ASSET PRICING

Doron Avramov

Professor of Finance

Fall 2011

Hebrew University of Jerusalem

Course Description

This course is intended for Ph.D as well as advanced master level students and it provides a comprehensive introduction to asset pricing and financial econometrics.

The following topics will be covered:

- Single and multi factor asset pricing models: theoretical aspects and empirical evidence.
- Time series market anomalies: the equity premium, risk free rate, and excess volatility puzzles.
- Cross sectional market anomalies: size, value, price momentum, earnings momentum, dispersion, credit risk, accrual, asset growth, capital investment, as well as total volatility and idiosyncratic volatility effects, among many others.
- Long horizon investment return.
- Beta pricing models versus the pricing kernel representation.
- Estimating and evaluating asset pricing models.
- Time-series versus cross-section methods in empirical asset pricing.
- Bayesian methods in finance.
- Understanding stock return predictability.
- Asset allocation when returns are IID and non IID.
- Investing in mutual funds and hedge funds.

Course materials

The Econometrics of Financial Markets, by John Y. Campbell, Andrew W. Lo, and A. Craig MacKinlay, Princeton University Press, 1997

Asset Pricing, by John H. Cochrane, Princeton University Press, 2005

Published and working papers in finance and economic journals as listed in the reference list

Course requirements

It is mandatory to attend all sessions. The course grade will be based on class participation (30%), paper presentations (30%), and written assignments (40%)

1 Asset pricing models

1.1 Overview

- We will address theoretical and empirical aspects of the time series and cross section properties of expected stock returns.
- Statistically, an expected asset return can be formulated as

$$\mathbb{E}(r_{i,t}) = \alpha_i + \beta_i'\mathbb{E}(f_t) \tag{1}$$

  where f denotes a set of K factor mimicking portfolios, β_i is a K vector of factor loadings, and α_i reflects the expected return component unexplained by factors, or model mispricing.
- For example, the market model is a statistical model with f being represented by the market portfolio.
- An asset pricing model intends to identify economic [(I)CAPM] or statistical [APT] common factors that *eliminate* α.
- In the absence of alpha, expected return differential across assets is triggered by factor loadings only.
- The presence of asset mispricing could give rise to additional cross sectional differences in expected returns.
- Empirical evidence shows that
  - α is there based on various factor specifications;
  - α varies cross sectionally with firm level variables such as size, the book-to-market ratio, turnover, and past returns;
  - α varies over time with both business conditions (proxied by variables such as the default spread and the term spread) and firm level variables.

1.2 CAPM

- The CAPM of Sharpe (1964) and Lintner (1965) has originated the literature on asset pricing models.
- The CAPM is a general equilibrium model in a single-period economy.
- The model imposes an economic restriction on the statistical structure in (1)
- The unconditional version of the CAPM says that

$$\begin{aligned} \mathbb{E}(r_{i,t}) &= \text{cov}(r_{i,t}, r_{m,t}) \frac{\mathbb{E}(r_{m,t})}{\text{var}(r_{m,t})}, && (2) \\ &= \beta_{i,m} \mathbb{E}(r_{m,t}), && (3) \end{aligned}$$

  where $r_{m,t}$ is excess return on the market portfolio at time t.
- Later, we will cover conditional versions of the CAPM wherein either mean returns or the co-variances or both are time varying.
- The CAPM measures an asset risk as the covariance of the asset return with the market portfolio return.
- The higher the covariance the less desirable the asset is; thus the asset price is lower which amounts to higher expected return.
- The market price of risk, common to all assets, is determined in equilibrium by the risk aversion of investors.

- For practical implications such as
  - estimating the cost of equity capital
  - testing the CAPM
  - evaluating performance of active management such as mutual funds, hedge funds, and investment newsletters

  the market portfolio is commonly proxied by the value weighted NYSE portfolio, S&P500 index, or CRSP index (AMEX, NYSE, and NASDAQ combined).
- Overall, the CAPM is nice, simple, and intuitive.
- But there are just too many empirical and theoretical drawbacks.

1.2.1 Empirical drawbacks

Below there is a comprehensive list of documented market anomalies in US financial markets.

a. The CAPM is at odds with anomalous patterns in the cross section of stock returns - both corporate and asset pricing anomalies.

