the t-test and the Q-test are infeasible since the procedures depend on an unknown nuisance parameter c, which cannot be estimated consistently. Intuitively, the degree of persistence, controlled by the parameter c, influences the distribution of test statistics that depend on the persistent predictor variable. This must be accounted for by adjusting either the critical values of the test (e.g. t-test) or the value of the test statistic itself (e.g. Q-test). Cavanagh, Elliott, and Stock (1995) discuss several (sup-bound, Bonferroni, and Scheffe-type) methods of making tests that depend on c feasible. Here, we will discuss the Bonferroni method.

To construct a Bonferroni confidence interval, we first construct a $100(1-\alpha_1)\%$ confidence interval for ρ, denoted as $C_\rho(\alpha_1)$. (We parameterize the degree of persistence by ρ rather than c since this is the more natural choice in the following.) Then for each value of ρ in the confidence interval, we construct a $100(1-\alpha_2)\%$ confidence interval for β given ρ, denoted as $C_{\beta|\rho}(\alpha_2)$. A confidence interval that does not depend on ρ can be obtained by

$$C_\beta(\alpha) = \bigcup_{\rho \in C_\rho(\alpha_1)} C_{\beta|\rho}(\alpha_2).$$

By Bonferroni's inequality, this confidence interval has coverage of at least $100(1-\alpha)\%$, where $\alpha = \alpha_1 + \alpha_2$.

This approach is conservative in the sense that the actual coverage rate of $C_\beta(\alpha)$

can be greater than $100(1-\alpha)\%$. This can be seen from the equality

$$\begin{aligned}\Pr(\beta \notin C_\beta(\alpha)) &= \Pr(\beta \notin C_\beta(\alpha)|\rho \in C_\rho(\alpha_1)) \Pr(\rho \in C_\rho(\alpha_1)) \\ &\quad + \Pr(\beta \notin C_\beta(\alpha)|\rho \notin C_\rho(\alpha_1)) \Pr(\rho \notin C_\rho(\alpha_1)).\end{aligned}$$

Since $\Pr(\beta \notin C_\beta(\alpha)|\rho \notin C_\rho(\alpha_1))$ is unknown, the Bonferroni method bounds it by one as the worst case. In addition, the inequality $\Pr(\beta \notin C_\beta(\alpha)|\rho \in C_\rho(\alpha_1)) \leq \alpha_2$ is strict unless the conditional confidence intervals $C_{\beta|\rho}(\alpha_2)$ do not depend on ρ. Because these worst case conditions are unlikely to hold in practice, the inequality

$$\Pr(\beta \notin C_\beta(\alpha)) \leq \alpha_2(1-\alpha_1) + \alpha_1 \leq \alpha$$

is likely to be strict, resulting in a conservative confidence interval.

To implement the Bonferroni confidence interval in practice, Cavanagh, Elliott, and Stock suggest inverting the augmented Dickey-Fuller t-statistic to first construct $C_\rho(\alpha_1)$. They then suggest inverting the conventional t-statistic for testing β, using the appropriate critical values based on its asymptotic distribution (3.18). The two t-statistics are correlated, which tends to increase the coverage rate of the confidence interval. Cavanagh, Elliott, and Stock suggest adjusting α_1 and α_2 to achieve an exact test of the desired significance level. Torous, Valkanov, and Yan (2001) have applied this method to test predictability in US data.

A natural question that arises is whether there is a more efficient method of constructing the Bonferroni confidence interval. Since there is no UMP test for an autoregressive unit root, there is no uniformly most accurate confidence interval for ρ. However, the DF-GLS test is more powerful than the augmented Dickey-Fuller test. Hence, we invert the DF-GLS statistic to obtain a tighter confidence interval for ρ.

In addition, we know that the Q-test is a more powerful test of β given ρ than the t-test. In fact, it is UMP conditional on an ancillary statistic when ρ is known. We can therefore obtain a more accurate confidence interval $C_{\beta|\rho}(\alpha_2)$ by inverting the Q-test. Because the statistic (3.19) is asymptotically normal under the null, an equal-tailed α_2-level confidence interval is simply $C_{\beta|\rho}(\alpha_2) = [\underline{\beta}(\rho, \alpha_2), \overline{\beta}(\rho, \alpha_2)]$, where

$$\widehat{\beta}(\rho) = \frac{\sum_{t=1}^{T} x_{t-1}^{\mu}[r_t - \frac{\sigma_{ue}}{\sigma_e \omega}(x_t - \rho x_{t-1})] + \frac{T}{2}\frac{\sigma_{ue}}{\sigma_e \omega}(\omega^2 - \sigma_v^2)}{\sum_{t=1}^{T} x_{t-1}^{\mu 2}}, \tag{3.21}$$

$$\underline{\beta}(\rho, \alpha_2) = \widehat{\beta}(\rho) - z_{\alpha_2/2}\sigma_u \left(\frac{1-\delta^2}{\sum_{t=1}^{T} x_{t-1}^{\mu 2}} \right)^{1/2}, \tag{3.22}$$

$$\overline{\beta}(\rho, \alpha_2) = \widehat{\beta}(\rho) + z_{\alpha_2/2}\sigma_u \left(\frac{1-\delta^2}{\sum_{t=1}^{T} x_{t-1}^{\mu 2}} \right)^{1/2}. \tag{3.23}$$

Let $C_\rho(\alpha_1) = [\underline{\rho}(\underline{\alpha}_1), \overline{\rho}(\overline{\alpha}_1)]$ denote the confidence interval for ρ, where $\underline{\alpha}_1 = \Pr(\rho < \underline{\rho}(\underline{\alpha}_1))$, $\overline{\alpha}_1 = \Pr(\rho > \overline{\rho}(\overline{\alpha}_1))$, and $\alpha_1 = \underline{\alpha}_1 + \overline{\alpha}_1$. Then the Bonferroni confidence interval is given by

$$C_\beta(\alpha) = [\underline{\beta}(\overline{\rho}(\overline{\alpha}_1), \alpha_2), \overline{\beta}(\underline{\rho}(\underline{\alpha}_1), \alpha_2)]. \tag{3.24}$$

Hence, we have a closed-form expression for the confidence interval of β that is easy to compute.

As discussed above, the Bonferroni confidence interval can be quite conservative. As suggested by Cavanagh, Elliott, and Stock, the significance levels α_1 and α_2 can be adjusted to achieve a test of desired significance level $\widetilde{\alpha} \leq \alpha$. To do so, we first fix α_2. Then for each δ, we numerically search over a grid for c to find the $\overline{\alpha}_1$ such that

$$\Pr(\underline{\beta}(\overline{\rho}(\overline{\alpha}_1), \alpha_2) > \beta) \leq \widetilde{\alpha}/2, \tag{3.25}$$

with equality at some c. We then repeat the same procedure for $\underline{\alpha}_1$ and

$$\Pr(\overline{\beta}(\underline{\rho}(\underline{\alpha}_1), \alpha_2) < \beta) \leq \widetilde{\alpha}/2. \tag{3.26}$$

In Table 3.2, we report the values of $\underline{\alpha}_1$ and $\overline{\alpha}_1$ for selected values of δ when $\widetilde{\alpha} = \alpha_2 = 0.10$, computed over the grid $c \in [-50, 5]$. The table can be used to construct a 10% Bonferroni confidence interval for β (equivalently, a 5% one-sided Q-test for predictability). Note that $\underline{\alpha}_1$ and $\overline{\alpha}_1$ are increasing in δ, so the Bonferroni inequality has more slack and the unadjusted Bonferroni test is more conservative the smaller is δ in absolute value. In order to implement the Bonferroni test using Table 3.2, one needs the confidence belts for the DF-GLS statistic. In Campbell (2003, Tables 2–11), we provide lookup tables which report the appropriate confidence interval for c, $C_c(\alpha_1) = [\underline{c}(\underline{\alpha}_1), \overline{c}(\overline{\alpha}_1)]$, given the values of the DF-GLS statistic and δ. Then the confidence interval $C_\rho(\alpha_1) = 1 + C_c(\alpha_1)/T$ for ρ results in a 10% Bonferroni confidence interval for β.

Our computational results indicate that in general the inequalities (3.25) and (3.26) are close equalities when c is large and more slack when c is small. For For right-tailed tests, the probability (3.25) can be small as 4.0% for some values of c and δ. For left-tailed tests, the probability (3.26) can be as small as 1.2%. This suggests that even the adjusted Bonferroni Q-test is conservative (i.e. undersized) when $c < 5$. The assumption that the predictor variable is never explosive (i.e. $c \leq 0$) allows us to further tighten the Bonferroni confidence interval. In our judgment, however, the magnitude of the resulting power gain is not sufficient to justify the loss of robustness against explosive roots.

Table 3.2: Significance Level of the DF-GLS Confidence Interval for the Bonferroni Q-test

This table reports the significance level of the confidence interval for the largest autoregressive root ρ, computed by inverting the DF-GLS test, which sets the size of the one-sided Bonferroni Q-test to 5%. Using the notation (3.24), the confidence interval $C_\rho(\alpha_1) = [\underline{\rho}(\underline{\alpha}_1), \overline{\rho}(\overline{\alpha}_1)]$ for ρ results in a 90% Bonferroni confidence interval $C_\beta(0.1)$ for β when $\alpha_2 = 0.1$.

δ	$\underline{\alpha}_1$	$\overline{\alpha}_1$	δ	$\underline{\alpha}_1$	$\overline{\alpha}_1$
-0.999	0.050	0.055	-0.500	0.080	0.280
-0.975	0.055	0.080	-0.475	0.085	0.285
-0.950	0.055	0.100	-0.450	0.085	0.295
-0.925	0.055	0.115	-0.425	0.090	0.310
-0.900	0.060	0.130	-0.400	0.090	0.320
-0.875	0.060	0.140	-0.375	0.095	0.330
-0.850	0.060	0.150	-0.350	0.100	0.345
-0.825	0.060	0.160	-0.325	0.100	0.355
-0.800	0.065	0.170	-0.300	0.105	0.360
-0.775	0.065	0.180	-0.275	0.110	0.370
-0.750	0.065	0.190	-0.250	0.115	0.375
-0.725	0.065	0.195	-0.225	0.125	0.380
-0.700	0.070	0.205	-0.200	0.130	0.390
-0.675	0.070	0.215	-0.175	0.140	0.395
-0.650	0.070	0.225	-0.150	0.150	0.400
-0.625	0.075	0.230	-0.125	0.160	0.405
-0.600	0.075	0.240	-0.100	0.175	0.415
-0.575	0.075	0.250	-0.075	0.190	0.420
-0.550	0.080	0.260	-0.050	0.215	0.425
-0.525	0.080	0.270	-0.025	0.250	0.435

3.3.4 Power under Local-to-Unity Asymptotics

Infeasible Tests

In this section, we first examine the power of the t-test and Q-test under local-to-unity asymptotics. Although these tests assume knowledge of c and are thus infeasible, their power provide benchmarks for assessing the power of feasible tests. When the predictor variable contains a local-to-unit root, OLS estimators $\widehat{\beta}$ and $\widehat{\rho}$ are consistent at the rate T, rather than $\sqrt{T}$. We therefore consider a sequence of alternatives of the form $\beta = \beta_0 + b/T$ for some fixed constant b. Details on the computation of the power functions are in Appendix A.5.

Figure 3.4 plots the power functions for the t-test (using the appropriate critical value that depends on c) and the Q-test. Under local-to-unity asymptotics, power functions are not symmetric in b. We only report the power for right-tailed tests (i.e. $b > 0$) since this is the region where the conventional t-test is size distorted (recall the discussion in Section 3.3.2). The results, however, are qualitatively similar for left-tailed tests. We consider various combinations of c (-2 and -20) and δ (-0.95 and -0.75), which are in the relevant region of the parameter space when the predictor variable is a valuation ratio. The nuisance parameters are normalized as $\sigma_u^2 = \omega^2 = 1$.

As expected, the power function for the Q-test dominates that for the t-test. The difference is especially large when $\delta = -0.95$. When the correlation between the innovations is large, there are large power gains from subtracting off the part of the innovation to returns that is correlated with the innovation to the predictor variable. These results confirm analogous calculations under first-order asymptotics, reported in Figure 3.1.

To assess the importance of the power gain, we compute the Pitman efficiency, which is the ratio of the sample sizes at which two tests achieve the same level of

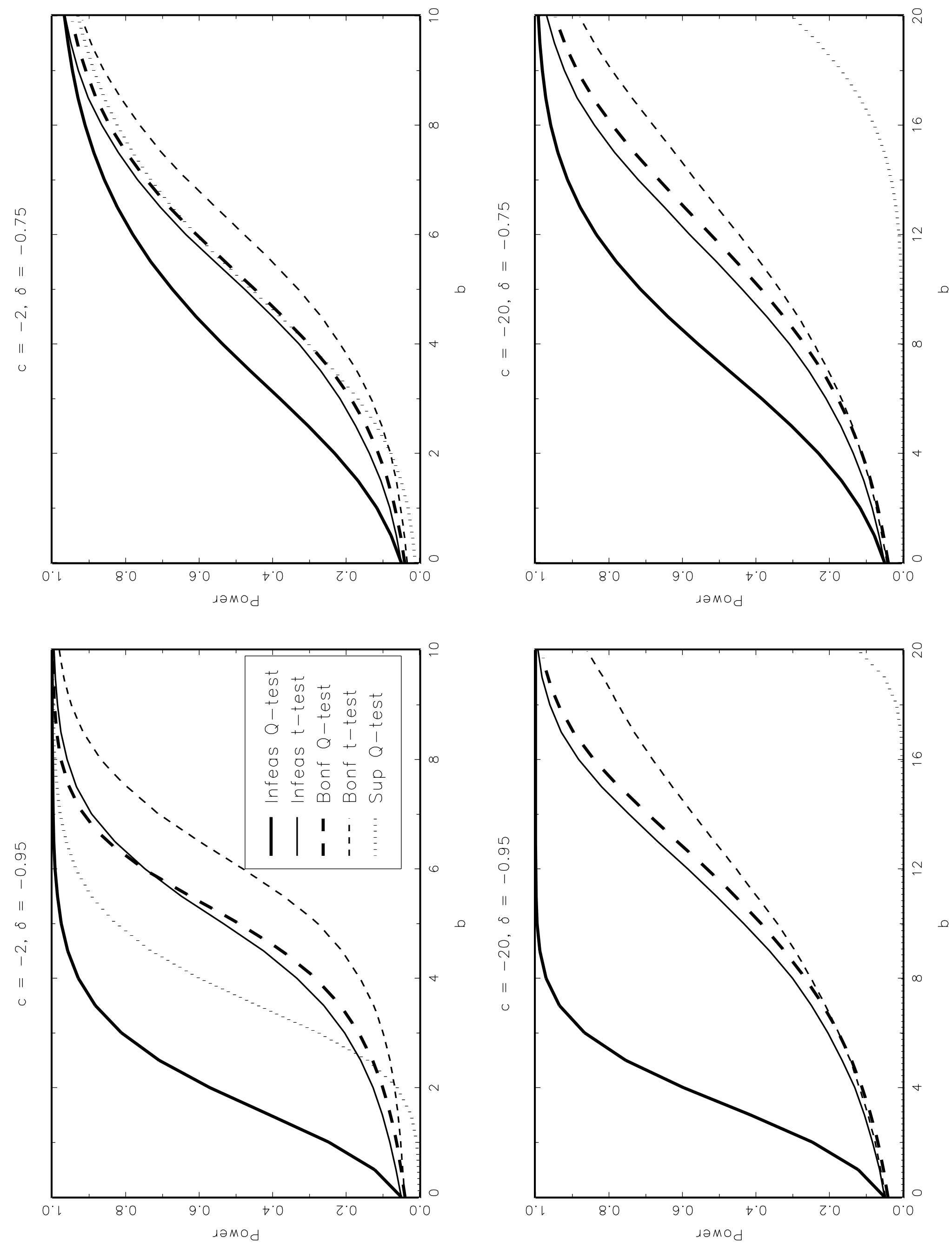

Figure 3.4: Local Asymptotic Power under Local-to-Unity Asymptotics

This figure plots the power of the infeasible Q-test and t-test that assume knowledge of the local-to-unity parameter, the Bonferroni Q-test and t-test, and the sup-bound Q-test. The null hypothesis is $\beta = \beta_0$ against the local alternatives $b = T(\beta - \beta_0) > 0$. $c = -2, -20$ is the local-to-unity parameter, and $\delta = -0.95, -0.75$ is the correlation between the innovations to returns and the predictor variable.

power (e.g. 50%) along a sequence of local alternatives. Consider the case $c = -2$ and $\delta = -0.95$. To compute the Pitman efficiency of the t-test relative to the Q-test, note first that the t-test achieves 50% power when $b = 4.8$. On the other hand, the Q-test achieves 50% power when $b = 1.8$. Following the discussion in Stock (1994, p. 2775), the Pitman efficiency of the t-test relative to the Q-test is $4.8/1.8 \approx 2.7$. This means that to achieve 50% power, the t-test asymptotically requires 170% more observations than the Q-test.

As was the case under first-order asymptotics, the power function for the Q-test corresponds to the Gaussian power envelope for conditional tests under model (3.13) and Assumption 1. Unlike the case for first-order asymptotics, however, the power function does not correspond to the power envelope for all (including unconditional) tests. This is because the ancillary statistic $T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}$ has a non-degenerate asymptotic distribution under local-to-unity asymptotics (see Lemma 1). The optimal test statistic is therefore a weighted sum of two statistics, where the weights depend on the alternative b. Consequently, there is no UMP test against a one-sided alternative. Although not reported in Figure 3.4, our calculations show that the power of the Q-test comes very close (essentially tangent) to the power envelope.

When the innovations are not normal, there are in principle tests that more efficient than the Q-test. However, the Q-test is asymptotically more efficient than the t-test even if the innovations are non-normal (i.e. Assumption 2). This illustrates the fact that the Gaussian likelihood function and the Neyman-Pearson Lemma are useful tools for deriving efficient tests even if the error distribution is unknown.

Feasible Tests

We now analyze the power properties of several feasible tests that have been proposed. Figure 3.4 reports the power of the Bonferroni t-test (Cavanagh, Elliott, and Stock

1995) and the Bonferroni Q-test described in the last section.[6]

In all cases considered, the Bonferroni Q-test dominates the Bonferroni t-test. In fact, the power of the Bonferroni Q-test comes very close to that of the infeasible t-test. The power gains of the Bonferroni Q-test over the Bonferroni t-test are larger the closer is c to zero and the larger is δ in absolute value. When $c = -2$ and $\delta = -0.95$, the Pitman efficiency is 1.2, which means that the Bonferroni t-test requires 20% more observations than the Bonferroni Q-test to achieve 50% power.

In addition to the Bonferroni tests, we also consider the power of Lewellen's (2003) test which is the Q-test that assumes $\rho = 1$. In our notation (3.24), Lewellen's confidence interval corresponds to $[\underline{\beta}(1, \alpha_2), \overline{\beta}(1, \alpha_2)]$. This test can be interpreted as a sup-bound Q-test, provided that the parameter space is restricted to $c \in (-\infty, 0]$, since $Q(\beta_0, \rho)$ is decreasing in ρ when $\delta < 0$. By construction, the sup-bound Q-test is the most powerful test when $c = 0$. When $c = -2$ and $\delta = -0.95$, the sup-bound Q-test is undersized when b is small and has good power when $b \gg 0$. When $c = -2$ and $\delta = -0.75$, the power of the sup-bound Q-test is close to that of the Bonferroni Q-test. When $c = -20$, the sup-bound Q-test has very poor power. In some sense, the comparison of the sup-bound Q-test with the Bonferroni tests is unfair because the size of the sup-bound test is greater than 5% when the true autoregressive root is explosive (i.e. $c > 0$), while the Bonferroni tests have the correct size even in the presence of explosive roots.

Against left-sided local alternatives (i.e. $b < 0$), the sup-bound t-test, which is the t-test that uses conventional critical values, has correct albeit conservative size. (Recall from Section 3.3.2 that the left-tailed t-test is undersized.) Although we do not report the power functions, our computations indicate the Bonferroni tests

[6]The numerical procedure described in Section 3.3.3 for the Bonferroni Q-test is also applied to the Bonferroni t-test. The significance levels $\overline{\alpha}_1$ and $\underline{\alpha}_1$ used in constructing the Dickey-Fuller confidence interval for ρ are chosen to result in a 5% one-sided test for β, uniformly in $c \in [-50, 5]$.

(based on either the t-test or the Q-test) are less undersized than the sup-bound t-test. Hence, the Bonferroni tests have better power, especially when the predictor variable is persistent (i.e. $c = -2$). The two Bonferroni tests have similar power although the t-test version has better power when the degree of persistence is low (i.e. $c = -20$).

As revealed by Figure 3.4, all the feasible tests considered here are biased.[7] That is, the power of the test can be less than the size, for alternatives b sufficiently close to zero. Recently, there has been progress in the development of unbiased tests for predictive regressions (see Jansson and Moreira (2003) and Polk, Thompson, and Vuolteenaho (2003)). Whether unbiasedness is a desirable property in constructing powerful tests of predictability remains to be seen.

We conclude that the Bonferroni Q-test has important power advantages over the other feasible tests. Against right-sided alternatives, it has better power than the Bonferroni t-test, especially when the predictor variable is highly persistent, and it has much better power than the sup-bound Q-test when the predictor variable is less persistent.

3.4 Predictability of Stock Returns

In this section, we implement our test of predictability in US equity data. We then relate our findings to previous empirical results in the literature.

3.4.1 Data

We use four different series of stock returns, dividend-price ratio, and earnings-price ratio. The first is annual S&P 500 index data (1871–2002) from Global Financial Data

[7]See Lehmann (1986, Chapter 4) for a textbook treatment of unbiasedness for hypothesis testing.

since 1926 and Shiller (2000) before then. (The latter is available from the author's webpage.) The other three series are annual, quarterly, and monthly NYSE/AMEX value-weighted index data (1926–2002) from the Center for Research in Security Prices (CRSP).