- Prominent asset pricing anomalies include the size, book-to-market, past return (short and long run reversal and intermediate term momentum), earnings momentum, dispersion, accruals, credit risk (level and changes), and total volatility and idiosyncratic volatility effects.
- There are many other documented anomalies, some of which are described below.

- The size and value effects
  - Basu (1977), Banz (1981), and Fama and French (FF) (1992) suggest that the CAPM cannot explain the effect of size, the dividend yield, the earnings yield, and the book-to-market on average returns.
    * Size effect: higher average returns on small stocks than large stocks. Beta cannot explain the difference.
    * Value effect: higher average returns on value stocks than growth stocks. Beta cannot explain the difference.

      Value firms have high E/P, B/P, D/P, or CF/P. Why value? Because physical assets can be purchased at low prices.

      Growth firms exhibit low ratios. Why growth? Because high price relative to fundamentals reflects capitalized growth opportunities.
- Past return anomalies:
  - Lehmann (1990) and Jegadeesh (1990) show that contrarian strategies that exploit the short-run return reversals in individual stocks generate abnormal returns of about 1.7% per week and 2.5% per month, respectively.
  - Jegadeesh and Titman (1993) and a great body of subsequent work uncover abnormal returns to momentum-based strategies focusing on investment horizons of 3, 6, 9, and 12 months. (Later we dedicate a special section for momentum.)
  - DeBondt and Thaler (1985, 1987) document long run reversals

- Accounting based anomalies:
  - Ball and Brown (1968) document the post-earnings-announcement drift, also known as earnings momentum. This anomaly refers to the fact that firms reporting unexpectedly high earnings subsequently outperform firms reporting unexpectedly low earnings. The superior performance lasts for about nine months after the earnings announcements.
  - Sloan (1996) shows that firms with relatively high (low) levels of accruals to total assets experience negative (positive) future abnormal returns. In particular, taking long (short) position in firms with the lowest (highest) accruals yields about 10% per year.
  - Diether, Malloy, and Scherbina (2002) suggest that firms with high dispersion in analysts' earnings forecasts earn less than firms with low dispersion.
- There are anomalous patterns related to credit conditions:
  - Level: Dichev (1998), Campbell, Hilscher, and Szilagyi (2008), and Avramov, Chordia, Jostova, and Philipov (2011) demonstrate a *negative* cross-sectional correlation between credit risk and future stock returns.
  - Changes: Dichev and Piotroski (2001) document abnormal stock price declines following credit rating downgrades. Interestingly, stock prices do not advance following rating upgrades.
- Asset Growth
  - Cooper, Gulen, and Schill (2008) find companies that grow their total asset more earn lower subsequent returns.

- They suggest that this phenomenon is due to investor initial overreaction to changes in future business prospects implied by asset expansions.
- Asset growth can be measured as the annual percentage change in total assets based on the most recent balance sheet including quarterly financial statements.

- Capital Investment
  - Titman, Wei, and Xie (2004) document a negative relation between capital investments and returns.
  - Correspondingly, an investment based factor had been used by Chen, Novy-Marx, and Zhang (2010) in their new three factor model.
  - Capital investment to assets is the annual change in gross property, plant, and equipment plus the annual change in inventories divided by lagged book value of assets.
  - Changes in property, plants, and equipment capture capital investment in long-lived assts used in operations many years such as buildings, machinery, furniture, and other equipment.
  - Changes in inventories capture working capital investment in short-lived assets used in a normal business cycle.

- Return on Assets (ROA):
  - Fama and French (2006) find that more profitable firms have higher expected returns than less profitable firms.
  - Chen, Novy-Marx, and Zhang (2010) show that firms with higher past return-on-assets earn abnormally higher subsequent returns.

- Wang and Yu (2010) find that the anomaly exists primarily among firms with high arbitrage costs and high information uncertainty, suggesting that mispricing is a culprit.
- ROA is typically measured as income before extraordinary items divided by one quarter lagged total assets.

- Gross Profitability Premium: Novy-Marx (2010) discovers that sorting on gross-profit-to-assets creates abnormal benchmark-adjusted returns, with more profitable firms having higher returns than less profitable ones.

- Volatility effects:
  - Ang, Hodrik, Xing, and Zhang (2006, 2009) show negative cross section relation between idiosyncratic volatility and average return in both US and global markets.
  - Also total volatility is negatively related to future returns.

- Counter intuitive relations: the dispersion, credit risk, and IV effects apparently violate the risk-return tradeoff.

  However, in a theoretical work Avramov, Cederburg, and Hore (2011) show that the dispersion, credit risk, and volatility effects are all related and they are perfectly consistent with rational asset pricing.

- The turnover effect.
  - Higher turnover is followed by lower future return. See, for example, Avramov and Chordia (RFS 2006).
  - It should be noted that Lee and Swaminathan (2000) find that high turnover stocks exhibit features of high growth stocks.

- Turnover can be constructed using various methods. Here are a few methods. For any trading day within a particular month, compute the daily volume in either $ or the number of traded stocks or the number of transactions. Then divide the volume by the market capitalization or by the number of outstanding stocks. Finally, use the daily average, within a trading month, of the volume/market capitalization ratio as the monthly turnover.

• F-Score

  - The extensively used F-Score of Piotroski (2000) also indicates the CAPM violation.
  - The FSOCRE is designed to identify firms with the strongest improvement in their overall financial conditions while meeting a minimum level of financial performance.
  - High F-score firms demonstrate distinct improvement along a variety of financial dimensions, while low FSCORE firms exhibit poor fundamentals along these same dimensions.
  - The FSCORE is computed as the sum of nine components which are either zero or one as explained in Piotroski (2000).
  - The overall FSCORE thus ranges between zero and nine, where a low (high) score represents a firm with very few (many) good signals about its financial conditions.

• Economic links and predictable returns

  - Cohen and Frazzini (2008) show that stocks do not promptly incorporate news about economically related firms.

- A long short strategy that capitalizes on this effect generates abnormal return of about 1.5% per month.
- This phenomenon establishes evidence against market efficiency in general and the CAPM in particular.

- Are anomalies pervasive?
  - Lo and MacKinlay (1990) claim that the size effect may very well be the result of unconscious, exhaustive search for a portfolio formation creating with the aim of rejecting the CAPM.
  - Lettau and Ludvigson (1999) suggest that the size and value effects are consistent with a conditional version of the (C)CAPM with the consumption to wealth ratio (cay) being the conditioning variable.
  - Avramov, Chordia, and Goyal (2006) show that implementing short term reversal strategies yields profits that do not survive transactions costs.
  - Avramov, Chordia, Jostova, and Philipov (2007a,b) (2011) show that the momentum, dispersion, and credit risk effects concentrate in a very small portion of high credit risk high illiquid stocks.
  - Others claim that momentum is robust (more later). Even then, it should be noted that over the last decade momentum seems to loose its momentum.
  - The accruals anomaly seems to be pervasive.
  - Two very relevant papers which analyze asset pricing anomalies in a unified framework are Fama and French (2008) and Avramov, Chordia, Jostova, and Philipov (2011).

b. Multifactor extensions of the CAPM such as Chen, Roll, and Ross (1986), Fama and French (1993), and Pastor and Stambaugh (2003) associate expected returns with a tendency to move with multiple risk factors.

   – On one hand, multifactor models dominate the explanation of the cross section dispersion in expected returns.
   – At the same time, however,
     * Multifactor models need not perform better out-of-sample.
     * The economic appeal of the additional factors is often questionable.

c. Mutual funds and hedge funds seem to outperform benchmark indexes, even after controlling for market risk. However, multifactor extensions explain most fund persistence (Carhart (1997) focuses on mutual funds and Fung and Hsieh (2001, 2004) on hedge funds).

d. Roll (1977) claims that the CAPM is untestable since the market return is unobservable and it cannot be measured accurately. Using proxy for the market factor could deliver unreliable inference.

1.2.2 Theoretical drawbacks of the CAPM

a. The CAPM assumes that the average investor cares only about the performance of his investment portfolio.

   – But eventual wealth could come from both investment, labor, and entrepreneurial incomes.

- An additional factor is therefore needed.
- The CAPM says that two stocks that are equally sensitive to market movements must have the same expected return.
- But if one stock performs better in recessions it would be more desirable for most investors who may actually lose their jobs or get lower salaries in recessions.
- The investors will therefore bid up the price of that stock, thereby lowering expected return.
- Thus, procyclical stocks should offer higher average returns than countercyclical stocks, even if both stocks have the same market beta.
- Put another way, covariation with recessions seems to matter in determining expected returns.
- You may correctly argue that the market tends to go down in recessions.
- But recessions tend to be unusually severe or mild for a given level of market returns.

b. The CAPM applies to single-period — myopic — investors. Merton (1973) introduces a multi-period version of the CAPM - that is the intertemporal CAPM (ICAPM).