Following Campbell and Shiller (1988b), the dividend-price ratio is computed as dividends over the past year divided by the current price, and the earnings-price ratio is computed as a moving average of earnings over the past ten years divided by the current price. Since earnings data are not available for the CRSP series, we instead use the corresponding earnings-price ratio from S&P 500. Earnings are available at quarterly frequency since 1935, and annual frequency before then. Shiller (2000) constructs monthly earnings by linear extrapolation. We instead assign quarterly earnings to each month of the quarter since 1935 and annual earnings to each month of the year before then.

To compute excess returns of stocks over a riskfree return, we use the 1-month T-bill rate for the monthly series and the 3-month T-bill rate for the quarterly series. For the annual series, the riskfree return is the return from rolling over the 3-month T-bill every quarter. Since 1926, the T-bill rates are from the CRSP Indices database. For our longer S&P 500 series, we augment this with US Commercial Paper Rates (New York City) from Macaulay (1938), available through NBER's webpage.

For the three CRSP series, we consider the sub-sample 1952–2002 in addition to the full sample. This allows us to add two additional predictor variables, the 3-month T-bill rate and the long-short yield spread. Following Fama and French (1989), the long yield used in computing the yield spread is Moody's Seasoned Aaa Corporate Bond Yield. The short rate used is the 1-month T-bill rate. Although data are available before 1952, the nature of the interest rate is very different then due to the Fed's policy of pegging the interest rate. Following the usual convention, excess returns and the predictor variables are all in logs.

3.4.2 Persistence of Predictor Variables

In Table 3.3, we report the 95% confidence interval for the autoregressive root ρ (and the corresponding c) for the log dividend-price ratio $(d - p)$, the log earnings-price ratio $(e - p)$, the 3-month T-bill rate (r_3), and the long-short yield spread $(y - r_1)$. The confidence interval is computed by the method described in Section 3.3.2. The autoregressive lag length $p \in [1, \overline{p}]$ for the predictor variable is estimated by the Bayes Information Criterion (BIC). We set the maximum lag length $\overline{p}$ to 4 for annual, 6 for quarterly, and 8 for monthly data. The estimated lag lengths are reported in the fourth column of Table 3.3.

All of the series are highly persistent, often containing a unit root in the confidence interval. An interesting feature of the confidence intervals for the valuation ratios ($d - p$ and $e - p$) is that they are sensitive to whether the sample period includes data after 1994. The confidence interval for the sample through 1994 (Panel B) is always less than that for the full sample through 2002 (Panel A). The source of this difference can be explained by Figure 3.2, which is a time series plot of the valuation ratios at quarterly frequency. Around 1994, these valuation ratios begin to drift down to historical lows, making the processes look more nonstationary. The least persistent series is the yield spread, whose confidence interval never contains a unit root.

The high persistence of these predictor variables suggests that first-order asymptotics, which implies that the t-statistic is approximately normal in large samples, may be misleading. As discussed in Section 3.3.2, whether conventional inference based on the t-test is reliable also depends on the correlation δ between the innovations to excess returns and the predictor variable. Hence, we report point estimates

Table 3.3: Estimates of the Model Parameters

This table reports estimates of the parameters for the predictive regression model. Returns are for annual S&P 500 and annual, quarterly, and monthly CRSP value-weighted index. The predictor variables are the log dividend-price ratio $(d-p)$, the log earnings-price ratio $(e-p)$, the 3-month T-bill rate (r_3), and the long-short yield spread $(y-r_1)$. p is the estimated autoregressive lag length for the predictor variable, and δ is the estimated correlation between the innovations to returns and the predictor variable. The last two columns are the 95% confidence intervals for the largest autoregressive root (ρ) and the corresponding local-to-unity parameter (c) for each of the predictor variables, computed using the DF-GLS statistic.

Series	Sample (Obs)	Variable	p	δ	DF-GLS	95% CI: ρ	95% CI: c
				A. Full Sample			
S&P 500	1880–2002	$d-p$	3	-0.846	-0.855	[0.949,1.033]	[-6.107,4.020]
	(123)	$e-p$	1	-0.962	-2.888	[0.768,0.965]	[-28.262,-4.232]
Annual	1926–2002	$d-p$	1	-0.721	-1.033	[0.903,1.050]	[-7.343,3.781]
	(77)	$e-p$	1	-0.957	-2.229	[0.748,1.000]	[-19.132,-0.027]
Quarterly	1926–2002	$d-p$	1	-0.942	-1.696	[0.957,1.007]	[-13.081,2.218]
	(305)	$e-p$	1	-0.986	-2.191	[0.939,1.000]	[-18.670,0.145]
Monthly	1926–2002	$d-p$	2	-0.950	-1.657	[0.986,1.003]	[-12.683,2.377]
	(913)	$e-p$	1	-0.987	-1.859	[0.984,1.002]	[-14.797,1.711]

(continued on the next page)

Series	Sample (Obs)	Variable	p	δ	DF-GLS	95% CI: ρ	95% CI: c
				B. Sample through 1994			
S&P 500	1880–1994	$d-p$	3	-0.836	-2.002	[0.854,1.010]	[-16.391,1.079]
	(115)	$e-p$	1	-0.958	-3.519	[0.663,0.914]	[-38.471,-9.789]
Annual	1926–1994	$d-p$	1	-0.693	-2.081	[0.745,1.010]	[-17.341,0.690]
	(69)	$e-p$	1	-0.959	-2.859	[0.591,0.940]	[-27.808,-4.074]
Quarterly	1926–1994	$d-p$	1	-0.941	-2.635	[0.910,0.991]	[-24.579,-2.470]
	(273)	$e-p$	1	-0.988	-2.827	[0.900,0.986]	[-27.322,-3.844]
Monthly	1926–1994	$d-p$	2	-0.948	-2.551	[0.971,0.998]	[-23.419,-1.914]
	(817)	$e-p$	2	-0.983	-2.600	[0.970,0.997]	[-24.105,-2.240]
				C. Sample from 1952			
Annual	1952–2002	$d-p$	1	-0.749	-0.462	[0.917,1.087]	[-4.131,4.339]
	(51)	$e-p$	1	-0.955	-1.522	[0.773,1.056]	[-11.354,2.811]
		r_3	1	0.006	-1.762	[0.725,1.040]	[-13.756,1.984]
		$y-r_1$	1	-0.243	-3.121	[0.363,0.878]	[-31.870,-6.100]
Quarterly	1952–2002	$d-p$	1	-0.977	-0.392	[0.981,1.022]	[-3.844,4.381]
	(204)	$e-p$	1	-0.980	-1.195	[0.958,1.017]	[-8.478,3.539]
		r_3	4	-0.095	-1.572	[0.941,1.013]	[-11.825,2.669]
		$y-r_1$	2	-0.100	-2.765	[0.869,0.983]	[-26.375,-3.347]
Monthly	1952–2002	$d-p$	1	-0.967	-0.275	[0.994,1.007]	[-3.365,4.451]
	(612)	$e-p$	1	-0.982	-0.978	[0.989,1.006]	[-6.950,3.857]
		r_3	2	-0.071	-1.569	[0.981,1.004]	[-11.801,2.676]
		$y-r_1$	1	-0.066	-4.368	[0.911,0.968]	[-54.471,-19.335]

of δ in the fifth column of Table 3.3. As expected, the correlations for the valuation ratios are negative and large. This is because movements in stock returns and these valuation ratios mostly come from movements in the stock price. The large magnitude of δ suggests that inference based on the conventional t-test leads to large size distortions. More formally, we fail to reject the null hypothesis that the size distortion is greater than 2.5% using the pretest described in Section 3.3.2. For the interest rate variables (r_3 and $y - r_1$), δ is much smaller. For these predictor variables, the pretest rejects the null hypothesis, which suggests that the conventional t-test leads to approximately correct inference.

3.4.3 Testing the Predictability of Returns

In this section, we construct valid confidence intervals for β to test the predictability of returns. Based on the power analysis in Section 3.3.4, our preferred test is the Bonferroni Q-test.

Summary of the Methodology

Our methodology and empirical findings can most easily be explained by the following graphical method, which can be implemented as a series of OLS regressions.

1. Run the regression (3.12) to obtain the standard error for $\widehat{\beta}$, denoted as $\mathrm{SE}(\widehat{\beta})$. Run the regression (3.17) to obtain the coefficients $\widehat{\psi}_i$ ($i = 1, \ldots, p-1$). Using the OLS residuals $\widehat{u}_t$ and $\widehat{e}_t$, compute $\widehat{\sigma}_u^2 = (T-2)^{-1} \sum_{t=1}^{T} \widehat{u}_t^2$, $\widehat{\sigma}_e^2 = (T-2)^{-1} \sum_{t=1}^{T} \widehat{e}_t^2$, $\widehat{\sigma}_{ue} = (T-2)^{-1} \sum_{t=1}^{T} \widehat{u}_t \widehat{e}_t$, $\widehat{\delta} = \widehat{\sigma}_{ue} / (\widehat{\sigma}_u \widehat{\sigma}_e)$, and $\widehat{\omega}^2 = \widehat{\sigma}_e^2 / (1 - \sum_{i=1}^{p-1} \widehat{\psi}_i)^2$.

2. Run the regression (3.15) to obtain the standard error for $\widehat{\rho}$, denoted as $\mathrm{SE}(\widehat{\rho})$. Using the OLS residuals $\widehat{v}_t$, compute $\widehat{\sigma}_v^2 = (T-2)^{-1} \sum_{t=1}^{T} \widehat{v}_t^2$.

3. Compute the DF-GLS statistic as follows. Regress $(x_0, x_1 - \rho_{GLS}x_0, \ldots, x_T - \rho_{GLS}x_{T-1})'$ onto $(1, 1-\rho_{GLS}, \ldots, 1-\rho_{GLS})'$, where $\rho_{GLS} = 1 - 7/T$, to obtain the coefficient μ_{GLS}. Let $\overline{x}_t = x_t - \mu_{GLS}$. Run the regression (3.17) using $\overline{x}_t$ *without the intercept*. The t-statistic for θ is the DF-GLS statistic.

4. Given the value of the DF-GLS statistic and $\widehat{\delta}$, use the lookup tables in Campbell (2003) to find the appropriate confidence interval $[\underline{c}, \overline{c}]$ for c. The confidence interval for ρ is $[\underline{\rho}, \overline{\rho}] = [1 + \underline{c}/T, 1 + \overline{c}/T]$.

5. For each $\rho \in [\underline{\rho}, \overline{\rho}]$, compute an equal-tailed 90% confidence interval for β given ρ as follows. Run the regression (3.12) with $r_t - (\widehat{\sigma}_{ue}/(\widehat{\sigma}_e\widehat{\omega}))(x_t - \rho x_{t-1})$ instead of r_t. Let $\widehat{\beta}(\rho)$ denote the coefficient on x_{t-1}. The confidence interval for β given ρ is $[\underline{\beta}(\rho), \overline{\beta}(\rho)]$, where

$$\begin{aligned}\underline{\beta}(\rho) &= \widehat{\beta}(\rho) + \frac{T-2}{2}\frac{\widehat{\sigma}_{ue}}{\widehat{\sigma}_e\widehat{\omega}}\left(\frac{\widehat{\omega}^2}{\widehat{\sigma}_v^2} - 1\right)\mathrm{SE}(\widehat{\rho})^2 - z_{0.05}(1-\widehat{\delta}^2)^{1/2}\mathrm{SE}(\widehat{\beta}), \\ \overline{\beta}(\rho) &= \widehat{\beta}(\rho) + \frac{T-2}{2}\frac{\widehat{\sigma}_{ue}}{\widehat{\sigma}_e\widehat{\omega}}\left(\frac{\widehat{\omega}^2}{\widehat{\sigma}_v^2} - 1\right)\mathrm{SE}(\widehat{\rho})^2 + z_{0.05}(1-\widehat{\delta}^2)^{1/2}\mathrm{SE}(\widehat{\beta}).\end{aligned}$$

6. Plot $[\underline{\beta}(\rho), \overline{\beta}(\rho)]$ against ρ for all $\rho \in [\underline{\rho}, \overline{\rho}]$.

In step 1, the autoregressive lag length p can be estimated consistently by BIC. When the predictor variable is an AR(1) (i.e. $p = 1$), regressions (3.15) and (3.17) are equivalent, so step 2 can be eliminated. In addition, the formulas in step 5 simplify since $\widehat{\omega}^2 = \widehat{\sigma}_v^2 = \widehat{\sigma}_e^2$ in that case. In practice, we only need to compute the confidence interval $[\underline{\beta}(\rho), \overline{\beta}(\rho)]$ at the end points of $[\underline{\rho}, \overline{\rho}]$ since $\underline{\beta}(\rho)$ and $\overline{\beta}(\rho)$ are linear in ρ. The 90% Bonferroni confidence interval $[\underline{\beta}(\overline{\rho}), \overline{\beta}(\underline{\rho})]$ corresponds to a 10% two-sided test or a 5% one-sided test for predictability.

In reporting our confidence interval for β, we will scale it by $\widehat{\sigma}_e/\widehat{\sigma}_u$. In other words, we report the confidence interval for $\widetilde{\beta} = (\sigma_e/\sigma_u)\beta$ instead of β. Although this normalization does not affect inference, it is a more natural way to report the empirical results for two reasons. First, $\widetilde{\beta}$ has a natural interpretation as the coefficient in (3.12) when the innovations in (3.12) and (3.17) are normalized to have unit variance. This is in the spirit of the simple regression model (3.1)–(3.2), which assumed unit variance in innovations. Second, by the equality

$$\widetilde{\beta} = \frac{\sigma(\mathbf{E}_{t-1}r_t - \mathbf{E}_{t-2}r_t)}{\sigma(r_t - \mathbf{E}_{t-1}r_t)},$$

$\widetilde{\beta}$ can be interpreted as the standard deviation of the change in expected returns relative to the standard deviation of the innovation to returns. To simplify notation, we will use β to denote $\widetilde{\beta}$ throughout the rest of the essay.

Empirical Findings

In Figure 3.5, we plot the Bonferroni confidence interval, using the annual and quarterly CRSP series (1926–2002), when the predictor variable is the dividend-price ratio or the earnings-price ratio. The thick lines represent the confidence interval based on the Bonferroni Q-test, and the thin lines represent the confidence interval based on the Bonferroni t-test. Because of the asymmetry in the null distribution of the t-statistic, the confidence interval for ρ used for the right-tailed Bonferroni t-test differs from that used for the left-tailed test. This explains why the length of the lower bound of the interval, corresponding to the right-tailed test, can differ from the upper bound, corresponding to the left-tailed test. The application of the Bonferroni Q-test is new, but the Bonferroni t-test has been applied previously by Torous, Valkanov, and Yan (2001). We report the latter for the purpose of comparison.

For the annual dividend-price ratio in Panel A, the Bonferroni confidence interval for β based on the Q-test lies strictly above zero. Hence, we can reject the null $\beta = 0$

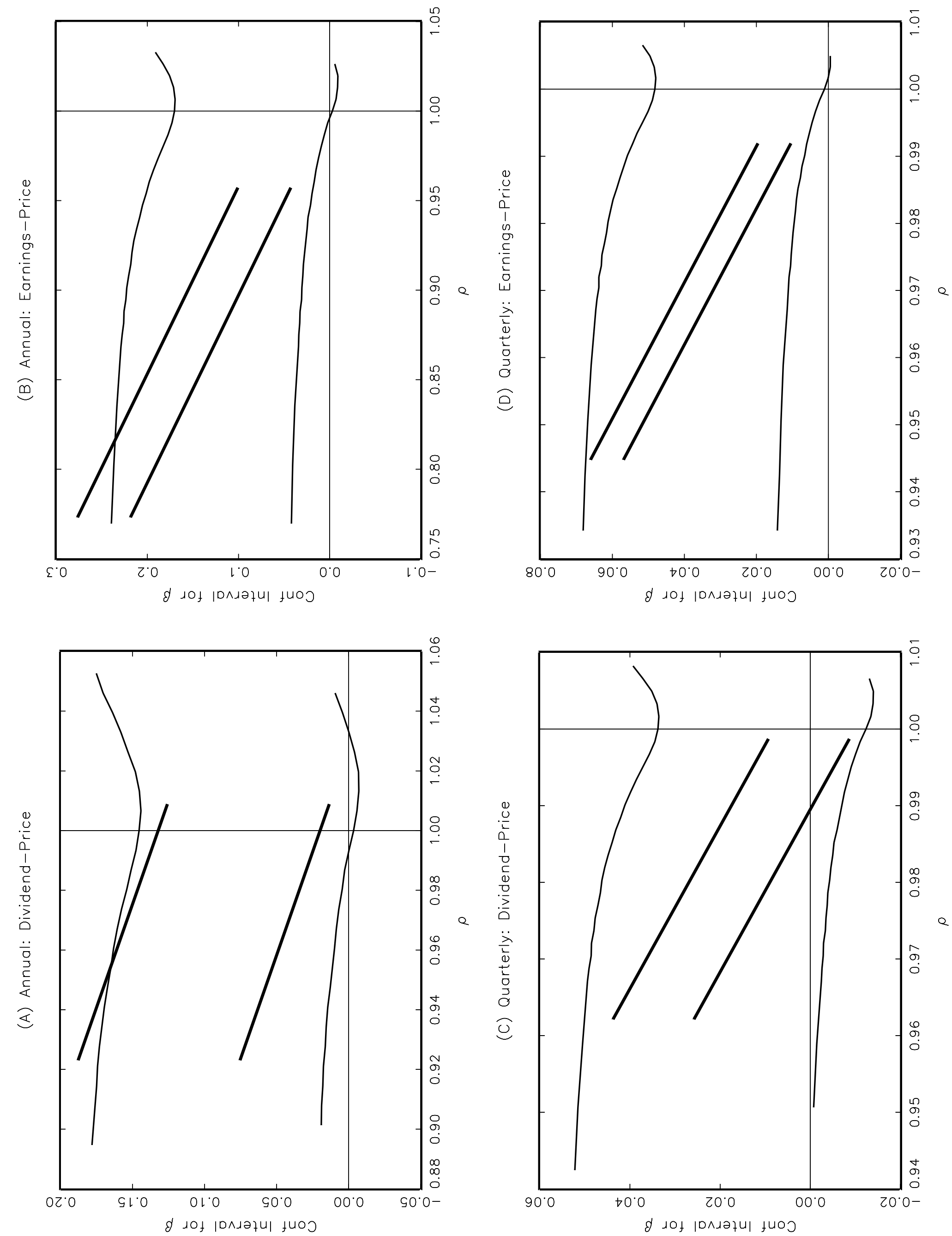

Figure 3.5: Bonferroni Confidence Interval for the Valuation Ratios

This figure plots the 90% confidence interval for β over the confidence interval for ρ. The significance level for ρ is chosen to result in a 90% Bonferroni confidence interval for β. The thick (thin) line is the confidence interval for β computed by inverting the Q-test (t-test). Returns are for annual and quarterly CRSP value-weighted index (1926–2002). The predictor variables are the log dividend-price ratio and the log earnings-price ratio.

against the alternative $\beta > 0$ at the 5% level. The Bonferroni confidence interval based on the t-test, however, includes $\beta = 0$. Hence, we cannot reject the null of no predictability using the Bonferroni t-test. This can be interpreted in light of the power comparisons in Figure 3.4. From Table 3.3, $\widehat{\delta} = -0.721$ and the confidence interval for c is $[-7.343, 3.781]$. In this region of the parameter space, the Bonferroni Q-test is more powerful than the Bonferroni t-test against right-sided alternatives, resulting in a tighter confidence interval.

For the quarterly dividend-price ratio in Panel C, the evidence for predictability is weaker. In the relevant range of the confidence interval for ρ, the confidence interval for β contains zero for both the Bonferroni Q-test and t-test, although the confidence interval is again tighter for the Q-test. Using the Bonferroni Q-test, the confidence interval for β lies above zero when $\rho \leq 0.988$. This means that if the true ρ is less than 0.988, we can reject the null hypothesis $\beta = 0$ against the alternative $\beta > 0$ at the 5% level. On the other hand, if $\rho > 0.988$, the confidence interval includes $\beta = 0$, so we cannot reject the null. Since there is uncertainty over the true value of ρ, we cannot reject the null of no predictability.

In Panel B, we test for predictability in annual data using the earnings-price ratio as the predictor variable. We find that stock returns are predictable with the Bonferroni Q-test, but not with the Bonferroni t-test. In Panel D, we obtain the same results at the quarterly frequency. Again, the Bonferroni Q-test gives tighter confidence intervals due to better power, which is empirically relevant for detecting predictability.

In Figure 3.6, we repeat the same exercise as Figure 3.5, using the quarterly CRSP series in the sub-sample 1952–2002. We report the plots for all four of our predictor variables: A) the dividend-price ratio, B) the earnings-price ratio, C) the T-bill rate, and D) the yield spread.

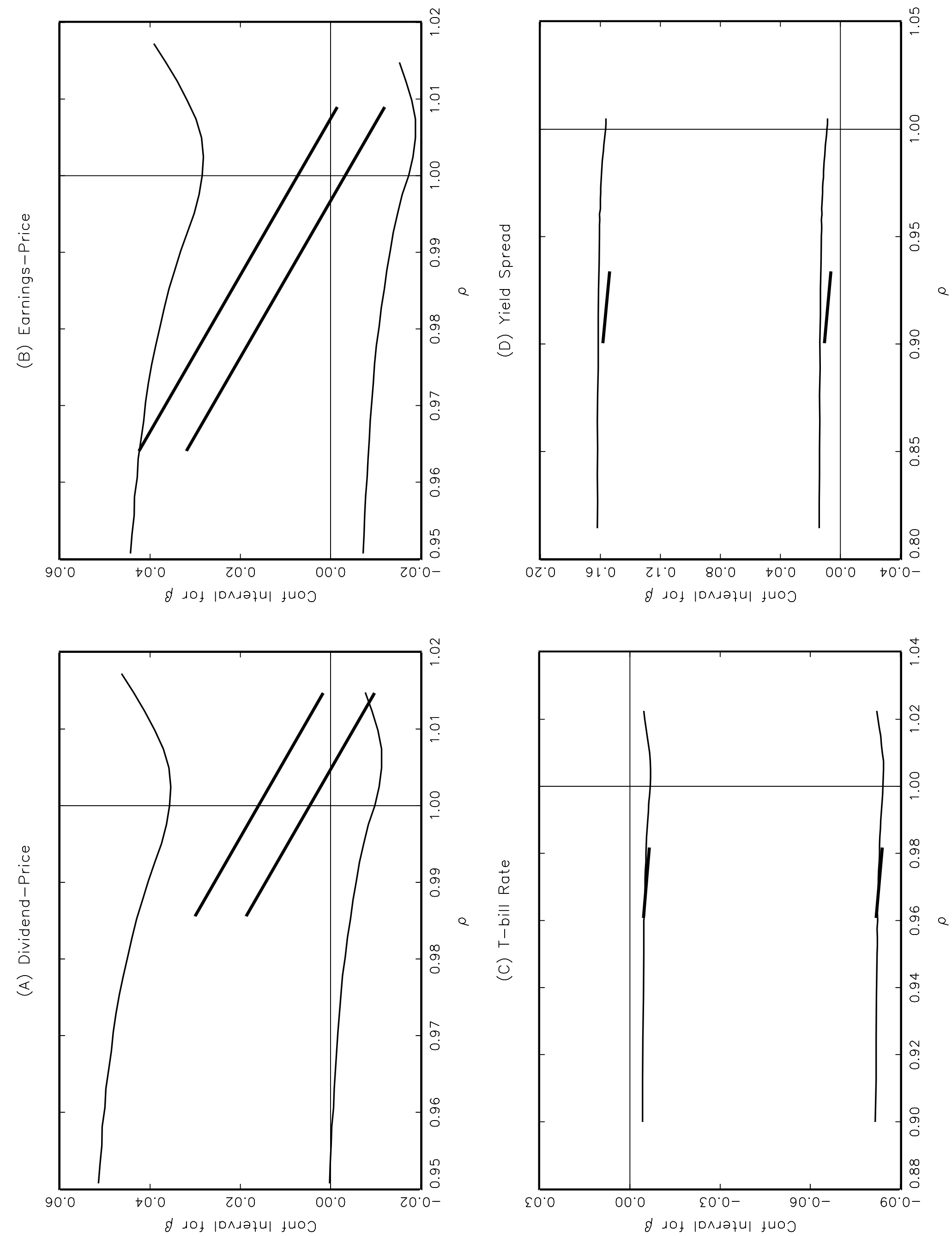

Figure 3.6: Bonferroni Confidence Interval for the Sample from 1952

This figure plots the 90% confidence interval for β over the confidence interval for ρ. The significance level for ρ is chosen to result in a 90% Bonferroni confidence interval for β. The thick (thin) line is the confidence interval for β computed by inverting the Q-test (t-test). Returns are for quarterly CRSP value-weighted index (1952–2002). The predictor variables are the log dividend-price ratio, the log earnings-price ratio, the 3-month T-bill rate, and the long-short yield spread.

For the dividend-price ratio, we find evidence for predictability if $\rho \leq 1.004$. This means that if we are willing to rule out explosive roots, confining attention to the area of the figure to the left of the vertical line at $\rho = 1$, we can conclude that returns are predictable with the dividend-price ratio. The confidence interval for ρ, however, includes explosive roots, so we cannot impose $\rho \leq 1$ without using prior information about the behavior of the dividend-price ratio.

The earnings-price ratio is a less successful predictor variable in this sub-sample. We find that ρ must be less than 0.997 before we can conclude that the earnings-price ratio predicts returns. Taking account of the uncertainty in the true value of ρ, we cannot reject the null hypothesis $\beta = 0$. The weaker evidence for predictability in the period since 1952 is partly due to the fact that the valuation ratios appear more persistent when restricted to this sub-sample. The confidence intervals therefore contain rather large values of ρ that were excluded in Figure 3.5.

For the T-bill rate, the Bonferroni confidence interval for β lies strictly below zero for both the Q-test and the t-test over the entire confidence interval for ρ. For the yield spread, the evidence for predictability is similarly strong, with the confidence interval strictly above zero over the entire range of ρ. The power advantage of the Bonferroni Q-test over the Bonferroni t-test is small when δ is small in absolute value, so these tests result in very similar confidence intervals.

In Table 3.4, we report the complete set of results in tabular form. In the fifth column of the table, we report the 90% Bonferroni confidence intervals for β using the t-test. In the sixth column, we report the 90% Bonferroni confidence interval using the Q-test. In terms of Figures 3.5–3.6, we simply report the minimum and maximum values of β for each of the tests.

Focusing first on the full-sample results in Panel A, the Bonferroni Q-test rejects the null of no predictability for the earnings-price ratio $(e - p)$ at all frequencies. For

Table 3.4: Tests of Predictability

This table reports statistics used to infer the predictability of returns. Returns are for annual S&P 500 and annual, quarterly, and monthly CRSP value-weighted index. (See Table 3.3 for the sample periods and the number of observations.) The predictor variables are the log dividend-price ratio ($d-p$), the log earnings-price ratio ($e-p$), the 3-month T-bill rate (r_3), and the long-short yield spread ($y-r_1$). The third and fourth columns report the t-statistic and the point estimate $\widehat{\beta}$ from an OLS regression of returns onto the predictor variable. The next two columns report the 90% Bonferroni confidence intervals for β using the t-test and Q-test, respectively. The final column reports the lower bound of the confidence interval for β based on the Q-test at $\rho = 1$.

Series	Variable	t-stat	$\widehat{\beta}$	90% CI: β		Low CI β
				t-test	Q-test	($\rho = 1$)
			A. Full Sample			
S&P 500	$d-p$	1.977	0.093	[-0.040,0.136]	[-0.032,0.115]	-0.017
	$e-p$	2.772	0.131	[-0.002,0.190]	[0.043,0.224]	-0.023
Annual	$d-p$	2.534	0.125	[-0.007,0.178]	[0.014,0.188]	0.020
	$e-p$	2.770	0.169	[-0.009,0.240]	[0.042,0.277]	0.002
Quarterly	$d-p$	2.060	0.034	[-0.014,0.052]	[-0.009,0.044]	-0.010
	$e-p$	2.908	0.049	[-0.001,0.068]	[0.010,0.066]	0.002
Monthly	$d-p$	1.706	0.009	[-0.006,0.014]	[-0.005,0.010]	-0.005
	$e-p$	2.662	0.014	[-0.001,0.019]	[0.002,0.018]	0.001
			B. Sample through 1994			
S&P 500	$d-p$	2.243	0.141	[-0.035,0.218]	[-0.048,0.183]	-0.080
	$e-p$	3.331	0.196	[0.063,0.273]	[0.094,0.326]	-0.029
Annual	$d-p$	2.993	0.212	[0.025,0.304]	[0.056,0.332]	0.011
	$e-p$	3.409	0.279	[0.048,0.380]	[0.126,0.448]	0.012
Quarterly	$d-p$	2.304	0.053	[-0.004,0.083]	[-0.006,0.076]	-0.027
	$e-p$	3.506	0.079	[0.018,0.107]	[0.027,0.109]	0.005
Monthly	$d-p$	1.790	0.013	[-0.004,0.022]	[-0.007,0.017]	-0.013
	$e-p$	3.185	0.022	[0.002,0.030]	[0.005,0.028]	0.000

(continued on the next page)

Series	Variable	t-stat	$\widehat{\beta}$	90% CI: β		Low CI β
				t-test	Q-test	($\rho = 1$)
			C. Sample from 1952			
Annual	$d-p$	2.289	0.124	[-0.023,0.178]	[-0.007,0.183]	0.020
	$e-p$	1.733	0.114	[-0.078,0.178]	[-0.031,0.229]	-0.025
	r_3	-1.143	-0.095	[-0.229,0.045]	[-0.231,0.042]	—
	$y-r_1$	1.124	0.136	[-0.087,0.324]	[-0.075,0.359]	-0.156
Quarterly	$d-p$	2.236	0.036	[-0.011,0.051]	[-0.010,0.030]	0.005
	$e-p$	1.777	0.029	[-0.019,0.044]	[-0.012,0.042]	-0.003
	r_3	-1.766	-0.042	[-0.084,-0.004]	[-0.084,-0.004]	-0.086
	$y-r_1$	1.991	0.090	[0.009,0.162]	[0.006,0.158]	-0.002
Monthly	$d-p$	2.259	0.012	[-0.004,0.017]	[-0.004,0.010]	0.001
	$e-p$	1.754	0.009	[-0.006,0.014]	[-0.004,0.012]	-0.001
	r_3	-2.431	-0.017	[-0.030,-0.006]	[-0.030,-0.006]	-0.030
	$y-r_1$	2.963	0.047	[0.020,0.072]	[0.020,0.072]	0.016

the dividend-price ratio ($d-p$), we fail to reject the null except for the annual CRSP series. Using the Bonferroni t-test, we always fail to reject the null due to its poor power relative to the Bonferroni Q-test.

In the sub-sample through 1994, reported in Panel B, the results are qualitatively similar. In particular, the Bonferroni Q-test finds predictability with the earnings-price ratio at all frequencies. Interestingly, the Bonferroni t-test also finds predictability in this sub-sample, although the confidence intervals are still wider than those of the Bonferroni Q-test. In this sub-sample, the evidence for predictability is sufficiently strong that a relatively inefficient test can also find predictability.

In Panel C, we report the results for the sub-sample since 1952. In this sub-sample, we cannot reject the null hypothesis for the valuation ratios ($d-p$ and $e-p$). For the T-bill rate and the yield spread (r_3 and $y-r_1$), however, we reject the null hypothesis except at annual frequency.

As we have seen in Figure 3.6, the weak evidence for predictability in this sub-sample arises from the fact that the confidence intervals for ρ contain explosive roots.

If we could obtain tighter confidence intervals for ρ that exclude these values, the lower bound of the confidence intervals for β would rise, strengthening the evidence for predictability. In the last column of Table 3.4, we report the lower bound of the confidence interval for β at $\rho = 1$. This corresponds to Lewellen's (2003) sup-bound Q-test, which restricts the parameter space to $\rho \leq 1$. In terms of Figures 3.5–3.6, this is equivalent to discarding the region of the plots where $\rho > 1$. Under this restriction, the lower bound of the confidence interval for the dividend-price ratio lies above zero at all frequencies. The dividend-price ratio therefore predicts returns in the subsample since 1952 provided that its autoregressive root is not explosive, consistent with Lewellen's findings.

To summarize the empirical results, we find reliable evidence for predictability with the earnings-price ratio, the T-bill rate, and the yield spread. The evidence for predictability with the dividend-price ratio is weaker, and we do not find unambiguous evidence for predictability using our Bonferroni Q-test. The Bonferroni Q-test gives tighter confidence intervals than the Bonferroni t-test due to better power. The power gain is empirically important in the full sample through 2002.

3.4.4 Connection to Previous Empirical Findings

The empirical literature on the predictability of returns is rather large, and in this section, we attempt to interpret the main findings in light of our analysis in the last section.

t-test

The earliest and the most intuitive approach to testing predictability is to run the predictive regression and to compute the t-statistic. One would then reject the null hypothesis $\beta = 0$ against the alternative $\beta > 0$ at the 5% level if the t-statistic is greater than 1.645. In the third column of Table 3.4, we report the t-statistics from the predictive regressions. Using the conventional critical value, the t-statistics are mostly "significant", often greater than two and sometimes greater than three. Comparing the full sample through 2002 (Panel A) and the sub-sample through 1994 (Panel B), the evidence for predictability appears to have weakened in the last eight years. In the late 1990's, stock returns were high when the valuation ratios were at historical lows. Hence, the evidence for predictability "went in the wrong direction".

However, one may worry about statistical inference that is so sensitive to an addition of 8 observations to a sample of 115 (for S&P 500) or an addition of 32 data points to a sample of 273 (for quarterly CRSP). In fact, this sensitivity is evidence for the failure of first-order asymptotics. Intuitively, when a predictor variable is persistent, its sample moments can change dramatically with an addition of a few data points. Since the t-statistic measures the covariance of excess returns with the lagged predictor variable, its value is sensitive to persistent deviations in the predictor variable from the mean. This is what happened in the late 1990's when valuation ratios reached historical lows. Tests that are derived from local-to-unity asymptotics take this persistence into account and hence lead to correct inference.

Using the Bonferroni Q-test, which is robust to the persistence problem, we find that the earnings-price ratio predicts returns in both the full sample and the sub-sample through 1994. There appears to be some empirical content in the claim that the evidence for predictability has weakened, with the Bonferroni confidence interval based on the Q-test shifting toward zero. Using the Bonferroni confidence interval based on the t-test, we reject the null of no predictability in the sub-sample through 1994, but not in the full sample. The "weakened" evidence for predictability in the recent years puts a premium on the efficiency of test procedures.

As additional evidence for the failure of first-order asymptotics, we report the OLS point estimates of β in the fourth column of Table 3.4. As equations (3.22)–(3.23) show, the point estimate $\widehat{\beta}$ does not necessarily lie in the center of the robust confidence interval for β. Indeed, $\widehat{\beta}$ for the valuation ratios are usually closer to the upper bound of the Bonferroni confidence interval based on the Q-test, and in a few cases (dividend-price ratio in Panel C), $\widehat{\beta}$ falls strictly above the confidence interval. This is a consequence of the upward finite-sample bias of the OLS estimator arising from the persistence of these predictor variables (see Stambaugh (1999) and Lewellen (2003)).

One way to interpret the t-test based on the conventional critical value (1.645 for a 5% one-sided test) is the Bayesian interpretation. Suppose $\delta = -0.9$, which is a reasonable value for the valuation ratios. As reported in Table 3.1, the unknown persistence parameter c must be less than -70 for the size distortion of the t-test to be less than 2.5%. Hence, if a researcher has prior information that $c < -70$, he can proceed with the t-test using the critical value 1.645. Our empirical findings in Figures 3.5–3.6 confirm that there is strong evidence for predictability with the valuation ratios when $\rho \ll 1$. The problem with such inference is that the lower bound of the confidence interval for c is much greater than -70, so it is hard to reconcile the prior belief in a low c with the observed persistence of the valuation ratios.

For the interest rate variables, the correlation δ is sufficiently small that conventional inference based on the t-test leads to approximately correct inference. This is confirmed in Panel C of Table 3.4, where inference based on the conventional t-test agrees with that based on the Bonferroni Q-test.

Long-Horizon Tests

Some authors, notably Fama and French (1988) and Campbell and Shiller (1988b), have explored the behavior of stock returns at lower frequencies by regressing long-horizon returns onto financial variables. In annual data, the dividend-price ratio has a smaller autoregressive root than it does in monthly data and is less persistent in that sense. Over several years, the ratio has an even smaller autoregressive root. Unfortunately, this does not eliminate the statistical problem caused by persistence because the effective sample size shrinks as one increases the horizon of the regression.

Recently, a number of authors have pointed out that the finite-sample distribution of the long-horizon regression coefficient and its associated t-statistic can be quite different from the asymptotic distribution due to persistence in the regressor and overlap in the returns data. (See Hodrick (1992), Nelson and Kim (1993), Ang and Bekaert (2001) for computational results and Valkanov (2003) and Torous, Valkanov, and Yan (2001) for analytical results.) Accounting for these problems, Torous, Valkanov, and Yan (2001) find no evidence for predictability at long horizons using many of the popular predictor variables. In fact, they find no evidence for predictability at any horizon or time period, except at quarterly and annual frequencies in 1952–1994.

Long-horizon regressions can also be understood as a way to reduce the noise in stock returns, because under the alternative hypothesis that returns are predictable, the variance of the return increases less than proportionally with the investment horizon (see Campbell, Lo, and MacKinlay (1997, Chapter 7) and Campbell (2001)). The procedures developed in this essay and in Lewellen (2003) have the important advantage that they reduce noise not only under the alternative, but also under the null. Thus they increase power against local alternatives, while long-horizon regression tests do not.

Other Tests

In this section, we discuss three recent papers that have taken the issue of persistence seriously to develop tests that have the correct size even if the predictor variable is highly persistent or I(1).

Lewellen (2003) proposes to test the predictability of returns by computing the Q-statistic evaluated at $\beta_0 = 0$ and $\rho = 1$ (i.e. $Q(0, 1)$). His test procedure rejects $\beta = 0$ against the one-sided alternative $\beta > 0$ at the α-level if $Q(0, 1) > z_\alpha$. Since the null distribution of $Q(0, 1)$ is standard normal under local-to-unity asymptotics, Lewellen's test procedure has correct size as long as $\rho = 1$. If $\rho \neq 1$, this procedure does not in general have the correct size. However, Lewellen's procedure is a valid (although conservative) one-sided test as long as $\delta \leq 0$ and $\rho \leq 1$. As we have shown in Panel C of Table 3.4, the 5% one-sided test using the monthly dividend-price ratio rejects when $\rho = 1$, confirming Lewellen's empirical findings.

Based on finance theory, it is reasonable to assume that the dividend-price ratio is mean reverting, at least in the very long run. However, we may not necessarily want to impose Lewellen's parametric assumption that the dividend-price ratio is an AR(1) with $\rho \leq 1$. In the absence of knowledge of the true data-generating process, the purpose of the parametric model (3.15)–(3.16) is to provide a flexible framework to approximate the dynamics of the predictor variable in finite samples. Allowing for the possibility that $\rho > 1$ can be an important part of that flexibility, especially in light of the recent behavior of the dividend-price ratio. In addition, we allow for possible short-run dynamics in the predictor variable by considering an AR(p), which Lewellen rules out by imposing a strict AR(1).

Another issue that arises with Lewellen's test is that of power. As shown in Figure 3.4, the test can have poor power when the predictor variable is stationary (i.e. $\rho < 1$). For instance, the annual earnings-price ratio for the S&P 500 index has 95% confidence interval [0.768,0.965] for ρ. As reported in Panel A of Table 3.4, the lower bound of the confidence interval for β using the Bonferroni Q-test is 0.043, rejecting the null of no predictability. However, the Q-test at $\rho = 1$ results in a lower bound of -0.023, failing to reject the null. Therefore, the poor power of Lewellen's procedure leads to the false conclusion that the earning-price ratio does not predict returns at annual frequency. Similarly, Lewellen's procedure always leads to wider confidence intervals than the Bonferroni Q-test in the sub-sample through 1994, when the valuation ratios are less persistent.

Torous, Valkanov, and Yan (2001) develop a test of predictability that is conceptually similar to ours, constructing Bonferroni confidence intervals for β. One difference from our approach is that they construct the confidence interval for ρ using the augmented Dickey-Fuller test, rather than the more powerful DF-GLS test of Elliott, Rothenberg, and Stock (1996). The second difference is that they use the long-horizon t-test, instead of the more powerful Q-test, for constructing the confidence interval of β given ρ. Their choice of the long-horizon t-test is motivated by their objective of highlighting the pitfalls of long-horizon regressions.

A key insight in Torous, Valkanov, and Yan (2001) is that the evidence for the predictability of returns depends critically on the unknown degree of persistence of the predictor variable. Because we cannot estimate the degree of persistence consistently, the evidence for predictability can be ambiguous. This point is illustrated in Figures 3.5–3.6, where we find that the dividend-price ratio predicts returns if its autoregressive root ρ is sufficiently small. In this essay, we have confirmed their finding that the evidence for predictability by the dividend-price ratio is weak once its persistence has been properly accounted for.

A different approach to dealing with the problem of persistence is to ignore the data on predictor variables and to base inference solely on the returns data. Under the null that returns are not predictable by a persistent predictor variable, returns should behave like a stationary process. Under the alternative of predictability, returns should have a unit or a near-unit root. Using this approach, Lanne (2002) fails to reject the null of no predictability. However, his test is conservative in the sense that it has poor power when the predictor variable is persistent but not close enough to being integrated.[8] Lanne's empirical finding agrees with ours and those of Torous, Valkanov, and Yan (2001). From Figures 3.5–3.6, we see that the valuation ratios predict returns provided that their degree of persistence is sufficiently small. In addition, we find evidence for predictability with the yield spread, which has a relatively small degree of persistence compared to the valuation ratios. Lanne's test would fail to detect predictability by less persistent variables like the yield spread.

3.5 Conclusion

The hypothesis that stock returns are predictable at long horizons has been called a "new fact in finance" (Cochrane 1999). That the predictability of stock returns is now widely accepted by financial economists is remarkable given the long tradition of the "random walk" model of stock prices. In this essay, we have shown that there is indeed evidence for predictability, but it is more challenging to detect than previous studies may have suggested. Most popular and economically sensible candidates for predictor variables (such as the dividend-price ratio, earnings-price ratio, or measures of the interest rate) are highly persistent. When the predictor variable is persistent, the distribution of the t-statistic can be nonstandard, which can lead to over-rejection

[8] In fact, Campbell, Lo, and MacKinlay (1997, Chapter 7) construct an example in which returns are univariate white noise but are predictable using a stationary variable with an arbitrary autoregressive coefficient.

of the null hypothesis using conventional critical values.

In this essay, we have developed a pretest to determine when the conventional t-test leads to misleading inferences. Using the pretest, we find that the t-test leads to correct inference for the short-term interest rate and the long-short yield spread. Persistence is not a problem for these interest rate variables because their innovations have sufficiently low correlation with the innovations to stock returns. Using the t-test with conventional critical values, we find that these interest rate variables predict returns in the post-1952 sample.

For the dividend-price ratio and the smoothed earnings-price ratio, persistence is an issue since their innovations are highly correlated with the innovations to stock returns. Using our pretest, we find that the conventional t-test can lead to misleading inferences for these valuation ratios. In this essay, we have developed an efficient test of predictability that leads to correct inference regardless of the degree of persistence of the predictor variable. Over the full sample, our test reveals that the earnings-price ratio reliably predicts returns at various frequencies (annual to monthly), while the dividend-price ratio predicts returns only at annual frequency. In the post-1952 sample, the evidence for predictability is weaker, but the dividend-price ratio predicts returns if we can rule out explosive autoregressive roots.

Taken together, these results suggest that there is a predictable component in stock returns, but one that is difficult to detect without careful use of efficient statistical tests.