- In the ICAPM setup, the demand for risky assets is attributed not only to the mean variance component, as in the CAPM, but also to hedging against unfavorable shifts in the investment opportunity set. The hedging demand is something extra relative to the CAPM.
- Merton points out that asset's risk should be measured by its covariance with the marginal utility of investors,

which need not be the same as the covariance with the market return.

– Merton shows that multiple state variables that are sources of priced risk are required to explain the cross section variation in expected returns.

– In such intertemporal models, equilibrium expected returns on risky assets may differ from the riskless rate even when they have no systematic (market) risk.

– But Merton does not say which other state variables are priced - this gives license to fish factors that work well in explaining the data but at the same time they void any economic content.

– Most prominent in the class of intertemporal models is the consumption CAPM of Rubinstein (1976), Lucas (1978), Breeden (1979), and Grossman and Shiller (1981).

– The greatness of the CCAPM is that it is to identify an economically based asset pricing factor.

– Indeed, the CCAPM is a single state variable model; real consumption growth is the single factor. See details below.

1.3 The Consumption CAPM — CCAPM

- The CCAPM is a theoretically motivated model. One of the few cases wherein theory invokes empirical studies - not the other way around.

- The CCAPM replaces the market factor by consumption growth.

- Let us derive a simple discrete time version of the CCAPM.

- The derivation is based on the textbook treatment of Cochrane (2007).
- We use the power utility to describe investor preferences:

$$u(c_t) = \frac{1}{1-\gamma} c_t^{1-\gamma} \text{ for } \gamma \neq 1, \tag{4}$$

$$u(c_t) = \log(c_t) \text{ for } \gamma = 1. \tag{5}$$

- Notation:
  - c_t denotes consumption at time t
  - γ is the relative risk aversion parameter.
- The investor must decide how to allocate her wealth between consumption and saving.
- The investor can freely buy or sell any amount of a security whose current price is p_t and next-period payoff is x_{t+1} ($x_{t+1} = p_{t+1} + d_{t+1}$).
- How much will she buy or sell?
- To find the answer, consider (wlog) a two-period investor whose income at time s is e_s, and let y be the amount of security she chooses to buy.
- The investor's problem is

$$\max_y u(c_t) + \mathbb{E}_t [\rho u(c_{t+1})] \quad s.t. \tag{6}$$

$$c_t = e_t - p_t y, \tag{7}$$

$$c_{t+1} = e_{t+1} + x_{t+1} y, \tag{8}$$

where ρ denotes the impatience parameter, also called the subjective discount factor.

- Substituting the constraints into the objective, and setting the derivative with respect to y equal to zero, we obtain the first order condition for an optimal consumption and portfolio choice,

$$p_t u'(c_t) = \mathbb{E}_t \left[\rho u'(c_{t+1}) x_{t+1} \right]. \tag{9}$$

- The left hand side in (9) reflects the loss in utility from consuming less as the investor buys an additional unit of the asset.

- The right hand side describes the expected increase in utility obtained from the extra payoff at $t+1$ attributed to this additional unit of the asset.

- A well-know representation for the first order condition is obtained by dividing both sides of (9) by $p_t u'(c_t)$,

$$1 = \mathbb{E}_t \left(\xi_{t+1} R_{t+1} \right), \tag{10}$$

where $\xi_{t+1} = \frac{\rho u'(c_{t+1})}{u'(c_t)}$ stands for the pricing kernel, also known as the marginal rate of substitution or the stochastic discount factor, and $R_{t+1} = \frac{x_{t+1}}{p_t}$ denotes the gross return on the security.

- The relation in equation (10) is the fundamental discount factor view of asset pricing theories.

1.3.1 Several Insights about the First Order Condition

- Observe from (10) that the gross riskfree rate, the rate known at time t, which is uncorrelated with the discount factor, is given by $R^f_{t,t+1} = 1/\mathbb{E}_t(\xi_{t+1})$.

- When investor preferences are described by the power utility function, as in equation (4), the pricing kernel takes the form $\xi_{t+1} = \rho(c_{t+1}/c_t)^{-\gamma}$.