Appendix

A.1 Estimation of Linear Factor Models

A.1.1 Cross-Sectional Regression

An intuitive way to estimate the linear factor model

$$\mathbf{E}[R_{it} - R_{0t}] = b'\Sigma_{fi} \tag{A.1}$$

is by a cross-sectional regression. The idea is to first estimate the covariance between the factors and the return on asset i as

$$\widehat{\mu}_f = \frac{1}{T}\sum_{t=1}^{T} f_t, \tag{A.2}$$

$$\widehat{\Sigma}_{fi} = \frac{1}{T}\sum_{t=1}^{T}(f_t - \widehat{\mu}_f)(R_{it} - R_{0t}). \tag{A.3}$$

Then estimate the factor risk prices b by the cross-sectional regression

$$\frac{1}{T}\sum_{t=1}^{T}[R_{it} - R_{0t}] = b'\widehat{\Sigma}_{fi}. \tag{A.4}$$

To relate this estimator to the GMM estimator based on the moments

$$\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} e_1(z_t,\theta) \\ e_2(z_t,\theta) \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} R_t - R_{0t}\iota - (R_t - R_{0t}\iota)(f_t-\mu_f)'b \\ f_t - \mu_f \end{bmatrix}, \tag{A.5}$$

consider the class of weighting matrices

$$W(\kappa) = \begin{bmatrix} I_N & 0 \\ 0 & \kappa I_F \end{bmatrix}, \tag{A.6}$$

where $\kappa > 0$ is a constant. As $\kappa \to \infty$, the GMM estimator approaches the cross-sectional regression. To see this explicitly, note that $T^{-1}\sum_{t=1}^{T} e_2(z_t,\theta) = 0$ results in the estimator (A.2) for the factor mean. Substituting $\widehat{\mu}_f$ for μ_f in $T^{-1}\sum_{t=1}^{T} e_1(z_t,\theta)$ and minimizing with respect to b is equivalent to running the cross-sectional regression (A.4).

Of course, setting the second moment $T^{-1}\sum_{t=1}^{T} e_2(z_t,\theta) = 0$ requires the weighting matrix $W(\infty)$. However, since this weighting matrix is degenerate, the resulting one-step estimator need not be consistent, and consequently, its finite sample bias can be large. Moreover, the resulting two-step estimator need not be efficient. Intuitively, the cross-sectional regression puts too little weight on valuable information about the factor mean and risk prices contained in the first moment $T^{-1}\sum_{t=1}^{T} e_1(z_t,\theta)$.

A.1.2 First-Stage Estimates of Consumption-Based Models

I now examine how the choice of the weighting matrix affects first-stage estimates of consumption-based linear factor models. Although formal tests of linear factor models are based on efficient second-stage estimates, first-stage measures of fit (e.g. mean absolute pricing error and R^2) have a long tradition in the finance literature.

I examine two cases of the weighting matrix (A.6), $\kappa = N/F$ and $\kappa = 100,000$. In estimating linear factor models, the number of test assets N is typically an order of magnitude greater than the number of factors F. Thus, the identity matrix (i.e. $\kappa = 1$) tends to put too little weight on $T^{-1}\sum_{t=1}^{T} e_2(z_t, \theta)$, allowing for too much "freedom" in the estimated factor mean. The weighting matrix $\kappa = N/F$ has the intuitive appeal of putting equal weight on the N moments used to identify the factor risk prices and the F moments used to identify the factor means. The weighting matrix $\kappa = 100,000$ is deliberately chosen to force GMM to set the factor mean equal to the sample mean (A.2) up to the first three decimal places. This is in the spirit of the cross-sectional regression (A.4), which corresponds to the degenerate weighting matrix $W(\infty)$.

Table A.1 reports first-stage estimates of the CCAPM and the durable consumption model. Comparing the estimates for the CCAPM, both the estimated risk price and the factor mean are smaller when $\kappa = N/F$. For $\kappa = 100,000$, the estimated risk price for nondurable consumption is 159, which appears to be biased upward compared to an efficient estimate of 106 (Table 1.2). The difference in measures of fit for the two weighting matrices is rather dramatic. The R^2 is 38% for $\kappa = N/F$, compared to -18% for $\kappa = 100,000$. Treating the sample mean of nondurable consumption growth as the population mean apparently results in a poor fit of the CCAPM through poor (implied) estimates of nondurable consumption betas.

The results are similar for the durable consumption model. For $\kappa = 100,000$, the estimated risk price for nondurable consumption is 175, which appears to be biased upward compared to an efficient estimate of 122 (Table 1.2). The estimated risk price for durable consumption is 41, which appears to be biased severely downward, compared to an efficient estimate of 197 (Table 1.2). For $\kappa = N/F$, the GMM estimate for the mean of durable consumption growth, 0.63%, is somewhat smaller

Table A.1: First-Stage GMM Estimates

The table reports the estimated factor risk prices and the factor means for the CCAPM and the durable consumption model. The test assets are the 25 Fama-French portfolios sorted by size and book-to-market equity. Estimation is by GMM, using the weighting matrix (A.6) with $\kappa = N/F$ or $\kappa = 100,000$. HAC standard errors in parentheses.

Factor	Parameter	CCAPM		Durable Model	
		$\kappa = N/F$	$\kappa = 100,000$	$\kappa = N/F$	$\kappa = 100,000$
Nondurables	Risk Price	117.316	159.447	112.944	174.513
		(106.520)	(143.823)	(63.735)	(81.402)
	Mean (%)	0.276	0.513	0.343	0.513
		(0.592)	(0.068)	(0.086)	(0.068)
Durables	Risk Price			187.536	41.498
				(118.033)	(126.655)
	Mean (%)			0.632	0.915
				(0.151)	(0.158)
MAE (%)		0.329	0.467	0.198	0.456
R^2		0.382	-0.180	0.770	-0.172

than the sample mean 0.92%. However, the estimate is within two standard errors of the sample mean. Again, the difference in measures of fit for the two weighting matrices is dramatic. The R^2 is 77% for $\kappa = N/F$, but -17% for $\kappa = 100,000$.

This illustrates the pitfalls of treating the sample factor mean as the population mean in estimating consumption-based linear factor models. The cross-sectional regression, which is equivalent to using a degenerate weighting matrix, results in poor estimates and inferences of the model.

A.2 Portfolios Sorted by Consumption Betas

A.2.1 Portfolio Formation

The portfolios are formed using returns on ordinary common equity, traded in NYSE, AMEX, or Nasdaq, in the CRSP Monthly Stock database. In June of each year t, nondurable and durable consumption betas are computed for each stock using quarterly returns from January of $t-5$ through December of $t-1$. Stocks with return data missing in any quarter are dropped from the sample. Then 25 portfolios are formed by independently sorting stocks into quintiles based on the nondurable and durable consumption betas. The value-weighted portfolio returns are then tracked from July of t through June of $t+1$.

A.2.2 Test of Linear Factor Models

Table A.2 reports estimates of the durable consumption model using the portfolios sorted by consumption betas. Without the intratemporal FOC, the point estimate of the risk price for nondurable consumption is 108. The estimate of the risk price for durable consumption is 98, which is significantly different from zero, implying a

Table A.2: Estimation of Linear Factor Models with Portfolios Sorted by Consumption Betas

The test assets are 25 portfolios formed by independently sorting stocks into quintiles based on nondurable and durable consumption betas. See notes to Table 1.2.

Factor Price	CAPM	Fama-French	CCAPM	Durable Model	
				Without FOC	With FOC
Market	2.118	3.760			
	(0.541)	(0.930)			
SMB		-4.257			
		(1.748)			
HML		0.639			
		(1.970)			
Nondurables			98.052	107.784	129.480
			(16.142)	(7.447)	(7.236)
Durables				98.116	119.716
				(13.410)	(17.995)
γ				205.901	249.196
				(16.731)	(24.725)
α (if $\rho = 0$)				0.479	0.482
				(0.035)	(0.026)
MAE (%)	0.336	0.295	0.362	0.247	0.279
R^2	-0.795	-0.299	-0.852	0.136	-0.187
J-test	24.562	16.724	24.143	17.640	32.586
	(0.430)	(0.778)	(0.453)	(0.777)	(0.113)

rejection of the CCAPM. The R^2 is 14%, and the J-test fails to reject the model. The results are similar when the intratemporal FOC is imposed, although the first-stage measures of fit are somewhat worse.

A.3 OCE Preferences

A.3.1 Derivation of the Euler Equation

Let $B_t = (B_{0t}, \dots, B_{Nt})'$. The current-period value function is

$$K_{t-1}(W_{t-1}, D_{t-2}) = \max_{C_{t-1},E_{t-1},B_{t-1}} \frac{v_{t-1}^{1-\sigma}}{1-\sigma} + \beta J_{t-1}(W_t, D_{t-1}), \tag{A.7}$$

where the continuation value function is

$$J_s(W_t, D_{t-1}) = \max_{C_t,E_t,B_t} \frac{\mathbf{E}_s[v_t^{1-\gamma}]^{\frac{1-\sigma}{1-\gamma}}}{1-\sigma} + \beta J_s(W_{t+1}, D_t). \tag{A.8}$$

The FOC with respect to $B_{i,t-1}$ is

$$-v_{t-1}^{-\sigma} v_{C,t-1} + \beta \frac{\partial J_{t-1}(W_t, D_{t-1})}{\partial B_{i,t-1}} = 0. \tag{A.9}$$

Differentiating the continuation value function with respect to $B_{i,t-1}$,

$$\frac{\partial J_{t-1}(W_t, D_{t-1})}{\partial B_{i,t-1}} = \mathbf{E}_{t-1}[v_t^{1-\gamma}]^{\frac{\gamma-\sigma}{1-\gamma}} \mathbf{E}_{t-1}[v_t^{-\gamma} v_{C,t} R_{it}]. \tag{A.10}$$

Combining equations (A.9) and (A.10) results in equation (1.42).

A.3.2 Log-Linear Approximation for the Riskfree Rate

Taking the log of both sides of equation (1.42) for the riskfree asset $i = 0$,

$$\frac{\gamma-\sigma}{1-\sigma} \log \mathbf{E}_{t-1}\left[\left(\frac{v_t}{v_{t-1}}\right)^{1-\gamma}\right] + \log \mathbf{E}_{t-1}[M_t R_{0t}] = 0. \tag{A.11}$$

Combining the second-order log-linear approximation

$$\log \mathbf{E}_{t-1}\left[\left(\frac{v_t}{v_{t-1}}\right)^{1-\gamma}\right] \approx (1-\gamma)\mathbf{E}_{t-1}[(1-\alpha)\Delta c_t + \alpha \Delta d_t] + \frac{(1-\gamma)^2}{2}\mathrm{Var}_{t-1}((1-\alpha)\Delta c_t + \alpha\Delta d_t) \quad \text{(A.12)}$$

with equation (1.31) results in equation (1.43).

A.4 Weak-Instrument Asymptotic Distribution of Similar Tests

This appendix derives the asymptotic distributions of similar tests under weak-instrument asymptotics (Staiger and Stock 1997). The asymptotic representations can then be used to plot the power functions for the similar tests that appear in Stock, Wright, and Yogo (2002, Figures 2–3). Let $\xrightarrow{p}$ denote convergence in probability and $\xrightarrow{d}$ denote convergence in distribution. Following Staiger and Stock, I make the following assumptions.

Assumption 3 (Local-to-Zero). $\Pi = C/\sqrt{T}$, *where* C *is a fixed* $K_2 \times 1$ *vector.*

Assumption 4 (Moment Conditions). *The following limits hold jointly:*

1. $(u'u/T, V'u/T, V'V/T) \xrightarrow{p} (\sigma_{uu}, \Sigma_{Vu}, \Sigma_{VV})$;
2. $\overline{Z}'\overline{Z}/T \xrightarrow{p} Q = \begin{bmatrix} Q_{XX} & Q_{XZ} \\ Q_{ZX} & Q_{ZZ} \end{bmatrix}$;
3. $(X'u/\sqrt{T}, Z'u/\sqrt{T}, X'V/\sqrt{T}, Z'V/\sqrt{T}) \xrightarrow{d} (\Psi_{Xu}, \Psi_{Zu}, \Psi_{XV}, \Psi_{ZV})$, *where* $\Psi = (\Psi'_{Xu}, \Psi'_{Zu}, \Psi'_{XV}, \Psi'_{ZV})' \sim \mathbf{N}(0, \Sigma \otimes Q)$.

As noted by Staiger and Stock, Assumption 4 can be derived from weak primitive assumptions that are reasonable in the present context of estimating the linearized Euler equation. Let $\Upsilon = Q_{ZZ} - Q_{ZX}Q_{XX}^{-1}Q_{XZ}$ and $\lambda = \Upsilon^{1/2}C\Sigma_{VV}^{-1/2}$. Define the vector

$$\begin{bmatrix} z_u \\ z_V \end{bmatrix} = \begin{bmatrix} \Upsilon^{-1/2}(\Psi_{Zu} - Q_{ZX}Q_{XX}^{-1}\Psi_{Xu})\sigma_{uu}^{-1/2} \\ \Upsilon^{-1/2}(\Psi_{ZV} - Q_{ZX}Q_{XX}^{-1}\Psi_{XV})\Sigma_{VV}^{-1/2} \end{bmatrix}$$

$$\sim \mathbf{N}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \otimes I_{K_2}\right). \tag{A.13}$$

Under Assumptions 3 and 4, Staiger and Stock (1997, Theorem 1(e)) show that $F \xrightarrow{d} (\lambda+z_V)'(\lambda+z_V)/K_2$. In other words, the first-stage F-statistic is asymptotically $O_p(1)$ under weak-instrument asymptotics. In contrast, the F-statistic becomes arbitrarily large under the conventional first-order asymptotics with fixed Π.

The following lemma derives the weak instrument asymptotic distributions of the statistics $\mathcal{S}$ and $\mathcal{T}$ (see (2.23) and (2.24)).

Lemma 2. *Suppose that Assumptions 3 and 4 hold. Let* $\overline{\Delta} = \sigma_{uu}^{-1/2}\Sigma_{VV}^{1/2}(\beta_0 - \beta)$ *and* $S_1(s) = 1 - 2\rho s + s^2$ *for any scalar* s. *Then*

$$\mathcal{S} \xrightarrow{d} \mathcal{S}^* = \frac{z_u - (\lambda + z_V)\overline{\Delta}}{S_1(\overline{\Delta})^{1/2}}$$

$$\sim \mathbf{N}\left(\frac{-\lambda\overline{\Delta}}{S_1(\overline{\Delta})^{1/2}}, I_{K_2}\right), \tag{A.14}$$

$$\mathcal{T} \xrightarrow{d} \mathcal{T}^* = \frac{(\lambda + z_V) + z_u\overline{\Delta} - \rho[z_u + (\lambda + z_V)\overline{\Delta})]}{(1-\rho^2)^{1/2}S_1(\overline{\Delta})^{1/2}}$$

$$\sim \mathbf{N}\left(\frac{\lambda(1 - \rho\overline{\Delta})}{(1-\rho^2)^{1/2}S_1(\overline{\Delta})^{1/2}}, I_{K_2}\right), \tag{A.15}$$

where $\mathcal{S}^*$ *and* $\mathcal{T}^*$ *are independent.*

Proof. Applying Staiger and Stock (1997, Lemma A1), $\widehat{\Omega} \xrightarrow{p} \Omega$ and $(Z^{\perp\prime}Z^{\perp})^{-1/2}Z^{\perp\prime}\overline{Y}^{\perp} \xrightarrow{d} w'$, where

$$w = \sigma_{uu}^{1/2} \begin{bmatrix} z_u + \sigma_{uu}^{-1/2}\Sigma_{VV}^{1/2}\beta(\lambda + z_V) \\ \sigma_{uu}^{-1/2}\Sigma_{VV}^{1/2}(\lambda + z_V) \end{bmatrix}.$$

This then implies that

$$\mathcal{S} \xrightarrow{d} \mathcal{S}^* = \frac{w'b_0}{(b_0'\Omega b_0)^{1/2}}, \tag{A.16}$$

$$\mathcal{T} \xrightarrow{d} \mathcal{T}^* = \frac{w'\Omega^{-1}a_0}{(a_0'\Omega^{-1}a_0)^{1/2}}. \tag{A.17}$$

Establishing the equivalence of these expressions to those that appear in the statement of the lemma requires the intermediate steps $b_0'\Omega b_0 = \sigma_{uu}S_1(\overline{\Delta})$, $a_0'\Omega^{-1}a_0 = [(1 - \rho^2)\Sigma_{VV}]^{-1}S_1(\overline{\Delta})$, and

$$\Omega^{-1}a_0 = \frac{1}{(1-\rho^2)\Sigma_{VV}} \begin{bmatrix} \sigma_{uu}^{-1/2}\Sigma_{VV}^{1/2}(\overline{\Delta} - \rho) \\ 1 - \rho\overline{\Delta} - \sigma_{uu}^{-1/2}\Sigma_{VV}^{1/2}\beta(\overline{\Delta} - \rho) \end{bmatrix}.$$

□

Note that $\mathcal{S}$ is asymptotically pivotal and independent of $\mathcal{T}$ under the null hypothesis (i.e. $\overline{\Delta} = 0$). A straightforward application of Lemma 2 to the AR statistic (2.25) results in

$$AR(\beta_0) \xrightarrow{d} \frac{\mathcal{S}^{*\prime}\mathcal{S}^*}{K_2} \sim \frac{\chi^2_{K_2}(\overline{\Delta}\lambda'\lambda\overline{\Delta}/S_1(\overline{\Delta}))}{K_2}, \tag{A.18}$$

which was shown by Staiger and Stock (1997, Theorem 5). The asymptotic distributions of the LM and LR statistics can similarly be obtained by application of Lemma 2 to (2.26) and (2.27), respectively.

Note that the asymptotic distributions of AR, LM, and LR statistics are completely determined by the matrix $[\mathcal{S}^*, \mathcal{T}^*]'[\mathcal{S}^*, \mathcal{T}^*]$, which has a noncentral Wishart distribution $\mathbf{W}_2(K_2, I_2, \Lambda)$ (see Phillips (1983)) with noncentrality matrix

$$\Lambda = \lambda'\lambda \begin{bmatrix} \frac{\overline{\Delta}^2}{S_1(\overline{\Delta})} & \frac{-\overline{\Delta}(1-\rho\overline{\Delta})}{(1-\rho^2)^{1/2}S_1(\overline{\Delta})} \\ \frac{-\overline{\Delta}(1-\rho\overline{\Delta})}{(1-\rho^2)^{1/2}S_1(\overline{\Delta})} & \frac{(1-\rho\overline{\Delta})^2}{(1-\rho^2)S_1(\overline{\Delta})} \end{bmatrix}. \tag{A.19}$$

Hence, the asymptotic distributions only depend on the number of instruments K_2, the concentration parameter $\lambda'\lambda$, the degree of simultaneity ρ, and $\overline{\Delta}$. The parameter $\overline{\Delta}$ has a natural interpretation as the distance between the null hypothesis β_0 and the true value β when the IV regression model, (2.14) and (2.15), is normalized to have unit variance.

A.5 Asymptotic Distribution of Test Statistics under the Local Alternative

In this appendix, we derive the asymptotic distribution of the t-statistic and the Q-statistic under the local alternative $\beta = \beta_0 + b/T$. These asymptotic representations are used to compute the power functions of the various test procedures in Section 3.3.4.

The t-statistic can be written as

$$t(\beta_0) = \frac{b(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}{\sigma_u} + \frac{T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu} u_t}{\sigma_u (T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}.$$

By Lemma 1 (see also Cavanagh, Elliott, and Stock (1995)),

$$t(\beta_0) \Rightarrow \frac{b\omega\kappa_c}{\sigma_u} + \delta\frac{\tau_c}{\kappa_c} + (1-\delta^2)^{1/2} Z, \tag{A.20}$$

where Z is a standard normal random variable independent of $(W_e(s), J_c(s))$.

Let $\widetilde{\rho} = 1 + \widetilde{c}/T$ and define the statistic

$$Q(\beta_0, \widetilde{\rho}) = \frac{\sum_{t=1}^{T} x_{t-1}^{\mu}[r_t - \beta_0 x_{t-1} - \frac{\sigma_{ue}}{\sigma_e \omega}(x_t - \widetilde{\rho} x_{t-1})] + \frac{T}{2}\frac{\sigma_{ue}}{\sigma_e \omega}(\omega^2 - \sigma_v^2)}{\sigma_u(1-\delta^2)^{1/2}(\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}. \tag{A.21}$$

The three types of Q-test considered in Section 3.3.4 correspond to special cases of this test statistic.

1. Infeasible Q-test: $\widetilde{\rho} = \rho = 1 + c/T$.

2. Bonferroni Q-test: $\widetilde{\rho} = \overline{\rho} = 1 + \overline{c}/T$, where $\overline{c}$ depends on the DF-GLS statistic and δ.

3. Sup-bound Q-test: $\widetilde{\rho} = 1$.

The statistic (A.21) can be written as

$$\begin{aligned} Q(\beta_0, \widetilde{\rho}) \quad = \quad & \frac{b(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}{\sigma_u(1-\delta^2)^{1/2}} + \frac{\delta(\widetilde{c} - c)(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}{\omega(1-\delta^2)^{1/2}} \\ & + \frac{T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}(u_t - \frac{\sigma_{ue}}{\sigma_e \omega} v_t) + \frac{1}{2}\frac{\sigma_{ue}}{\sigma_e \omega}(\omega^2 - \sigma_v^2)}{\sigma_u(1-\delta^2)^{1/2}(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}. \end{aligned}$$

By Lemma 1,

$$Q(\beta_0, \rho) \Rightarrow \frac{b\omega\kappa_c}{\sigma_u(1-\delta^2)^{1/2}} + \frac{\delta(\widetilde{c} - c)\kappa_c}{(1-\delta^2)^{1/2}} + Z. \tag{A.22}$$

The power function for the right-tailed test (i.e. $b > 0$) is therefore given by

$$\pi_Q(b) = \mathbf{E}\left[\Phi\left(z_\alpha - \frac{b\omega\kappa_c}{\sigma_u(1-\delta^2)^{1/2}} - \frac{\delta(\widetilde{c} - c)\kappa_c}{(1-\delta^2)^{1/2}}\right)\right], \tag{A.23}$$

where the expectation is taken over the distribution of $(W_e(s), J_c(s))$.

Following Stock (1991, Appendix B), the limiting distributions (A.20) and (A.22) are approximated by Monte Carlo simulation. We generate 20,000 realizations of the

Gaussian AR(1) (i.e. model (3.2)) with $T = 500$, $\rho = 1 + c/T$, and $x_0 = 0$. The distribution of κ_c is approximated by $(T^{-2} \sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}$, and τ_c is approximated by $T^{-1} \sum_{t=1}^{T} x_{t-1}^{\mu} e_t$.

A.6 Tables for Implementing the Test for Predictability

This appendix contains tables necessary for implementing the econometric methods in Chapter 3.

Table A.3 reports equal-tailed 95%, 90%, and 80% confidence intervals for c, based on the DF-GLS statistic. The corresponding confidence interval for ρ (the largest autoregressive root of the predictor variable) can be computed as $[\underline{\rho}, \overline{\rho}] = [1 + \underline{c}/T, 1 + \overline{c}/T]$.

Tables A.4–A.13 are lookup tables necessary for implementing the Bonferroni Q-test. Given the value of the DF-GLS statistic and the estimate of δ (correlation between the innovations to returns and the predictor variable), the tables report the appropriate confidence interval for c. The confidence interval $[\underline{\rho}, \overline{\rho}] = [1+\underline{c}/T, 1+\overline{c}/T]$ for ρ results in a 10% Bonferroni confidence interval for β (i.e. 5% one-sided test for predictability). See Section 3.4.3 for details on the implementation of the method.

Table A.3: Confidence Interval for c Based on the DF-GLS Statistic

DF-GLS	95%		90%		80%	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	-0.627	4.992	-0.282	4.260	0.106	3.462
0.9	-0.753	4.954	-0.395	4.222	0.021	3.418
0.8	-0.892	4.916	-0.509	4.177	-0.084	3.373
0.7	-1.036	4.879	-0.639	4.132	-0.196	3.328
0.6	-1.196	4.841	-0.771	4.088	-0.314	3.284
0.5	-1.371	4.803	-0.919	4.043	-0.439	3.238
0.4	-1.560	4.765	-1.080	3.999	-0.577	3.188
0.3	-1.763	4.723	-1.255	3.948	-0.729	3.139
0.2	-2.005	4.678	-1.440	3.896	-0.887	3.089
0.1	-2.221	4.632	-1.653	3.845	-1.053	3.040
0.0	-2.534	4.587	-1.894	3.794	-1.239	2.986
-0.1	-2.819	4.542	-2.142	3.741	-1.451	2.914
-0.2	-3.116	4.495	-2.430	3.675	-1.692	2.841
-0.3	-3.463	4.436	-2.732	3.609	-1.970	2.769
-0.4	-3.876	4.376	-3.051	3.543	-2.248	2.683
-0.5	-4.289	4.316	-3.442	3.470	-2.587	2.593
-0.6	-4.758	4.256	-3.870	3.385	-2.944	2.502
-0.7	-5.293	4.166	-4.332	3.299	-3.374	2.375
-0.8	-5.779	4.072	-4.861	3.199	-3.840	2.246
-0.9	-6.415	3.967	-5.423	3.080	-4.369	2.062
-1.0	-7.122	3.827	-6.031	2.935	-4.940	1.884
-1.1	-7.836	3.683	-6.720	2.738	-5.559	1.695
-1.2	-8.528	3.532	-7.450	2.541	-6.239	1.437
-1.3	-9.362	3.352	-8.209	2.270	-6.961	1.133
-1.4	-10.199	3.132	-9.048	2.000	-7.761	0.795
-1.5	-11.147	2.870	-9.895	1.762	-8.493	0.429
-1.6	-12.102	2.583	-10.801	1.384	-9.368	0.035
-1.7	-13.124	2.203	-11.733	1.018	-10.236	-0.412
-1.8	-14.155	1.884	-12.721	0.598	-11.184	-0.823
-1.9	-15.242	1.492	-13.732	0.152	-12.142	-1.348

(continued on the next page)

DF-GLS	95%		90%		80%	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-16.365	1.087	-14.833	-0.312	-13.158	-1.887
-2.1	-17.574	0.594	-15.966	-0.793	-14.191	-2.465
-2.2	-18.783	0.104	-17.135	-1.377	-15.298	-3.075
-2.3	-19.991	-0.413	-18.319	-1.981	-16.427	-3.742
-2.4	-21.328	-1.020	-19.541	-2.598	-17.593	-4.420
-2.5	-22.704	-1.510	-20.838	-3.242	-18.818	-5.118
-2.6	-24.112	-2.243	-22.188	-3.970	-20.076	-5.924
-2.7	-25.457	-2.846	-23.586	-4.711	-21.357	-6.690
-2.8	-26.912	-3.609	-24.961	-5.410	-22.695	-7.544
-2.9	-28.436	-4.304	-26.359	-6.214	-24.090	-8.420
-3.0	-29.943	-5.118	-27.860	-7.041	-25.472	-9.247
-3.1	-31.536	-5.936	-29.298	-8.066	-26.894	-10.253
-3.2	-33.097	-6.837	-30.832	-8.836	-28.342	-11.226
-3.3	-34.717	-7.833	-32.450	-9.830	-29.799	-12.225
-3.4	-36.430	-8.621	-34.036	-10.826	-31.332	-13.331
-3.5	-38.121	-9.597	-35.669	-11.803	-32.926	-14.434
-3.6	-39.948	-10.627	-37.388	-12.916	-34.559	-15.535
-3.7	-41.707	-11.626	-39.168	-14.057	-36.197	-16.676
-3.8	-43.497	-12.720	-40.906	-15.110	-37.834	-17.847
-3.9	-45.364	-13.870	-42.639	-16.190	-39.660	-19.036
-4.0	-47.207	-14.964	-44.481	-17.412	-41.379	-20.263
-4.1	-49.147	-16.058	-46.362	-18.624	-43.127	-21.640
-4.2	-51.119	-17.293	-48.211	-19.827	-44.920	-22.950
-4.3	-53.097	-18.514	-50.144	-21.166	-46.779	-24.251
-4.4	-55.107	-19.717	-52.078	-22.513	-48.674	-25.623
-4.5	-57.120	-21.067	-54.094	-23.837	-50.534	-27.018
-4.6	-59.173	-22.422	-56.039	-25.237	-52.521	-28.458
-4.7	-61.304	-23.698	-58.063	-26.666	-54.518	-29.985
-4.8	-63.574	-25.123	-60.181	-28.113	-56.448	-31.511
-4.9	-65.635	-26.584	-62.272	-29.528	-58.491	-33.053
-5.0	-67.777	-28.030	-64.454	-31.057	-60.606	-34.615

Table A.4: Confidence Interval for c in the Bonferroni Q-test: $\delta \in (-1.0, -0.9)$

DF-GLS	$\delta = -0.999$		$\delta = -0.975$		$\delta = -0.950$		$\delta = -0.925$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	-0.282	4.148	-0.228	3.720	-0.228	3.462	-0.228	3.294
0.9	-0.395	4.106	-0.340	3.676	-0.340	3.418	-0.340	3.254
0.8	-0.509	4.063	-0.455	3.632	-0.455	3.373	-0.455	3.211
0.7	-0.639	4.021	-0.580	3.588	-0.580	3.328	-0.580	3.168
0.6	-0.771	3.977	-0.712	3.544	-0.712	3.284	-0.712	3.124
0.5	-0.919	3.931	-0.855	3.500	-0.855	3.238	-0.855	3.080
0.4	-1.080	3.885	-1.004	3.451	-1.004	3.188	-1.004	3.037
0.3	-1.255	3.840	-1.178	3.400	-1.178	3.139	-1.178	2.991
0.2	-1.440	3.794	-1.364	3.350	-1.364	3.089	-1.364	2.930
0.1	-1.653	3.747	-1.567	3.300	-1.567	3.040	-1.567	2.869
0.0	-1.894	3.684	-1.799	3.250	-1.799	2.986	-1.799	2.808
-0.1	-2.142	3.620	-2.062	3.184	-2.062	2.914	-2.062	2.747
-0.2	-2.430	3.557	-2.322	3.117	-2.322	2.841	-2.322	2.664
-0.3	-2.732	3.493	-2.637	3.051	-2.637	2.769	-2.637	2.580
-0.4	-3.051	3.421	-2.949	2.977	-2.949	2.683	-2.949	2.497
-0.5	-3.442	3.350	-3.308	2.884	-3.308	2.593	-3.308	2.398
-0.6	-3.870	3.279	-3.749	2.790	-3.749	2.502	-3.749	2.298
-0.7	-4.332	3.187	-4.209	2.682	-4.209	2.375	-4.209	2.174
-0.8	-4.861	3.081	-4.735	2.565	-4.735	2.246	-4.735	2.026
-0.9	-5.423	2.958	-5.293	2.413	-5.293	2.062	-5.293	1.869
-1.0	-6.031	2.784	-5.883	2.217	-5.883	1.884	-5.883	1.695
-1.1	-6.720	2.604	-6.562	1.999	-6.562	1.695	-6.562	1.472
-1.2	-7.450	2.384	-7.319	1.805	-7.319	1.437	-7.319	1.176
-1.3	-8.209	2.112	-8.065	1.552	-8.065	1.133	-8.065	0.870
-1.4	-9.048	1.870	-8.883	1.216	-8.883	0.795	-8.883	0.521
-1.5	-9.895	1.581	-9.723	0.891	-9.723	0.429	-9.723	0.148
-1.6	-10.801	1.200	-10.619	0.500	-10.619	0.035	-10.619	-0.259
-1.7	-11.733	0.849	-11.551	0.064	-11.551	-0.412	-11.551	-0.677
-1.8	-12.721	0.402	-12.527	-0.390	-12.527	-0.823	-12.527	-1.177
-1.9	-13.732	-0.036	-13.546	-0.815	-13.546	-1.348	-13.546	-1.697

(continued on the next page)

DF-GLS	$\delta = -0.999$		$\delta = -0.975$		$\delta = -0.950$		$\delta = -0.925$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-14.833	-0.537	-14.618	-1.360	-14.618	-1.887	-14.618	-2.221
-2.1	-15.966	-1.035	-15.733	-1.908	-15.733	-2.465	-15.733	-2.834
-2.2	-17.135	-1.565	-16.894	-2.509	-16.894	-3.075	-16.894	-3.483
-2.3	-18.319	-2.176	-18.072	-3.119	-18.072	-3.742	-18.072	-4.111
-2.4	-19.541	-2.800	-19.301	-3.802	-19.301	-4.420	-19.301	-4.853
-2.5	-20.838	-3.505	-20.571	-4.503	-20.571	-5.118	-20.571	-5.578
-2.6	-22.188	-4.177	-21.895	-5.214	-21.895	-5.924	-21.895	-6.333
-2.7	-23.586	-4.963	-23.311	-6.004	-23.311	-6.690	-23.311	-7.147
-2.8	-24.961	-5.694	-24.699	-6.799	-24.699	-7.544	-24.699	-8.057
-2.9	-26.359	-6.496	-26.059	-7.752	-26.059	-8.420	-26.059	-8.869
-3.0	-27.860	-7.388	-27.543	-8.547	-27.543	-9.247	-27.543	-9.811
-3.1	-29.298	-8.255	-28.992	-9.509	-28.992	-10.253	-28.992	-10.767
-3.2	-30.832	-9.108	-30.478	-10.437	-30.478	-11.226	-30.478	-11.730
-3.3	-32.450	-10.143	-32.089	-11.401	-32.089	-12.225	-32.089	-12.802
-3.4	-34.036	-11.135	-33.704	-12.429	-33.704	-13.331	-33.704	-13.917
-3.5	-35.669	-12.144	-35.327	-13.561	-35.327	-14.434	-35.327	-15.001
-3.6	-37.388	-13.271	-37.006	-14.652	-37.006	-15.535	-37.006	-16.091
-3.7	-39.168	-14.373	-38.773	-15.715	-38.773	-16.676	-38.773	-17.273
-3.8	-40.906	-15.428	-40.552	-16.878	-40.552	-17.847	-40.552	-18.464
-3.9	-42.639	-16.565	-42.239	-18.078	-42.239	-19.036	-42.239	-19.670
-4.0	-44.481	-17.799	-44.044	-19.284	-44.044	-20.263	-44.044	-20.977
-4.1	-46.362	-18.993	-45.908	-20.555	-45.908	-21.640	-45.908	-22.309
-4.2	-48.211	-20.204	-47.797	-21.920	-47.797	-22.950	-47.797	-23.597
-4.3	-50.144	-21.555	-49.710	-23.229	-49.710	-24.251	-49.710	-24.939
-4.4	-52.078	-22.911	-51.663	-24.565	-51.663	-25.623	-51.663	-26.312
-4.5	-54.094	-24.274	-53.659	-25.947	-53.659	-27.018	-53.659	-27.702
-4.6	-56.039	-25.659	-55.614	-27.379	-55.614	-28.458	-55.614	-29.222
-4.7	-58.063	-27.094	-57.614	-28.836	-57.614	-29.985	-57.614	-30.760
-4.8	-60.181	-28.531	-59.710	-30.326	-59.710	-31.511	-59.710	-32.288
-4.9	-62.272	-29.945	-61.771	-31.868	-61.771	-33.053	-61.771	-33.854
-5.0	-64.454	-31.539	-63.942	-33.403	-63.942	-34.615	-63.942	-35.419

Table A.5: Confidence Interval for c in the Bonferroni Q-test: $\delta \in [-0.9, -0.8)$

DF-GLS	$\delta = -0.900$		$\delta = -0.875$		$\delta = -0.850$		$\delta = -0.825$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	-0.183	3.161	-0.183	3.083	-0.183	3.013	-0.183	2.936
0.9	-0.290	3.121	-0.290	3.046	-0.290	2.969	-0.290	2.888
0.8	-0.406	3.081	-0.406	3.008	-0.406	2.919	-0.406	2.840
0.7	-0.524	3.041	-0.524	2.959	-0.524	2.870	-0.524	2.791
0.6	-0.658	3.002	-0.658	2.906	-0.658	2.821	-0.658	2.742
0.5	-0.797	2.947	-0.797	2.854	-0.797	2.772	-0.797	2.685
0.4	-0.945	2.892	-0.945	2.801	-0.945	2.716	-0.945	2.628
0.3	-1.111	2.837	-1.111	2.748	-1.111	2.654	-1.111	2.571
0.2	-1.294	2.782	-1.294	2.682	-1.294	2.593	-1.294	2.515
0.1	-1.493	2.719	-1.493	2.615	-1.493	2.531	-1.493	2.450
0.0	-1.717	2.646	-1.717	2.549	-1.717	2.465	-1.717	2.383
-0.1	-1.980	2.573	-1.980	2.479	-1.980	2.394	-1.980	2.316
-0.2	-2.232	2.499	-2.232	2.403	-2.232	2.323	-2.232	2.248
-0.3	-2.535	2.416	-2.535	2.327	-2.535	2.251	-2.535	2.146
-0.4	-2.862	2.333	-2.862	2.251	-2.862	2.142	-2.862	2.045
-0.5	-3.199	2.250	-3.199	2.136	-3.199	2.032	-3.199	1.941
-0.6	-3.632	2.122	-3.632	2.021	-3.632	1.919	-3.632	1.837
-0.7	-4.084	1.993	-4.084	1.894	-4.084	1.804	-4.084	1.723
-0.8	-4.591	1.852	-4.591	1.763	-4.591	1.658	-4.591	1.560
-0.9	-5.156	1.696	-5.156	1.583	-5.156	1.478	-5.156	1.355
-1.0	-5.758	1.499	-5.758	1.358	-5.758	1.235	-5.758	1.114
-1.1	-6.420	1.230	-6.420	1.091	-6.420	0.968	-6.420	0.857
-1.2	-7.185	0.945	-7.185	0.814	-7.185	0.684	-7.185	0.548
-1.3	-7.930	0.634	-7.930	0.478	-7.930	0.334	-7.930	0.203
-1.4	-8.712	0.273	-8.712	0.125	-8.712	-0.003	-8.712	-0.149
-1.5	-9.543	-0.099	-9.543	-0.274	-9.543	-0.429	-9.543	-0.565
-1.6	-10.452	-0.535	-10.452	-0.675	-10.452	-0.831	-10.452	-1.005
-1.7	-11.386	-0.978	-11.386	-1.143	-11.386	-1.302	-11.386	-1.467
-1.8	-12.346	-1.460	-12.346	-1.658	-12.346	-1.829	-12.346	-1.998
-1.9	-13.359	-2.000	-13.359	-2.175	-13.359	-2.373	-13.359	-2.558

(continued on the next page)

DF-GLS	$\delta = -0.900$		$\delta = -0.875$		$\delta = -0.850$		$\delta = -0.825$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-14.409	-2.563	-14.409	-2.754	-14.409	-2.957	-14.409	-3.135
-2.1	-15.524	-3.166	-15.524	-3.380	-15.524	-3.589	-15.524	-3.783
-2.2	-16.676	-3.830	-16.676	-4.043	-16.676	-4.216	-16.676	-4.429
-2.3	-17.855	-4.503	-17.855	-4.771	-17.855	-4.953	-17.855	-5.127
-2.4	-19.074	-5.185	-19.074	-5.439	-19.074	-5.660	-19.074	-5.864
-2.5	-20.322	-5.958	-20.322	-6.175	-20.322	-6.429	-20.322	-6.632
-2.6	-21.643	-6.723	-21.643	-6.963	-21.643	-7.246	-21.643	-7.463
-2.7	-23.032	-7.583	-23.032	-7.882	-23.032	-8.111	-23.032	-8.327
-2.8	-24.429	-8.435	-24.429	-8.695	-24.429	-8.954	-24.429	-9.157
-2.9	-25.790	-9.269	-25.790	-9.606	-25.790	-9.864	-25.790	-10.106
-3.0	-27.246	-10.264	-27.246	-10.554	-27.246	-10.815	-27.246	-11.065
-3.1	-28.706	-11.226	-28.706	-11.514	-28.706	-11.771	-28.706	-12.027
-3.2	-30.170	-12.218	-30.170	-12.548	-30.170	-12.848	-30.170	-13.131
-3.3	-31.773	-13.337	-31.773	-13.660	-31.773	-13.963	-31.773	-14.227
-3.4	-33.380	-14.435	-33.380	-14.739	-33.380	-15.025	-33.380	-15.294
-3.5	-34.994	-15.503	-34.994	-15.803	-34.994	-16.095	-34.994	-16.402
-3.6	-36.650	-16.631	-36.650	-16.963	-36.650	-17.287	-36.650	-17.589
-3.7	-38.396	-17.811	-38.396	-18.149	-38.396	-18.487	-38.396	-18.796
-3.8	-40.222	-19.013	-40.222	-19.359	-40.222	-19.698	-40.222	-20.015
-3.9	-41.891	-20.245	-41.891	-20.626	-41.891	-20.994	-41.891	-21.330
-4.0	-43.681	-21.594	-43.681	-21.958	-43.681	-22.307	-43.681	-22.630
-4.1	-45.507	-22.894	-45.507	-23.247	-45.507	-23.587	-45.507	-23.913
-4.2	-47.393	-24.185	-47.393	-24.569	-47.393	-24.929	-47.393	-25.256
-4.3	-49.294	-25.558	-49.294	-25.939	-49.294	-26.310	-49.294	-26.646
-4.4	-51.229	-26.947	-51.229	-27.341	-51.229	-27.722	-51.229	-28.090
-4.5	-53.225	-28.377	-53.225	-28.803	-53.225	-29.205	-53.225	-29.588
-4.6	-55.187	-29.912	-55.187	-30.325	-55.187	-30.724	-55.187	-31.116
-4.7	-57.180	-31.436	-57.180	-31.868	-57.180	-32.267	-57.180	-32.662
-4.8	-59.253	-32.992	-59.253	-33.425	-59.253	-33.828	-59.253	-34.221
-4.9	-61.330	-34.560	-61.330	-34.991	-61.330	-35.401	-61.330	-35.797
-5.0	-63.478	-36.123	-63.478	-36.593	-63.478	-37.029	-63.478	-37.451

Table A.6: Confidence Interval for c in the Bonferroni Q-test: $\delta \in [-0.8, -0.7)$

DF-GLS	$\delta = -0.800$		$\delta = -0.775$		$\delta = -0.750$		$\delta = -0.725$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	-0.140	2.859	-0.140	2.792	-0.140	2.728	-0.140	2.695
0.9	-0.243	2.814	-0.243	2.749	-0.243	2.678	-0.243	2.646
0.8	-0.359	2.769	-0.359	2.697	-0.359	2.628	-0.359	2.597
0.7	-0.475	2.719	-0.475	2.646	-0.475	2.578	-0.475	2.548
0.6	-0.607	2.664	-0.607	2.594	-0.607	2.529	-0.607	2.499
0.5	-0.743	2.610	-0.743	2.542	-0.743	2.475	-0.743	2.441
0.4	-0.893	2.555	-0.893	2.489	-0.893	2.416	-0.893	2.384
0.3	-1.050	2.501	-1.050	2.428	-1.050	2.358	-1.050	2.327
0.2	-1.227	2.436	-1.227	2.367	-1.227	2.300	-1.227	2.270
0.1	-1.428	2.372	-1.428	2.306	-1.428	2.238	-1.428	2.199
0.0	-1.647	2.307	-1.647	2.243	-1.647	2.157	-1.647	2.120
-0.1	-1.900	2.240	-1.900	2.155	-1.900	2.075	-1.900	2.041
-0.2	-2.161	2.149	-2.161	2.067	-2.161	1.994	-2.161	1.960
-0.3	-2.442	2.057	-2.442	1.979	-2.442	1.909	-2.442	1.878
-0.4	-2.785	1.964	-2.785	1.889	-2.785	1.824	-2.785	1.795
-0.5	-3.109	1.867	-3.109	1.799	-3.109	1.732	-3.109	1.691
-0.6	-3.519	1.770	-3.519	1.685	-3.519	1.598	-3.519	1.561
-0.7	-3.962	1.625	-3.962	1.540	-3.962	1.450	-3.962	1.404
-0.8	-4.460	1.457	-4.460	1.353	-4.460	1.262	-4.460	1.218
-0.9	-5.031	1.244	-5.031	1.139	-5.031	1.044	-5.031	0.999
-1.0	-5.639	1.007	-5.639	0.905	-5.639	0.810	-5.639	0.768
-1.1	-6.292	0.751	-6.292	0.633	-6.292	0.523	-6.292	0.473
-1.2	-7.018	0.421	-7.018	0.305	-7.018	0.204	-7.018	0.153
-1.3	-7.796	0.095	-7.796	-0.013	-7.796	-0.128	-7.796	-0.185
-1.4	-8.542	-0.282	-8.542	-0.413	-8.542	-0.537	-8.542	-0.590
-1.5	-9.395	-0.680	-9.395	-0.812	-9.395	-0.965	-9.395	-1.028
-1.6	-10.289	-1.139	-10.289	-1.262	-10.289	-1.409	-10.289	-1.473
-1.7	-11.227	-1.628	-11.227	-1.768	-11.227	-1.916	-11.227	-1.987
-1.8	-12.178	-2.149	-12.178	-2.298	-12.178	-2.457	-12.178	-2.534
-1.9	-13.191	-2.711	-13.191	-2.879	-13.191	-3.038	-13.191	-3.111