- Assuming lognormal consumption growth one can show that the continuously compounded riskfree rate is

$$\begin{aligned} r^f_{t,t+1} &= -\ln(\rho) - \ln\mathbb{E}_t \left[\exp(-\gamma\Delta \ln c_{t+1})\right], \\ &= -\ln(\rho) + \gamma\mathbb{E}_t(\Delta\ln c_{t+1}) - \frac{\gamma^2}{2}\sigma^2_t(\Delta\ln c_{t+1}). \quad (11) \end{aligned}$$

- To derive (11) we have used the useful relation that if x is normally distributed then

$$\mathbb{E}(e^{ax}) = e^{\mathbb{E}(ax)} e^{\frac{a^2}{2}\sigma^2(x)}. \quad (12)$$

- We can see from (11) that the EIS is $1/\gamma$ which creates some problems, as discussed below.

- A risky asset is one having nonzero correlation with the discount factor.

- Then, we obtain a *beta pricing model*

$$\mathbb{E}_t(r_{i,t+1}) = r^f_{t,t+1} + \underbrace{\left(\frac{\text{cov}_t(r_{i,t+1}, \xi_{t+1})}{\text{var}_t(\xi_{t+1})}\right)\left(-\frac{\text{var}_t(\xi_{t+1})}{\mathbb{E}_t(\xi_{t+1})}\right)}_{\textit{risk adjustment}}. \quad (13)$$

- In words, expected excess return on each security, stock, bond, or option, should be proportional to the coefficient in the regression of that return on the discount factor.

- The constant of proportionality, common to all assets, is the risk premium.

- Focusing on the power utility function and using a first order Taylor series expansion, we obtain

$$\mathbb{E}(r_{i,t+1}) \approx r^f + \beta_{i,\Delta c}\lambda_{\Delta c}, \tag{14}$$

where

$$\beta_{i,\Delta c} = \frac{\text{cov}(r_{i,t+1}, \Delta c)}{\text{var}(\Delta c)}, \tag{15}$$

$$\lambda_{\Delta c} = \gamma \text{var}(\Delta c). \tag{16}$$

- This is the discrete time version of the consumption CAPM.
- The relation is exact in continuos time.
- The asset's risk is defined as the covariance between the asset return and consumption growth.
- The risk premium is obtained as the product of the investor risk aversion and the volatility of consumption growth.
- Notice from equation (14) that the expected return on an asset is larger the larger the covariance between the return on that asset with consumption growth.
- Intuition: an asset doing badly in recessions (positive covariance) when the investor consumes little, is less desirable than an asset doing badly in expansions (negative covariance) when the investor feels wealthy and is consuming a great deal.

  The former asset will be sold for a lower price, thereby having a higher expected return.

1.3.2 Theoretically, the CCAPM appears preferable to the traditional CAPM

- It takes into account the dynamic nature of portfolio decisions
- It integrates the many forms of wealth beyond financial asset wealth
- Consumption should deliver the purest measure of good and bad times as investors consume less when their income prospects are low or if they think future returns will be bad.

Empirically, however, The CCAPM has disappointed as described below.

1.3.3 CCAPM: The Empirical Evidence

- Hansen and Singleton (1982, 1983) formulate a consumption-based model in which a representative agent has time-separable power utility of consumption.
- They reject the model on U.S. data, finding that it cannot simultaneously explain the time-variation of interest rates and the cross-sectional pattern of average returns on stocks and bonds.
- Wheatley (1988) rejects the model based on international data.
- Mankiw and Shapiro (1986) show that the CCAPM performs no better, and in many respects even worse, than the CAPM.
- They regress the average returns of the 464 NYSE stocks that were continuously traded from 1959 to 1982 on their market betas, on consumption growth betas, and on both betas.

- They find that the market betas are more strongly and robustly associated with the cross section of average returns, and that market beta drives out consumption beta in multiple regressions.
- Breeden, Gibbons, and Litzenberger (1989) find comparable performance of the CAPM and a model that uses a *mimicking portfolio* for consumption growth as the single factor.
- Cochrane (1996) finds that the static CAPM substantially outperforms the consumption CAPM in pricing size portfolios.
- Academics and practitioners typically use portfolio-based models, such as the CAPM or the Fama and French extension to correct for risk.
- So the CCAPM has not been a great success.
- Recently, Lettau and Ludvigson (2001) and Amir and Bansal (2004) suggest that *Consumption Strikes Back!*.
- Indeed, Parker and Julliard (2005), Bansal, Dittmar, and Lundblad (2005), Hansen, Heaton, and Li (2008), and Avramov, Cederburg, and Hore (2011) suggest that if systematic consumption risk is measured over long horizons the CCAPM is able to explain the cross section dispersion in average stock returns.