(continued on the next page)

DF-GLS	$\delta = -0.800$		$\delta = -0.775$		$\delta = -0.750$		$\delta = -0.725$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-14.222	-3.304	-14.222	-3.487	-14.222	-3.656	-14.222	-3.747
-2.1	-15.328	-3.976	-15.328	-4.125	-15.328	-4.263	-15.328	-4.350
-2.2	-16.461	-4.662	-16.461	-4.839	-16.461	-4.990	-16.461	-5.060
-2.3	-17.635	-5.303	-17.635	-5.520	-17.635	-5.695	-17.635	-5.778
-2.4	-18.852	-6.062	-18.852	-6.240	-18.852	-6.471	-18.852	-6.560
-2.5	-20.095	-6.835	-20.095	-7.045	-20.095	-7.269	-20.095	-7.362
-2.6	-21.406	-7.713	-21.406	-7.920	-21.406	-8.119	-21.406	-8.223
-2.7	-22.774	-8.542	-22.774	-8.754	-22.774	-8.983	-22.774	-9.060
-2.8	-24.181	-9.416	-24.181	-9.647	-24.181	-9.861	-24.181	-9.964
-2.9	-25.546	-10.349	-25.546	-10.579	-25.546	-10.808	-25.546	-10.914
-3.0	-26.983	-11.303	-26.983	-11.526	-26.983	-11.754	-26.983	-11.866
-3.1	-28.446	-12.297	-28.446	-12.551	-28.446	-12.813	-28.446	-12.943
-3.2	-29.887	-13.398	-29.887	-13.656	-29.887	-13.913	-29.887	-14.040
-3.3	-31.471	-14.485	-31.471	-14.740	-31.471	-14.986	-31.471	-15.105
-3.4	-33.080	-15.557	-33.080	-15.815	-33.080	-16.061	-33.080	-16.188
-3.5	-34.693	-16.694	-34.693	-16.977	-34.693	-17.233	-34.693	-17.358
-3.6	-36.321	-17.876	-36.321	-18.160	-36.321	-18.420	-36.321	-18.551
-3.7	-38.031	-19.086	-38.031	-19.370	-38.031	-19.637	-38.031	-19.770
-3.8	-39.856	-20.322	-39.856	-20.623	-39.856	-20.908	-39.856	-21.043
-3.9	-41.580	-21.629	-41.580	-21.925	-41.580	-22.211	-41.580	-22.347
-4.0	-43.334	-22.932	-43.334	-23.252	-43.334	-23.559	-43.334	-23.701
-4.1	-45.143	-24.242	-45.143	-24.580	-45.143	-24.885	-45.143	-25.026
-4.2	-47.012	-25.589	-47.012	-25.907	-47.012	-26.221	-47.012	-26.377
-4.3	-48.910	-27.002	-48.910	-27.344	-48.910	-27.676	-48.910	-27.821
-4.4	-50.839	-28.456	-50.839	-28.806	-50.839	-29.145	-50.839	-29.300
-4.5	-52.841	-29.941	-52.841	-30.293	-52.841	-30.648	-52.841	-30.817
-4.6	-54.807	-31.474	-54.807	-31.833	-54.807	-32.191	-54.807	-32.364
-4.7	-56.749	-33.024	-56.749	-33.393	-56.749	-33.740	-56.749	-33.915
-4.8	-58.834	-34.594	-58.834	-34.964	-58.834	-35.312	-58.834	-35.483
-4.9	-60.918	-36.194	-60.918	-36.573	-60.918	-36.946	-60.918	-37.129
-5.0	-63.033	-37.855	-63.033	-38.238	-63.033	-38.611	-63.033	-38.789

Table A.7: Confidence Interval for c in the Bonferroni Q-test: $\delta \in [-0.7, -0.6)$

DF-GLS	$\delta = -0.700$		$\delta = -0.675$		$\delta = -0.650$		$\delta = -0.625$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	-0.096	2.633	-0.096	2.577	-0.096	2.524	-0.052	2.499
0.9	-0.203	2.586	-0.203	2.531	-0.203	2.476	-0.161	2.448
0.8	-0.315	2.540	-0.315	2.483	-0.315	2.425	-0.272	2.398
0.7	-0.432	2.492	-0.432	2.429	-0.432	2.373	-0.390	2.347
0.6	-0.557	2.436	-0.557	2.376	-0.557	2.321	-0.509	2.297
0.5	-0.696	2.380	-0.696	2.322	-0.696	2.270	-0.646	2.245
0.4	-0.843	2.325	-0.843	2.268	-0.843	2.208	-0.789	2.178
0.3	-0.995	2.269	-0.995	2.204	-0.995	2.139	-0.947	2.112
0.2	-1.172	2.202	-1.172	2.133	-1.172	2.071	-1.118	2.045
0.1	-1.362	2.128	-1.362	2.062	-1.362	2.003	-1.301	1.976
0.0	-1.574	2.054	-1.574	1.991	-1.574	1.930	-1.503	1.904
-0.1	-1.819	1.978	-1.819	1.915	-1.819	1.858	-1.741	1.832
-0.2	-2.086	1.898	-2.086	1.839	-2.086	1.785	-2.021	1.760
-0.3	-2.356	1.819	-2.356	1.763	-2.356	1.694	-2.277	1.659
-0.4	-2.702	1.734	-2.702	1.656	-2.702	1.585	-2.603	1.553
-0.5	-3.019	1.613	-3.019	1.541	-3.019	1.468	-2.942	1.430
-0.6	-3.420	1.489	-3.420	1.399	-3.420	1.322	-3.330	1.289
-0.7	-3.875	1.320	-3.875	1.238	-3.875	1.157	-3.785	1.121
-0.8	-4.367	1.127	-4.367	1.045	-4.367	0.972	-4.277	0.931
-0.9	-4.920	0.911	-4.920	0.830	-4.920	0.753	-4.824	0.707
-1.0	-5.522	0.669	-5.522	0.572	-5.522	0.482	-5.417	0.435
-1.1	-6.166	0.373	-6.166	0.279	-6.166	0.185	-6.043	0.140
-1.2	-6.890	0.059	-6.890	-0.044	-6.890	-0.140	-6.769	-0.188
-1.3	-7.644	-0.307	-7.644	-0.429	-7.644	-0.533	-7.484	-0.576
-1.4	-8.396	-0.696	-8.396	-0.811	-8.396	-0.935	-8.259	-0.994
-1.5	-9.250	-1.142	-9.250	-1.258	-9.250	-1.377	-9.117	-1.437
-1.6	-10.134	-1.606	-10.134	-1.718	-10.134	-1.844	-9.985	-1.906
-1.7	-11.075	-2.116	-11.075	-2.253	-11.075	-2.383	-10.923	-2.447
-1.8	-12.018	-2.671	-12.018	-2.809	-12.018	-2.940	-11.864	-3.005
-1.9	-13.026	-3.248	-13.026	-3.396	-13.026	-3.548	-12.871	-3.620

(continued on the next page)

DF-GLS	$\delta = -0.700$		$\delta = -0.675$		$\delta = -0.650$		$\delta = -0.625$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-14.038	-3.894	-14.038	-4.035	-14.038	-4.159	-13.889	-4.218
-2.1	-15.140	-4.531	-15.140	-4.733	-15.140	-4.867	-14.970	-4.925
-2.2	-16.259	-5.200	-16.259	-5.388	-16.259	-5.558	-16.064	-5.627
-2.3	-17.438	-5.975	-17.438	-6.128	-17.438	-6.291	-17.248	-6.384
-2.4	-18.651	-6.708	-18.651	-6.879	-18.651	-7.063	-18.454	-7.164
-2.5	-19.895	-7.546	-19.895	-7.747	-19.895	-7.912	-19.699	-7.994
-2.6	-21.188	-8.406	-21.188	-8.585	-21.188	-8.748	-20.977	-8.846
-2.7	-22.528	-9.214	-22.528	-9.457	-22.528	-9.647	-22.293	-9.732
-2.8	-23.942	-10.172	-23.942	-10.377	-23.942	-10.572	-23.711	-10.674
-2.9	-25.300	-11.130	-25.300	-11.337	-25.300	-11.534	-25.088	-11.638
-3.0	-26.719	-12.098	-26.719	-12.332	-26.719	-12.549	-26.487	-12.662
-3.1	-28.200	-13.179	-28.200	-13.405	-28.200	-13.616	-27.952	-13.723
-3.2	-29.649	-14.258	-29.649	-14.477	-29.649	-14.689	-29.404	-14.795
-3.3	-31.196	-15.328	-31.196	-15.549	-31.196	-15.767	-30.926	-15.871
-3.4	-32.802	-16.437	-32.802	-16.685	-32.802	-16.923	-32.517	-17.033
-3.5	-34.425	-17.608	-34.425	-17.867	-34.425	-18.104	-34.153	-18.214
-3.6	-36.017	-18.813	-36.017	-19.075	-36.017	-19.312	-35.760	-19.422
-3.7	-37.724	-20.035	-37.724	-20.302	-37.724	-20.559	-37.441	-20.679
-3.8	-39.531	-21.316	-39.531	-21.582	-39.531	-21.857	-39.213	-21.981
-3.9	-41.273	-22.624	-41.273	-22.900	-41.273	-23.178	-40.963	-23.303
-4.0	-43.025	-23.965	-43.025	-24.241	-43.025	-24.515	-42.706	-24.641
-4.1	-44.822	-25.305	-44.822	-25.602	-44.822	-25.874	-44.508	-25.997
-4.2	-46.660	-26.694	-46.660	-27.005	-46.660	-27.305	-46.325	-27.442
-4.3	-48.545	-28.137	-48.545	-28.440	-48.545	-28.749	-48.226	-28.887
-4.4	-50.462	-29.604	-50.462	-29.903	-50.462	-30.211	-50.097	-30.358
-4.5	-52.466	-31.131	-52.466	-31.447	-52.466	-31.760	-52.094	-31.908
-4.6	-54.463	-32.690	-54.463	-33.000	-54.463	-33.300	-54.115	-33.445
-4.7	-56.353	-34.258	-56.353	-34.560	-56.353	-34.864	-56.001	-35.015
-4.8	-58.454	-35.817	-58.454	-36.136	-58.454	-36.472	-58.063	-36.633
-4.9	-60.538	-37.476	-60.538	-37.812	-60.538	-38.135	-60.177	-38.296
-5.0	-62.628	-39.135	-62.628	-39.463	-62.628	-39.798	-62.260	-39.964

Table A.8: Confidence Interval for c in the Bonferroni Q-test: $\delta \in [-0.6, -0.5)$

DF-GLS	$\delta = -0.600$		$\delta = -0.575$		$\delta = -0.550$		$\delta = -0.525$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	-0.052	2.443	-0.052	2.392	-0.012	2.347	-0.012	2.303
0.9	-0.161	2.394	-0.161	2.345	-0.122	2.302	-0.122	2.257
0.8	-0.272	2.345	-0.272	2.298	-0.231	2.256	-0.231	2.202
0.7	-0.390	2.296	-0.390	2.250	-0.349	2.198	-0.349	2.146
0.6	-0.509	2.246	-0.509	2.190	-0.469	2.138	-0.469	2.089
0.5	-0.646	2.183	-0.646	2.129	-0.603	2.079	-0.603	2.032
0.4	-0.789	2.119	-0.789	2.068	-0.742	2.019	-0.742	1.971
0.3	-0.947	2.056	-0.947	2.007	-0.898	1.955	-0.898	1.907
0.2	-1.118	1.992	-1.118	1.939	-1.063	1.889	-1.063	1.842
0.1	-1.301	1.922	-1.301	1.871	-1.242	1.823	-1.242	1.778
0.0	-1.503	1.852	-1.503	1.802	-1.450	1.756	-1.450	1.699
-0.1	-1.741	1.782	-1.741	1.727	-1.679	1.666	-1.679	1.609
-0.2	-2.021	1.695	-2.021	1.630	-1.952	1.573	-1.952	1.519
-0.3	-2.277	1.593	-2.277	1.534	-2.210	1.474	-2.210	1.411
-0.4	-2.603	1.489	-2.603	1.418	-2.516	1.355	-2.516	1.298
-0.5	-2.942	1.358	-2.942	1.293	-2.877	1.233	-2.877	1.169
-0.6	-3.330	1.220	-3.330	1.146	-3.238	1.084	-3.238	1.026
-0.7	-3.785	1.050	-3.785	0.984	-3.698	0.917	-3.698	0.849
-0.8	-4.277	0.855	-4.277	0.787	-4.181	0.719	-4.181	0.638
-0.9	-4.824	0.622	-4.824	0.537	-4.724	0.459	-4.724	0.386
-1.0	-5.417	0.350	-5.417	0.270	-5.324	0.186	-5.324	0.107
-1.1	-6.043	0.051	-6.043	-0.036	-5.932	-0.120	-5.932	-0.202
-1.2	-6.769	-0.289	-6.769	-0.389	-6.651	-0.484	-6.651	-0.570
-1.3	-7.484	-0.666	-7.484	-0.752	-7.385	-0.863	-7.385	-0.974
-1.4	-8.259	-1.096	-8.259	-1.194	-8.153	-1.294	-8.153	-1.399
-1.5	-9.117	-1.546	-9.117	-1.645	-8.998	-1.742	-8.998	-1.854
-1.6	-9.985	-2.035	-9.985	-2.146	-9.845	-2.253	-9.845	-2.370
-1.7	-10.923	-2.575	-10.923	-2.693	-10.773	-2.806	-10.773	-2.917
-1.8	-11.864	-3.132	-11.864	-3.266	-11.714	-3.390	-11.714	-3.522
-1.9	-12.871	-3.745	-12.871	-3.884	-12.713	-4.011	-12.713	-4.121

(continued on the next page)

DF-GLS	$\delta=-0.600$		$\delta=-0.575$		$\delta=-0.550$		$\delta=-0.525$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-13.889	-4.367	-13.889	-4.520	-13.738	-4.679	-13.738	-4.817
-2.1	-14.970	-5.053	-14.970	-5.185	-14.811	-5.333	-14.811	-5.487
-2.2	-16.064	-5.766	-16.064	-5.942	-15.900	-6.094	-15.900	-6.217
-2.3	-17.248	-6.546	-17.248	-6.688	-17.070	-6.815	-17.070	-6.954
-2.4	-18.454	-7.332	-18.454	-7.480	-18.261	-7.659	-18.261	-7.821
-2.5	-19.699	-8.165	-19.699	-8.323	-19.503	-8.508	-19.503	-8.647
-2.6	-20.977	-9.034	-20.977	-9.162	-20.776	-9.338	-20.776	-9.548
-2.7	-22.293	-9.907	-22.293	-10.084	-22.076	-10.263	-22.076	-10.444
-2.8	-23.711	-10.861	-23.711	-11.048	-23.484	-11.227	-23.484	-11.405
-2.9	-25.088	-11.827	-25.088	-12.014	-24.872	-12.209	-24.872	-12.398
-3.0	-26.487	-12.867	-26.487	-13.068	-26.262	-13.259	-26.262	-13.444
-3.1	-27.952	-13.922	-27.952	-14.125	-27.720	-14.318	-27.720	-14.509
-3.2	-29.404	-14.996	-29.404	-15.198	-29.177	-15.398	-29.177	-15.595
-3.3	-30.926	-16.078	-30.926	-16.299	-30.682	-16.524	-30.682	-16.738
-3.4	-32.517	-17.262	-32.517	-17.485	-32.257	-17.707	-32.257	-17.914
-3.5	-34.153	-18.450	-34.153	-18.674	-33.875	-18.898	-33.875	-19.113
-3.6	-35.760	-19.646	-35.760	-19.866	-35.496	-20.100	-35.496	-20.337
-3.7	-37.441	-20.923	-37.441	-21.155	-37.166	-21.401	-37.166	-21.628
-3.8	-39.213	-22.236	-39.213	-22.468	-38.916	-22.711	-38.916	-22.946
-3.9	-40.963	-23.558	-40.963	-23.800	-40.682	-24.032	-40.682	-24.280
-4.0	-42.706	-24.896	-42.706	-25.143	-42.408	-25.391	-42.408	-25.639
-4.1	-44.508	-26.262	-44.508	-26.528	-44.209	-26.798	-44.209	-27.056
-4.2	-46.325	-27.715	-46.325	-27.983	-45.998	-28.245	-45.998	-28.503
-4.3	-48.226	-29.167	-48.226	-29.434	-47.910	-29.701	-47.910	-29.966
-4.4	-50.097	-30.660	-50.097	-30.949	-49.762	-31.232	-49.762	-31.506
-4.5	-52.094	-32.204	-52.094	-32.496	-51.751	-32.768	-51.751	-33.032
-4.6	-54.115	-33.737	-54.115	-34.024	-53.802	-34.302	-53.802	-34.580
-4.7	-56.001	-35.319	-56.001	-35.606	-55.655	-35.893	-55.655	-36.190
-4.8	-58.063	-36.951	-58.063	-37.258	-57.704	-37.553	-57.704	-37.853
-4.9	-60.177	-38.611	-60.177	-38.928	-59.811	-39.226	-59.811	-39.519
-5.0	-62.260	-40.272	-62.260	-40.603	-61.893	-40.912	-61.893	-41.218

Table A.9: Confidence Interval for c in the Bonferroni Q-test: $\delta \in [-0.5, -0.4)$

DF-GLS	$\delta = -0.500$		$\delta = -0.475$		$\delta = -0.450$		$\delta = -0.425$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	-0.012	2.263	0.020	2.242	0.020	2.193	0.049	2.128
0.9	-0.122	2.212	-0.083	2.187	-0.083	2.140	-0.046	2.077
0.8	-0.231	2.156	-0.191	2.133	-0.191	2.087	-0.156	2.026
0.7	-0.349	2.101	-0.306	2.078	-0.306	2.034	-0.268	1.972
0.6	-0.469	2.045	-0.428	2.023	-0.428	1.979	-0.391	1.913
0.5	-0.603	1.989	-0.559	1.964	-0.559	1.918	-0.516	1.855
0.4	-0.742	1.925	-0.700	1.901	-0.700	1.857	-0.660	1.797
0.3	-0.898	1.861	-0.853	1.837	-0.853	1.796	-0.809	1.734
0.2	-1.063	1.797	-1.012	1.774	-1.012	1.730	-0.967	1.653
0.1	-1.242	1.727	-1.193	1.698	-1.193	1.648	-1.145	1.573
0.0	-1.450	1.641	-1.392	1.614	-1.392	1.565	-1.339	1.491
-0.1	-1.679	1.555	-1.617	1.530	-1.617	1.478	-1.556	1.395
-0.2	-1.952	1.462	-1.881	1.431	-1.881	1.376	-1.813	1.299
-0.3	-2.210	1.354	-2.149	1.326	-2.149	1.274	-2.093	1.192
-0.4	-2.516	1.245	-2.442	1.213	-2.442	1.153	-2.375	1.075
-0.5	-2.877	1.109	-2.812	1.080	-2.812	1.028	-2.740	0.945
-0.6	-3.238	0.964	-3.162	0.932	-3.162	0.872	-3.083	0.792
-0.7	-3.698	0.787	-3.610	0.759	-3.610	0.691	-3.523	0.590
-0.8	-4.181	0.563	-4.082	0.528	-4.082	0.460	-3.987	0.361
-0.9	-4.724	0.313	-4.625	0.281	-4.625	0.210	-4.515	0.099
-1.0	-5.324	0.028	-5.227	-0.012	-5.227	-0.085	-5.122	-0.193
-1.1	-5.932	-0.288	-5.831	-0.331	-5.831	-0.418	-5.735	-0.539
-1.2	-6.651	-0.653	-6.543	-0.694	-6.543	-0.775	-6.440	-0.912
-1.3	-7.385	-1.063	-7.289	-1.106	-7.289	-1.191	-7.187	-1.321
-1.4	-8.153	-1.503	-8.051	-1.544	-8.051	-1.628	-7.945	-1.755
-1.5	-8.998	-1.966	-8.879	-2.018	-8.879	-2.115	-8.753	-2.262
-1.6	-9.845	-2.485	-9.713	-2.537	-9.713	-2.644	-9.605	-2.803
-1.7	-10.773	-3.027	-10.624	-3.086	-10.624	-3.206	-10.492	-3.368
-1.8	-11.714	-3.640	-11.571	-3.696	-11.571	-3.808	-11.439	-3.971
-1.9	-12.713	-4.226	-12.558	-4.288	-12.558	-4.421	-12.414	-4.640

(continued on the next page)

DF-GLS	$\delta = -0.500$		$\delta = -0.475$		$\delta = -0.450$		$\delta = -0.425$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-13.738	-4.925	-13.578	-4.983	-13.578	-5.102	-13.430	-5.290
-2.1	-14.811	-5.616	-14.646	-5.681	-14.646	-5.820	-14.485	-6.037
-2.2	-15.900	-6.364	-15.747	-6.444	-15.747	-6.573	-15.591	-6.750
-2.3	-17.070	-7.119	-16.902	-7.205	-16.902	-7.352	-16.736	-7.571
-2.4	-18.261	-7.962	-18.078	-8.035	-18.078	-8.170	-17.904	-8.402
-2.5	-19.503	-8.790	-19.315	-8.881	-19.315	-9.033	-19.147	-9.213
-2.6	-20.776	-9.687	-20.582	-9.757	-20.582	-9.919	-20.408	-10.150
-2.7	-22.076	-10.616	-21.884	-10.698	-21.884	-10.870	-21.693	-11.116
-2.8	-23.484	-11.578	-23.267	-11.663	-23.267	-11.833	-23.061	-12.088
-2.9	-24.872	-12.586	-24.658	-12.680	-24.658	-12.858	-24.456	-13.117
-3.0	-26.262	-13.630	-26.050	-13.723	-26.050	-13.895	-25.852	-14.154
-3.1	-27.720	-14.700	-27.504	-14.796	-27.504	-14.972	-27.292	-15.239
-3.2	-29.177	-15.786	-28.943	-15.880	-28.943	-16.059	-28.740	-16.353
-3.3	-30.682	-16.945	-30.415	-17.045	-30.415	-17.236	-30.205	-17.532
-3.4	-32.257	-18.125	-32.003	-18.226	-32.003	-18.426	-31.771	-18.730
-3.5	-33.875	-19.331	-33.627	-19.437	-33.627	-19.638	-33.380	-19.941
-3.6	-35.496	-20.572	-35.246	-20.681	-35.246	-20.902	-35.011	-21.222
-3.7	-37.166	-21.856	-36.897	-21.956	-36.897	-22.194	-36.662	-22.520
-3.8	-38.916	-23.184	-38.629	-23.297	-38.629	-23.526	-38.356	-23.841
-3.9	-40.682	-24.524	-40.426	-24.644	-40.426	-24.867	-40.162	-25.189
-4.0	-42.408	-25.876	-42.136	-25.992	-42.136	-26.229	-41.879	-26.582
-4.1	-44.209	-27.306	-43.904	-27.433	-43.904	-27.671	-43.649	-28.035
-4.2	-45.998	-28.759	-45.716	-28.885	-45.716	-29.132	-45.444	-29.497
-4.3	-47.910	-30.236	-47.613	-30.366	-47.613	-30.631	-47.334	-31.019
-4.4	-49.762	-31.777	-49.470	-31.916	-49.470	-32.173	-49.193	-32.555
-4.5	-51.751	-33.302	-51.418	-33.431	-51.418	-33.687	-51.115	-34.067
-4.6	-53.802	-34.866	-53.456	-35.001	-53.456	-35.264	-53.139	-35.663
-4.7	-55.655	-36.482	-55.356	-36.628	-55.356	-36.901	-55.083	-37.313
-4.8	-57.704	-38.144	-57.361	-38.294	-57.361	-38.578	-57.034	-38.988
-4.9	-59.811	-39.822	-59.470	-39.962	-59.470	-40.257	-59.132	-40.666
-5.0	-61.893	-41.503	-61.556	-41.655	-61.556	-41.924	-61.221	-42.344

Table A.10: Confidence Interval for c in the Bonferroni Q-test: $\delta \in [-0.4, -0.3)$

DF-GLS	$\delta = -0.400$		$\delta = -0.375$		$\delta = -0.350$		$\delta = -0.325$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	0.049	2.086	0.078	2.047	0.106	1.991	0.106	1.954
0.9	-0.046	2.037	-0.008	1.999	0.021	1.938	0.021	1.902
0.8	-0.156	1.985	-0.119	1.944	-0.084	1.885	-0.084	1.850
0.7	-0.268	1.929	-0.230	1.889	-0.196	1.832	-0.196	1.797
0.6	-0.391	1.872	-0.351	1.833	-0.314	1.779	-0.314	1.743
0.5	-0.516	1.815	-0.475	1.778	-0.439	1.718	-0.439	1.673
0.4	-0.660	1.758	-0.619	1.713	-0.577	1.644	-0.577	1.602
0.3	-0.809	1.684	-0.769	1.636	-0.729	1.571	-0.729	1.532
0.2	-0.967	1.607	-0.926	1.560	-0.887	1.498	-0.887	1.453
0.1	-1.145	1.529	-1.098	1.481	-1.053	1.410	-1.053	1.368
0.0	-1.339	1.441	-1.287	1.390	-1.239	1.323	-1.239	1.283
-0.1	-1.556	1.346	-1.497	1.299	-1.451	1.232	-1.451	1.187
-0.2	-1.813	1.251	-1.746	1.199	-1.692	1.129	-1.692	1.085
-0.3	-2.093	1.140	-2.032	1.090	-1.970	1.025	-1.970	0.979
-0.4	-2.375	1.028	-2.313	0.976	-2.248	0.899	-2.248	0.851
-0.5	-2.740	0.889	-2.662	0.836	-2.587	0.766	-2.587	0.711
-0.6	-3.083	0.738	-2.999	0.673	-2.944	0.584	-2.944	0.530
-0.7	-3.523	0.530	-3.444	0.470	-3.374	0.379	-3.374	0.323
-0.8	-3.987	0.301	-3.914	0.240	-3.840	0.141	-3.840	0.077
-0.9	-4.515	0.035	-4.441	-0.034	-4.369	-0.135	-4.369	-0.201
-1.0	-5.122	-0.265	-5.014	-0.342	-4.940	-0.458	-4.940	-0.528
-1.1	-5.735	-0.614	-5.645	-0.682	-5.559	-0.794	-5.559	-0.873
-1.2	-6.440	-1.004	-6.342	-1.082	-6.239	-1.190	-6.239	-1.268
-1.3	-7.187	-1.411	-7.067	-1.503	-6.961	-1.619	-6.961	-1.691
-1.4	-7.945	-1.856	-7.854	-1.956	-7.761	-2.095	-7.761	-2.177
-1.5	-8.753	-2.357	-8.623	-2.453	-8.493	-2.601	-8.493	-2.697
-1.6	-9.605	-2.900	-9.500	-2.996	-9.368	-3.149	-9.368	-3.246
-1.7	-10.492	-3.474	-10.366	-3.585	-10.236	-3.746	-10.236	-3.844
-1.8	-11.439	-4.073	-11.317	-4.170	-11.184	-4.327	-11.184	-4.449
-1.9	-12.414	-4.777	-12.285	-4.870	-12.142	-5.009	-12.142	-5.123

(continued on the next page)

DF-GLS	$\delta = -0.400$		$\delta = -0.375$		$\delta = -0.350$		$\delta = -0.325$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-13.430	-5.422	-13.300	-5.540	-13.158	-5.703	-13.158	-5.832
-2.1	-14.485	-6.157	-14.342	-6.278	-14.191	-6.469	-14.191	-6.582
-2.2	-15.591	-6.868	-15.444	-6.979	-15.298	-7.212	-15.298	-7.351
-2.3	-16.736	-7.739	-16.579	-7.859	-16.427	-8.040	-16.427	-8.172
-2.4	-17.904	-8.541	-17.748	-8.657	-17.593	-8.865	-17.593	-9.025
-2.5	-19.147	-9.397	-18.981	-9.571	-18.818	-9.773	-18.818	-9.901
-2.6	-20.408	-10.310	-20.239	-10.471	-20.076	-10.691	-20.076	-10.845
-2.7	-21.693	-11.271	-21.528	-11.429	-21.357	-11.653	-21.357	-11.809
-2.8	-23.061	-12.249	-22.882	-12.417	-22.695	-12.659	-22.695	-12.819
-2.9	-24.456	-13.283	-24.272	-13.447	-24.090	-13.690	-24.090	-13.841
-3.0	-25.852	-14.331	-25.664	-14.501	-25.472	-14.763	-25.472	-14.925
-3.1	-27.292	-15.414	-27.094	-15.586	-26.894	-15.853	-26.894	-16.023
-3.2	-28.740	-16.539	-28.539	-16.726	-28.342	-17.012	-28.342	-17.193
-3.3	-30.205	-17.713	-29.989	-17.901	-29.799	-18.188	-29.799	-18.375
-3.4	-31.771	-18.920	-31.550	-19.117	-31.332	-19.404	-31.332	-19.586
-3.5	-33.380	-20.146	-33.154	-20.352	-32.926	-20.647	-32.926	-20.841
-3.6	-35.011	-21.427	-34.786	-21.624	-34.559	-21.916	-34.559	-22.123
-3.7	-36.662	-22.729	-36.428	-22.934	-36.197	-23.242	-36.197	-23.446
-3.8	-38.356	-24.051	-38.072	-24.265	-37.834	-24.586	-37.834	-24.793
-3.9	-40.162	-25.408	-39.912	-25.621	-39.660	-25.947	-39.660	-26.166
-4.0	-41.879	-26.817	-41.619	-27.045	-41.379	-27.393	-41.379	-27.607
-4.1	-43.649	-28.270	-43.386	-28.497	-43.127	-28.834	-43.127	-29.055
-4.2	-45.444	-29.727	-45.170	-29.949	-44.920	-30.292	-44.920	-30.540
-4.3	-47.334	-31.268	-47.046	-31.505	-46.779	-31.856	-46.779	-32.089
-4.4	-49.193	-32.794	-48.934	-33.023	-48.674	-33.363	-48.674	-33.589
-4.5	-51.115	-34.312	-50.819	-34.556	-50.534	-34.933	-50.534	-35.184
-4.6	-53.139	-35.916	-52.818	-36.163	-52.521	-36.567	-52.521	-36.831
-4.7	-55.083	-37.587	-54.798	-37.844	-54.518	-38.230	-54.518	-38.499
-4.8	-57.034	-39.257	-56.742	-39.524	-56.448	-39.921	-56.448	-40.179
-4.9	-59.132	-40.934	-58.812	-41.206	-58.491	-41.591	-58.491	-41.847
-5.0	-61.221	-42.626	-60.912	-42.898	-60.606	-43.286	-60.606	-43.540

Table A.11: Confidence Interval for c in the Bonferroni Q-test: $\delta \in [-0.3, -0.2)$

DF-GLS	$\delta = -0.300$		$\delta = -0.275$		$\delta = -0.250$		$\delta = -0.225$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	0.135	1.937	0.162	1.900	0.190	1.884	0.240	1.867
0.9	0.048	1.885	0.073	1.850	0.100	1.834	0.149	1.818
0.8	-0.051	1.833	-0.020	1.799	0.010	1.784	0.058	1.768
0.7	-0.163	1.781	-0.131	1.747	-0.101	1.728	-0.042	1.708
0.6	-0.278	1.721	-0.243	1.679	-0.213	1.660	-0.156	1.641
0.5	-0.403	1.652	-0.369	1.611	-0.337	1.593	-0.273	1.574
0.4	-0.535	1.582	-0.497	1.543	-0.467	1.526	-0.402	1.507
0.3	-0.691	1.512	-0.652	1.470	-0.615	1.449	-0.539	1.428
0.2	-0.849	1.430	-0.811	1.388	-0.770	1.367	-0.695	1.347
0.1	-1.010	1.346	-0.973	1.305	-0.934	1.286	-0.861	1.265
0.0	-1.196	1.262	-1.156	1.218	-1.114	1.195	-1.036	1.173
-0.1	-1.403	1.164	-1.355	1.120	-1.310	1.098	-1.230	1.078
-0.2	-1.639	1.064	-1.583	1.022	-1.527	1.002	-1.440	0.978
-0.3	-1.913	0.954	-1.856	0.905	-1.798	0.881	-1.696	0.859
-0.4	-2.192	0.827	-2.137	0.782	-2.089	0.761	-1.991	0.736
-0.5	-2.508	0.681	-2.446	0.626	-2.391	0.599	-2.280	0.574
-0.6	-2.884	0.505	-2.830	0.449	-2.775	0.419	-2.648	0.391
-0.7	-3.301	0.296	-3.236	0.240	-3.162	0.208	-3.029	0.177
-0.8	-3.771	0.046	-3.705	-0.014	-3.636	-0.047	-3.496	-0.078
-0.9	-4.300	-0.232	-4.226	-0.301	-4.142	-0.333	-3.995	-0.366
-1.0	-4.871	-0.559	-4.801	-0.622	-4.731	-0.655	-4.569	-0.685
-1.1	-5.476	-0.913	-5.403	-0.990	-5.334	-1.027	-5.191	-1.062
-1.2	-6.141	-1.309	-6.044	-1.391	-5.962	-1.430	-5.814	-1.467
-1.3	-6.862	-1.730	-6.772	-1.820	-6.684	-1.866	-6.520	-1.912
-1.4	-7.631	-2.219	-7.499	-2.305	-7.427	-2.350	-7.279	-2.394
-1.5	-8.397	-2.744	-8.302	-2.833	-8.210	-2.878	-8.032	-2.921
-1.6	-9.237	-3.297	-9.144	-3.393	-9.059	-3.444	-8.882	-3.491
-1.7	-10.117	-3.890	-10.001	-3.987	-9.910	-4.033	-9.740	-4.079
-1.8	-11.065	-4.506	-10.952	-4.643	-10.836	-4.708	-10.622	-4.762
-1.9	-12.015	-5.181	-11.902	-5.294	-11.785	-5.346	-11.563	-5.396

(continued on the next page)

DF-GLS	$\delta = -0.300$		$\delta = -0.275$		$\delta = -0.250$		$\delta = -0.225$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-13.038	-5.900	-12.913	-6.023	-12.789	-6.078	-12.549	-6.133
-2.1	-14.066	-6.633	-13.931	-6.734	-13.810	-6.784	-13.573	-6.841
-2.2	-15.164	-7.417	-15.023	-7.549	-14.899	-7.616	-14.648	-7.684
-2.3	-16.279	-8.236	-16.130	-8.371	-16.004	-8.442	-15.760	-8.505
-2.4	-17.450	-9.082	-17.298	-9.189	-17.168	-9.244	-16.906	-9.333
-2.5	-18.664	-9.962	-18.504	-10.106	-18.356	-10.174	-18.066	-10.248
-2.6	-19.918	-10.920	-19.765	-11.068	-19.601	-11.137	-19.310	-11.209
-2.7	-21.193	-11.886	-21.029	-12.032	-20.869	-12.107	-20.572	-12.181
-2.8	-22.513	-12.899	-22.323	-13.054	-22.160	-13.130	-21.857	-13.206
-2.9	-23.906	-13.919	-23.717	-14.082	-23.545	-14.164	-23.217	-14.246
-3.0	-25.296	-15.011	-25.114	-15.174	-24.936	-15.259	-24.595	-15.338
-3.1	-26.711	-16.116	-26.526	-16.287	-26.341	-16.380	-25.984	-16.464
-3.2	-28.153	-17.284	-27.964	-17.465	-27.781	-17.553	-27.429	-17.642
-3.3	-29.614	-18.469	-29.428	-18.658	-29.243	-18.747	-28.896	-18.838
-3.4	-31.134	-19.680	-30.938	-19.866	-30.744	-19.953	-30.391	-20.043
-3.5	-32.710	-20.937	-32.504	-21.127	-32.302	-21.223	-31.930	-21.315
-3.6	-34.335	-22.218	-34.133	-22.413	-33.944	-22.513	-33.560	-22.608
-3.7	-35.975	-23.545	-35.763	-23.739	-35.556	-23.834	-35.163	-23.928
-3.8	-37.617	-24.896	-37.404	-25.100	-37.196	-25.197	-36.795	-25.295
-3.9	-39.417	-26.275	-39.176	-26.493	-38.936	-26.598	-38.505	-26.703
-4.0	-41.153	-27.722	-40.935	-27.934	-40.711	-28.043	-40.291	-28.150
-4.1	-42.881	-29.165	-42.647	-29.378	-42.428	-29.481	-42.029	-29.592
-4.2	-44.665	-30.655	-44.411	-30.883	-44.188	-31.001	-43.750	-31.118
-4.3	-46.504	-32.201	-46.233	-32.423	-45.990	-32.545	-45.546	-32.656
-4.4	-48.409	-33.702	-48.156	-33.930	-47.902	-34.054	-47.418	-34.176
-4.5	-50.259	-35.307	-49.974	-35.554	-49.737	-35.680	-49.275	-35.801
-4.6	-52.243	-36.951	-51.955	-37.208	-51.695	-37.335	-51.218	-37.456
-4.7	-54.256	-38.618	-53.981	-38.879	-53.713	-39.006	-53.224	-39.128
-4.8	-56.178	-40.303	-55.904	-40.551	-55.651	-40.679	-55.181	-40.799
-4.9	-58.189	-41.970	-57.929	-42.216	-57.634	-42.353	-57.148	-42.477
-5.0	-60.310	-43.673	-60.031	-43.934	-59.739	-44.062	-59.211	-44.204

Table A.12: Confidence Interval for c in the Bonferroni Q-test: $\delta \in [-0.2, -0.1)$

DF-GLS	$\delta = -0.200$		$\delta = -0.175$		$\delta = -0.150$		$\delta = -0.125$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	0.264	1.834	0.306	1.818	0.347	1.802	0.387	1.787
0.9	0.175	1.786	0.225	1.770	0.269	1.755	0.308	1.737
0.8	0.082	1.734	0.130	1.713	0.177	1.693	0.224	1.674
0.7	-0.012	1.669	0.036	1.648	0.081	1.629	0.127	1.611
0.6	-0.127	1.603	-0.072	1.584	-0.018	1.566	0.030	1.548
0.5	-0.242	1.537	-0.188	1.519	-0.134	1.502	-0.081	1.481
0.4	-0.372	1.466	-0.312	1.446	-0.249	1.425	-0.199	1.405
0.3	-0.503	1.387	-0.446	1.368	-0.386	1.348	-0.327	1.328
0.2	-0.659	1.309	-0.592	1.291	-0.526	1.271	-0.464	1.252
0.1	-0.823	1.227	-0.746	1.206	-0.683	1.184	-0.619	1.162
0.0	-0.999	1.133	-0.928	1.113	-0.855	1.093	-0.784	1.073
-0.1	-1.192	1.039	-1.117	1.020	-1.044	1.001	-0.973	0.979
-0.2	-1.399	0.933	-1.320	0.911	-1.245	0.889	-1.177	0.868
-0.3	-1.648	0.816	-1.553	0.797	-1.468	0.777	-1.398	0.757
-0.4	-1.939	0.683	-1.842	0.658	-1.742	0.633	-1.655	0.609
-0.5	-2.229	0.527	-2.142	0.503	-2.053	0.476	-1.964	0.449
-0.6	-2.586	0.339	-2.472	0.313	-2.377	0.289	-2.288	0.264
-0.7	-2.972	0.119	-2.879	0.089	-2.782	0.062	-2.680	0.035
-0.8	-3.438	-0.137	-3.326	-0.165	-3.213	-0.192	-3.105	-0.220
-0.9	-3.940	-0.430	-3.825	-0.461	-3.713	-0.492	-3.601	-0.522
-1.0	-4.492	-0.745	-4.367	-0.781	-4.242	-0.819	-4.121	-0.857
-1.1	-5.112	-1.133	-4.976	-1.167	-4.854	-1.200	-4.732	-1.236
-1.2	-5.746	-1.539	-5.602	-1.573	-5.464	-1.608	-5.341	-1.642
-1.3	-6.438	-2.003	-6.287	-2.041	-6.134	-2.079	-5.985	-2.116
-1.4	-7.185	-2.476	-7.002	-2.523	-6.856	-2.568	-6.717	-2.614
-1.5	-7.956	-3.004	-7.814	-3.047	-7.633	-3.091	-7.453	-3.138
-1.6	-8.787	-3.588	-8.578	-3.636	-8.409	-3.683	-8.271	-3.731
-1.7	-9.655	-4.168	-9.485	-4.212	-9.235	-4.261	-9.097	-4.310
-1.8	-10.523	-4.849	-10.330	-4.893	-10.137	-4.936	-9.966	-4.977
-1.9	-11.466	-5.499	-11.269	-5.560	-11.079	-5.612	-10.901	-5.671

(continued on the next page)

DF-GLS	$\delta = -0.200$		$\delta = -0.175$		$\delta = -0.150$		$\delta = -0.125$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-12.445	-6.241	-12.228	-6.298	-12.023	-6.357	-11.841	-6.417
-2.1	-13.469	-6.947	-13.255	-7.002	-13.052	-7.086	-12.857	-7.163
-2.2	-14.534	-7.812	-14.303	-7.871	-14.086	-7.927	-13.888	-7.988
-2.3	-15.640	-8.613	-15.406	-8.664	-15.183	-8.715	-14.980	-8.778
-2.4	-16.781	-9.509	-16.539	-9.576	-16.300	-9.636	-16.085	-9.699
-2.5	-17.938	-10.391	-17.705	-10.465	-17.475	-10.535	-17.250	-10.606
-2.6	-19.175	-11.351	-18.923	-11.423	-18.678	-11.493	-18.438	-11.562
-2.7	-20.429	-12.331	-20.165	-12.408	-19.905	-12.482	-19.668	-12.555
-2.8	-21.707	-13.353	-21.446	-13.432	-21.186	-13.506	-20.934	-13.577
-2.9	-23.062	-14.402	-22.770	-14.486	-22.494	-14.566	-22.226	-14.643
-3.0	-24.445	-15.495	-24.131	-15.576	-23.843	-15.660	-23.569	-15.738
-3.1	-25.842	-16.632	-25.542	-16.718	-25.238	-16.804	-24.956	-16.890
-3.2	-27.262	-17.806	-26.950	-17.892	-26.649	-17.972	-26.366	-18.062
-3.3	-28.715	-19.011	-28.375	-19.097	-28.074	-19.179	-27.787	-19.267
-3.4	-30.208	-20.233	-29.862	-20.321	-29.568	-20.407	-29.269	-20.499
-3.5	-31.760	-21.500	-31.414	-21.588	-31.095	-21.678	-30.780	-21.771
-3.6	-33.379	-22.797	-33.018	-22.892	-32.669	-22.988	-32.326	-23.082
-3.7	-34.980	-24.119	-34.630	-24.223	-34.281	-24.324	-33.953	-24.420
-3.8	-36.596	-25.502	-36.227	-25.598	-35.859	-25.700	-35.537	-25.801
-3.9	-38.295	-26.920	-37.912	-27.024	-37.555	-27.126	-37.198	-27.229
-4.0	-40.088	-28.358	-39.704	-28.468	-39.317	-28.566	-38.941	-28.668
-4.1	-41.818	-29.802	-41.414	-29.905	-41.051	-30.007	-40.689	-30.122
-4.2	-43.546	-31.349	-43.135	-31.467	-42.770	-31.577	-42.393	-31.691
-4.3	-45.339	-32.878	-44.910	-32.994	-44.524	-33.102	-44.139	-33.208
-4.4	-47.190	-34.412	-46.759	-34.537	-46.346	-34.653	-45.950	-34.764
-4.5	-49.050	-36.031	-48.637	-36.160	-48.234	-36.280	-47.854	-36.395
-4.6	-50.969	-37.704	-50.505	-37.819	-50.063	-37.953	-49.666	-38.073
-4.7	-52.973	-39.375	-52.491	-39.491	-52.028	-39.615	-51.582	-39.735
-4.8	-54.939	-41.043	-54.486	-41.157	-54.061	-41.274	-53.598	-41.397
-4.9	-56.898	-42.729	-56.402	-42.851	-55.944	-42.977	-55.537	-43.104
-5.0	-58.955	-44.469	-58.437	-44.608	-57.967	-44.744	-57.545	-44.873