1.4 The Consumption CAPM and the Equity Premium Puzzle

- From a cross section perspective, the CCAPM fails if consumption beta is unable to explain why average returns differ across stocks, which is indeed the case.

- At the aggregate level (time-series perspective) the CCAPM leads to the so-called *equity premium puzzle* documented by Mehra and Prescott (1985) as well as the *riskfree puzzle*.

- To illustrate, let us manipulate the first order condition (10) (for notational clarity I suppress the time dependence)

$$\begin{aligned} 1 &= E(\xi R), && (17)\\ &= E(\xi)E(R) + cov(\xi, R), && (18)\\ &= E(\xi)E(R) + \rho_{\xi,R}\sigma(\xi)\sigma(R). && (19) \end{aligned}$$

- Dividing both sides of (19) by $E(\xi)\sigma(R)$ leads to

$$\frac{E(R) - R^f}{\sigma(R)} = -\rho_{\xi,R}\frac{\sigma(\xi)}{E(\xi)}), \qquad (20)$$

which implies that

$$\left|\frac{E(R) - R^f}{\sigma(R)}\right| \leq \frac{\sigma(\xi)}{E(\xi)} \left(= \sigma(\xi)R^f\right). \qquad (21)$$

- The left hand side in (21) is known as the Sharpe ratio.

- The highest Sharpe ratio is associated with portfolios lying on the mean-variance efficient frontier.

- Notice that the slope of the frontier is governed by the volatility of the discount factor.

- Under the CCAPM it follows that

$$\left|\frac{E(R^{mv}) - R^f}{\sigma(R^{mv})}\right| = \frac{\sigma\left[(c_{t+1}/c_t)^{-\gamma}\right]}{E\left[(c_{t+1}/c_t)^{-\gamma}\right]}. \qquad (22)$$

- When log consumption growth is normally distributed the right hand side of (22) can be shown to be equal to

$$\sqrt{e^{\gamma^2 \sigma^2 (\Delta \ln c_{t+1})}}, \tag{23}$$

which can be approximated by

$$\gamma \sigma(\Delta \ln c). \tag{24}$$

- In words, the slope of the mean-variance efficient frontier is higher if the economy is riskier, i.e., if consumption growth is more volatile, or if investors are more risk averse.
- Over the past several decades in the US, real stock returns have averaged 9% with a std. of about 20%, while the real return on T.bills has been about 1%.
- Thus, the historical annual market Sharpe ratio has been about 0.4.
- Moreover, aggregate nondurable and services consumption growth had a std. of 1%.
- This fact can only be reconciled with $\gamma = 50$.
- But the empirical estimates are between 2 and 10.
- This is the "equity premium puzzle." The historical Sharpe ratio is simply too large than the one obtained with reasonable risk aversion and consumption volatility estimates.

1.5 The riskfree rate puzzle

- Using the standard CCAPM framework also gives rise to the riskfree rate puzzle.

- Recall, we have shown that

$$r^f_{t,t+1} = -\ln(\rho) + \gamma\mathbb{E}_t(\Delta \ln c_{t+1}) - \frac{\gamma^2}{2}\sigma^2_t(\Delta \ln c_{t+1}). \quad (25)$$

- Using $\gamma = 2$ the riskfree rate should be around 5% to 6% per year.
- The actually observed rate is less than 1%.

1.6 How shall we resolve the equity premium and riskfree puzzles?

- Perhaps investors are much more risk averse than we may have thought.
  - This indeed resolves the equity premium puzzle.
  - But higher risk aversion parameter implies higher riskfree rate. So higher risk aversion reinforces the riskfree puzzle.
- Perhaps the stock returns over the last 50 years are good luck rather than an equilibrium compensation for risk.
- If so the equity premium will disappear in several decades as more reasonable returns are realized.
- Perhaps something is deeply wrong with the utility function specification and/or the use of aggregate consumption data.
  - Indeed, the CCAPM assumes that agents' preferences are time additive VNM representation (e.g., power).
  - Standard power utility preferences impose tight restrictions on the relation between the equity premium and the risk free rate.