Table A.13: Confidence Interval for c in the Bonferroni Q-test: $\delta \in [-0.1, 0)$

DF-GLS	$\delta = -0.100$		$\delta = -0.075$		$\delta = -0.050$		$\delta = -0.025$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
1.0	0.447	1.758	0.506	1.743	0.601	1.725	0.729	1.689
0.9	0.367	1.699	0.424	1.682	0.521	1.665	0.646	1.630
0.8	0.286	1.638	0.342	1.621	0.437	1.605	0.563	1.571
0.7	0.195	1.577	0.260	1.560	0.351	1.545	0.478	1.511
0.6	0.095	1.516	0.162	1.500	0.265	1.481	0.388	1.441
0.5	-0.006	1.444	0.062	1.425	0.165	1.406	0.298	1.368
0.4	-0.124	1.368	-0.047	1.350	0.061	1.331	0.200	1.295
0.3	-0.243	1.292	-0.171	1.275	-0.053	1.256	0.091	1.218
0.2	-0.382	1.211	-0.301	1.192	-0.182	1.171	-0.022	1.133
0.1	-0.526	1.123	-0.442	1.104	-0.318	1.085	-0.157	1.049
0.0	-0.692	1.035	-0.603	1.017	-0.464	0.999	-0.297	0.957
-0.1	-0.874	0.935	-0.779	0.914	-0.636	0.893	-0.452	0.854
-0.2	-1.074	0.827	-0.972	0.807	-0.824	0.788	-0.632	0.750
-0.3	-1.292	0.708	-1.188	0.683	-1.027	0.659	-0.830	0.613
-0.4	-1.530	0.562	-1.424	0.539	-1.263	0.517	-1.044	0.470
-0.5	-1.826	0.397	-1.705	0.372	-1.515	0.348	-1.303	0.302
-0.6	-2.161	0.209	-2.037	0.181	-1.843	0.155	-1.584	0.102
-0.7	-2.532	-0.018	-2.406	-0.046	-2.209	-0.073	-1.951	-0.128
-0.8	-2.958	-0.276	-2.828	-0.307	-2.610	-0.336	-2.337	-0.397
-0.9	-3.440	-0.584	-3.293	-0.614	-3.065	-0.646	-2.776	-0.707
-1.0	-3.951	-0.928	-3.806	-0.961	-3.572	-0.993	-3.263	-1.062
-1.1	-4.525	-1.306	-4.367	-1.342	-4.110	-1.376	-3.779	-1.442
-1.2	-5.153	-1.711	-4.982	-1.746	-4.722	-1.788	-4.351	-1.875
-1.3	-5.806	-2.191	-5.623	-2.227	-5.347	-2.265	-4.972	-2.345
-1.4	-6.510	-2.705	-6.323	-2.747	-6.004	-2.785	-5.626	-2.859
-1.5	-7.278	-3.236	-7.050	-3.283	-6.738	-3.329	-6.333	-3.418
-1.6	-8.037	-3.820	-7.851	-3.864	-7.485	-3.906	-7.057	-3.991
-1.7	-8.881	-4.413	-8.642	-4.464	-8.273	-4.518	-7.868	-4.636
-1.8	-9.740	-5.084	-9.530	-5.138	-9.121	-5.192	-8.673	-5.295
-1.9	-10.643	-5.782	-10.393	-5.840	-10.001	-5.895	-9.548	-6.001

(continued on the next page)

DF-GLS	$\delta = -0.100$		$\delta = -0.075$		$\delta = -0.050$		$\delta = -0.025$	
	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$	$\underline{c}$	$\overline{c}$
-2.0	-11.582	-6.520	-11.335	-6.570	-10.942	-6.617	-10.427	-6.719
-2.1	-12.569	-7.290	-12.301	-7.346	-11.883	-7.405	-11.363	-7.527
-2.2	-13.596	-8.111	-13.326	-8.169	-12.899	-8.227	-12.327	-8.349
-2.3	-14.671	-8.928	-14.376	-9.002	-13.926	-9.052	-13.352	-9.159
-2.4	-15.776	-9.822	-15.477	-9.882	-15.017	-9.938	-14.403	-10.060
-2.5	-16.928	-10.739	-16.610	-10.810	-16.120	-10.877	-15.499	-11.013
-2.6	-18.097	-11.691	-17.772	-11.759	-17.286	-11.830	-16.638	-11.966
-2.7	-19.324	-12.696	-18.989	-12.769	-18.471	-12.845	-17.813	-12.991
-2.8	-20.568	-13.726	-20.227	-13.801	-19.685	-13.874	-19.009	-14.019
-2.9	-21.832	-14.801	-21.496	-14.882	-20.953	-14.961	-20.225	-15.113
-3.0	-23.169	-15.893	-22.800	-15.975	-22.244	-16.060	-21.518	-16.220
-3.1	-24.540	-17.056	-24.137	-17.139	-23.556	-17.222	-22.824	-17.387
-3.2	-25.947	-18.233	-25.557	-18.317	-24.934	-18.399	-24.146	-18.571
-3.3	-27.361	-19.441	-26.972	-19.524	-26.346	-19.606	-25.558	-19.773
-3.4	-28.821	-20.680	-28.407	-20.770	-27.765	-20.860	-26.973	-21.042
-3.5	-30.326	-21.948	-29.903	-22.041	-29.242	-22.142	-28.408	-22.334
-3.6	-31.865	-23.272	-31.428	-23.370	-30.754	-23.466	-29.888	-23.645
-3.7	-33.483	-24.619	-33.019	-24.720	-32.306	-24.818	-31.414	-25.010
-3.8	-35.080	-25.991	-34.629	-26.094	-33.928	-26.198	-32.991	-26.404
-3.9	-36.717	-27.426	-36.228	-27.528	-35.494	-27.621	-34.581	-27.819
-4.0	-38.441	-28.875	-37.941	-28.979	-37.162	-29.076	-36.154	-29.285
-4.1	-40.194	-30.358	-39.694	-30.471	-38.894	-30.576	-37.860	-30.806
-4.2	-41.863	-31.927	-41.391	-32.031	-40.622	-32.133	-39.617	-32.355
-4.3	-43.606	-33.428	-43.103	-33.534	-42.328	-33.640	-41.294	-33.855
-4.4	-45.402	-34.995	-44.874	-35.111	-44.076	-35.221	-43.018	-35.448
-4.5	-47.270	-36.631	-46.729	-36.755	-45.864	-36.866	-44.798	-37.100
-4.6	-49.108	-38.311	-48.597	-38.434	-47.787	-38.545	-46.636	-38.773
-4.7	-50.978	-39.983	-50.416	-40.098	-49.582	-40.209	-48.530	-40.440
-4.8	-52.963	-41.640	-52.381	-41.764	-51.482	-41.883	-50.329	-42.116
-4.9	-54.952	-43.359	-54.394	-43.490	-53.483	-43.612	-52.275	-43.871
-5.0	-56.924	-45.145	-56.333	-45.275	-55.419	-45.409	-54.286	-45.666

Bibliography

Aït-Sahalia, Yacine, Jonathan A. Parker, and Motohiro Yogo, 2003, Luxury goods and the equity premium, *Journal of Finance*, forthcoming.

Amemiya, Takeshi, 1985, *Advanced Econometrics* (Harvard University Press: Cambridge, MA).

Anderson, T. W., and Herman Rubin, 1949, Estimation of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics* 20, 46–63.

Andrews, Donald W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.

Ang, Andrew, and Geert Bekaert, 2001, Stock return predictability: Is it there?, Working paper, Columbia Business School.

Attanasio, Orazio P., and Hamish Low, 2000, Estimating Euler equations, NBER Technical Working Paper No. 253.

Attanasio, Orazio P., and Guglielmo Weber, 1989, Intertemporal substitution, risk aversion and the Euler equation for consumption, *Economic Journal* 99, 59–73.

Billingsley, Patrick, 1999, *Convergence of Probability Measures* . Wiley Series in Probability and Statistics (John Wiley and Sons, Inc.: New York) second edn.

Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.

——— , Michael R. Gibbons, and Robert H. Litzenberger, 1989, Empirical test of the consumption-oriented CAPM, *Journal of Finance* 44, 231–262.

Breeden, Douglas T., and Robert H. Litzenberger, 1978, Prices of state-contingent claims implicit in option prices, *Journal of Business* 51, 621–651.

Caballero, Ricardo J., 1993, Durable goods: An explanation for their slow adjustment, *Journal of Political Economy* 101, 351–384.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

——— , 1996, Understanding risk and return, *Journal of Political Economy* 104, 298–345.

——— , 1998, Data appendix for 'asset prices, consumption, and the business cycle', Working paper, Harvard University.

——— , 2001, Why long horizons? A study of power against persistent alternatives, *Journal of Empirical Finance* 8, 459–491.

——— , 2003, Consumption-based asset pricing, in George M. Constantinides, Milton Harris, and René M. Stulz, ed.: *Handbook of the Economics of Finance*, vol. 1B . chap. 13, pp. 801–885 (Elsevier: Amsterdam).

——— , and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press: Princeton, NJ).

Campbell, John Y., and N. Gregory Mankiw, 1989, Consumption, income, and interest rates: Reinterpreting the time series evidence, in Olivier J. Blanchard, and Stanley Fischer, ed.: *NBER Macroeconomics Annual 1989* . pp. 185–216 (MIT Press: Cambridge, MA).

Campbell, John Y., and Robert J. Shiller, 1988a, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

——— , 1988b, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

Campbell, John Y., and Luis M. Viceira, 1999, Consumption and portfolio decisions when expected returns are time varying, *Quarterly Journal of Economics* 114, 433–495.

——— , 2002, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors* . Clarendon Lectures in Economics (Oxford University Press: New York).

Campbell, John Y., and Tuomo Vuolteenaho, 2002, Bad beta, good beta, Working paper, Harvard University.

Carroll, Christopher D., 2001, Death to the log-linearized consumption Euler equation! (and very poor health to the second-order approximation), *Advances in Macroeconomics* 1, article 6.

Cavanagh, Christopher L., Graham Elliott, and James H. Stock, 1995, Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131–1147.

Cochrane, John H., 1999, New facts in finance, NBER Working Paper No. 7169.

——— , 2001, *Asset Pricing* (Princeton University Press: Princeton, NJ).

Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho, 2003, The value spread, *Journal of Finance* 58, 609–641.

Cox, D. R., and D. V. Hinkley, 1974, *Theoretical Statistics* (Chapman and Hall: London).

Davis, James L., Eugene F. Fama, and Kenneth R. French, 2000, Characteristics, covariances, and average returns: 1929 to 1997, *Journal of Finance* 55, 389–406.

D'Avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.

den Haan, Wouter J., and Andrew T. Levin, 1997, A practitioner's guide to robust covariance matrix estimation, in G.S. Maddala, and C.R. Rao, ed.: *Handbook of Statistics*, vol. 15 . pp. 299–342 (Elsevier: Amsterdam).

——— , 2000, Robust covariance matrix estimation with data-dependent VAR prewhitening order, NBER Technical Working Paper No. 255.

Dickey, David A., and Wayne A. Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* 49, 1057–1072.

Dunn, Kenneth B., and Kenneth J. Singleton, 1986, Modeling the term structure of interest rates under non-separable utility and durability of goods, *Journal of Financial Economics* 17, 27–55.

Eichenbaum, Martin, and Lars Peter Hansen, 1990, Estimating models with intertemporal substitution using aggregate time series data, *Journal of Business and Economic Statistics* 8, 53–69.

Elliott, Graham, Thomas J. Rothenberg, and James H. Stock, 1996, Efficient tests for an autoregressive unit root, *Econometrica* 64, 813–836.

Elliott, Graham, and James H. Stock, 1994, Inference in time series regression when the order of integration of a regressor is unknown, *Econometric Theory* 10, 672–700.

——— , 2001, Confidence intervals for autoregressive coefficients near one, *Journal of Econometrics* 103, 155–181.

Epstein, Larry G., and Stanley E. Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–969.

——— , 1991, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis, *Journal of Political Economy* 99, 263–286.

Evans, G. B. A., and N. E. Savin, 1981, Testing for unit roots: 1, *Econometrica* 49, 753–779.

——— , 1984, Testing for unit roots: 2, *Econometrica* 52, 1241–1270.

Fama, Eugene F., 1991, Efficient capital markets: II, *Journal of Finance* 46, 1575–1617.

——— , and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–24.

——— , 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

——— , 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

——— , 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fuller, Wayne A., 1977, Some properties of a modification of the limited information estimator, *Econometrica* 45, 939–954.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Goyal, Amit, and Ivo Welch, 2003, Predicting the equity premium with dividend ratios, *Management Science* 49, 639–654.

Grossman, Sanford J., and Guy Laroque, 1990, Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods, *Econometrica* 58, 25–51.

Hall, Robert E., 1985, Real interest and consumption, NBER Working Paper No. 1694.

——— , 1988, Intertemporal substitution in consumption, *Journal of Political Economy* 96, 339–357.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

———, John Heaton, and Amir Yaron, 1996, Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics* 14, 262–280.

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269–1286.

———, 1983, Stochastic consumption, risk aversion, and the temporal behavior of asset returns, *Journal of Political Economy* 91, 249–265.

Harvey, Campbell R., 1989, Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics* 24, 289–317.

Hausman, Jerry A., Jinyong Hahn, and Guido Kuersteiner, 2001, Higher order MSE of jacknife 2SLS, Working paper, MIT.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.

Jansson, Michael, and Marcelo J. Moreira, 2003, Optimal inference in regression models with nearly integrated regressors, Working paper, Harvard University.

Kandel, Shmuel, and Robert F. Stambaugh, 1990, Expectations and volatility of consumption and asset returns, *Review of Financial Studies* 3, 207–232.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.

Kleibergen, Frank, 2002, Pivotal statistics for testing structural parameters in instrumental variables regression, *Econometrica* 70, 1781–1803.

Kocherlakota, Narayana R., 1990, On the 'discount' factor in growth economies, *Journal of Monetary Economics* 25, 43–47.

Kothari, S. P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169–203.

Lamont, Owen A., and Richard H. Thaler, 2003, Can the market add and subtract? Mispricing in tech stock carve-outs, *Journal of Political Economy* 111, 227–268.

Lanne, Markku, 2002, Testing the predictability of stock returns, *Review of Economics and Statistics* 84, 407–415.

Lehmann, E. L., 1986, *Testing Statistical Hypotheses* . Springer Texts in Statistics (Springer-Verlag: New York) second edn.

——— , 1999, *Elements of Large Sample Theory* . Springer Texts in Statistics (Springer-Verlag: New York) second edn.

Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Lewellen, Jonathan, 2003, Predicting returns with financial ratios, *Journal of Financial Economics*, forthcoming.

Lintner, John, 1965, Security prices, risk, and maximal gains from diversification, *Journal of Finance* 20, 587–615.

Lustig, Hanno, and Stijn Van Nieuwerburgh, 2002, Housing collateral, consumption insurance, and risk premia, Working paper, University of Chicago.

Macaulay, F. R., 1938, *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields, and Stock Prices in the United States Since 1856* (National Bureau of Economic Research: New York).

Mankiw, N. Gregory, and Matthew D. Shapiro, 1986a, Do we reject too often? Small sample properties of tests of rational expectations models, *Economics Letters* 20, 139–145.

——— , 1986b, Risk and return: Consumption beta versus market beta, *Review of Economics and Statistics* 68, 452–459.

Mehra, Rajnish, and Edward C. Prescott, 1985, The equity premium: A puzzle, *Journal of Monetary Economics* 15, 145–161.

Moreira, Marcelo J., 2001, Tests with correct size when instruments can be arbitrarily weak, Working paper, Harvard University.

——— , 2003, A conditional likelihood ratio test for structural models, *Econometrica* 71, 1027–1048.

Neely, Christopher J., Amlan Roy, and Charles H. Whiteman, 2001, Risk aversion versus intertemporal substitution: A case study of identification failure in the intertemporal consumption capital asset pricing model, *Journal of Business and Economic Statistics* 19, 395–403.

Nelson, Charles R., and Myung J. Kim, 1993, Predictable stock returns: The role of small sample bias, *Journal of Finance* 48, 641–661.

Nelson, Charles R., and Richard Startz, 1990, Some further results on the exact small sample properties of the instrumental variable estimator, *Econometrica* 58, 967–976.

Newey, Whitney K., and Daniel McFadden, 1994, Large sample estimation and hypothesis testing, in Robert F. Engle, and Daniel L. McFadden, ed.: *Handbook of Econometrics*, vol. 4 . chap. 36, pp. 2111–2245 (Elsevier: Amsterdam).

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Ogaki, Masao, and Carmen M. Reinhart, 1998, Measuring intertemporal substitution: The role of durable goods, *Journal of Political Economy* 106, 1078–1098.

Pakoš, Michal, 2003, Asset pricing with durable goods and non-homothetic preferences, Working paper, University of Chicago.

Parker, Jonathan A., and Christian Julliard, 2003, Consumption risk and cross-sectional returns, Working paper, Princeton University.

Phillips, Peter C. B., 1983, Exact small sample theory in the simultaneous equations model, in Zvi Griliches, and Michael D. Intriligator, ed.: *Handbook of Econometrics*, vol. 1 . chap. 8, pp. 449–516 (Elsevier: Amsterdam).

——— , 1987, Towards a unified asymptotic theory for autoregression, *Biometrika* 74, 535–547.

——— , and Pierre Perron, 1988, Testing for a unit root in time series regression, *Biometrika* 75, 335–346.

Piazessi, Monika, Martin Schneider, and Selale Tuzel, 2003, Housing, consumption, and asset pricing, Working paper, UCLA.

Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho, 2003, New forecasts of the equity premium, Working paper, Harvard University.

Richardson, Matthew, and James H. Stock, 1989, Drawing inferences from statistics based on multiyear asset returns, *Journal of Financial Economics* 25, 323–348.

Roll, Richard, 1977, A critique of the asset pricing theory's tests: Part I: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129–176.

Rothenberg, Thomas J., 1984, Approximating the distributions of econometric estimators and test statistics, in Zvi Griliches, and Michael D. Intriligator, ed.: *Handbook of Econometrics*, vol. 2 . chap. 15, pp. 881–935 (Elsevier: Amsterdam).

Rubinstein, Mark, 1976, The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics* 7, 407–425.

Selden, Larry, 1978, A new representation of preferences over "certain × uncertain" consumption pairs: The "ordinal certainty equivalent" hypothesis, *Econometrica* 46, 1045–1060.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Shiller, Robert J., 2000, *Irrational Exuberance* (Princeton University Press: Princeton, NJ).

Staiger, Douglas, and James H. Stock, 1997, Instrumental variables regression with weak instruments, *Econometrica* 65, 557–586.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Stock, James H., 1991, Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series, *Journal of Monetary Economics* 28, 435–459.

———, 1994, Unit roots, structural breaks and trends, in Robert F. Engle, and Daniel L. McFadden, ed.: *Handbook of Econometrics*, vol. 4 . chap. 46, pp. 2739–2841 (Elsevier Science Publishers: New York).

———, and Jonathan H. Wright, 2000, GMM with weak identification, *Econometrica* 68, 1055–1096.

———, and Motohiro Yogo, 2002, A survey of weak instruments and weak identification in generalized method of moments, *Journal of Business and Economic Statistics* 20, 518–529.

Stock, James H., and Motohiro Yogo, 2003, Testing for weak instruments in linear IV regression, Working paper, Harvard University.

Torous, Walter, Rossen Valkanov, and Shu Yan, 2001, On predicting stock returns with nearly integrated explanatory variables, *Journal of Business*, forthcoming.

Valkanov, Rossen, 2003, Long-horizon regressions: Theoretical results and applications, *Journal of Financial Economics* 68, 201–232.

Vissing-Jørgensen, Annette, 2002, Limited asset market participation and the elasticity of intertemporal substitution, *Journal of Political Economy* 110, 825–853.

Weil, Philippe, 1989, The equity premium puzzle and the risk-free rate puzzle, *Journal of Monetary Economics* 24, 401–421.

Whelan, Karl, 2000, A guide to the use of chain weighted aggregated NIPA data, Working paper, Federal Reserve Board.

Woodford, Michael D., 2003, *Interest and Prices: Foundations of a Theory of Monetary Policy* (Princeton University Press: Princeton).

Wooldridge, Jeffrey M., 1994, Estimation and inference for dependent processes, in Robert F. Engle, and Daniel L. McFadden, ed.: *Handbook of Econometrics*, vol. 4 . chap. 45, pp. 2639–2738 (Elsevier: Amsterdam).