- As shown earlier, EIS and the relative risk aversion parameter are reciprocals of each other.
- Economically they should not be tightly related.
- EIS is about deterministic consumption paths - it measures the willingness to exchange consumption today with consumption tomorrow for a given riskfree rate, whereas risk aversion is about preferences over random variables.
- Epstein and Zin (1989) and Weil (1990) entertain recursive intertemporal utility functions that separate risk aversion from elasticity of intertemporal substitution, thereby separating the equity premium from the riskfree rate.
- Duffie and Epstein (1992) introduces the Stochastic Differential Utility which is the continuous time version of Epstein-Zin-Weil.
- They show that under certain parameter restrictions, the riskfree rate actually diminishes with higher risk aversion.
- Empirically, recursive preferences perform better in matching the data.

• Reitz (1988) comes up with an interesting idea: he brings the possibility of low probability states of economic disaster and is able to explain the observed equity premium.

• Barro (2005) supports the Reitz's perspective.

• Weitzman (2007) proposes an elegant solution using a Bayesian framework to characterize the ex ante uncertainty about consumption growth.

• The asset pricing literature typically assumes that the growth rate is normally distributed.

$$g \sim N(\mu_g, \sigma_g^2). \tag{26}$$

- The literature also assumes that μ_g and σ_g are known to the agents in the economy.
- What if you assume that μ_g is known and σ_g is unknown.
- Moreover, you model σ_g as inverted gamma distributed random variable.
- Then g has the Student-t distribution.
- The student t distribution captures both the high equity premium and low riskfree rate.
- It should also be noted that there is some recent literature proposing alternative measures of consumption which purportedly help with measurement errors in consumption growth and work better empirically - see e.g., Jagannathan and Wang (2007), Savov (2011), and Da and Yun (2010).

1.7 Long Run Risk

- The LRR of Bansal and Yaron (2004) has also been a success.
- LRR models feature a small by highly persistent component in consumption that is hard to capture directly using consumption data.
- Never-the-less that small component is important for asset pricing.

- The persistent component is modeled either as stationary (Bansal and Yaron (2004)) or as co-integrated (Bansal, Dittmar, and Kiku (2009)) stochastic process.
- The model has been found useful in explaining the equity premium puzzle, size and book to market effects, momentum, long term reversals, risk premiums in bond markets, real exchange rate movements, among others (see a review paper by Bansal (2007)).
- However, all the evidence is based on calibration experiments or in-sample data fitting.
- Ferson, Nallareddy, and Xie (2011) examine the out of sample performance of the LRR paradigm.
- They examine stationary and co-integrated versions of the model.
- They find that the model performs comparably overall to the simple CAPM as well as a co-integrated version outperforms the stationary version.

1.8 Multifactor Models

- The poor empirical performance of the single factor (C)CAPM motivates a search for alternative asset pricing models.
- Either include more factors or go along with conditional models or both.
- Multiple factors have been inspired along the spirit of
  - The Arbitrage Pricing Theory — APT — (1976) of Ross

- The intertemporal CAPM — ICAPM — of Merton (1973).

- Distinguishing between the APT and ICAPM is often confusing.
- Cochrane (2001) argues that the biggest difference between them for empirical work is in the inspiration of factors:
  - The APT suggests a statistical analysis of the covariance matrix of returns to find factors that characterize common movements.
  - The ICAPM puts some economic meaning to the selected factors.
- For instance, Roll and Ross (1980) propose factor analysis of the covariance matrix of returns, and Connor and Korajczyk (1988) propose an asymptotic principle components.
- Later we will study a nice Bayesian approach for extracting common factors using a panel of stock returns.
- These statistically motivated approaches to extracting factors remain silent about the *economic* forces that determine factor risk prices.

1.8.1 The Fama-French Model

- The three-factor FF model is attractive from an empirical perspective.
- FF (1992, 1993) have shown that the cross-sectional variation in expected returns can be captured using the following factors:

1. the return on the market portfolio in excess of the risk free rate of return
2. a zero net investment (spread) portfolio that is long in small firm stocks and short in large firm stocks (SMB)
3. a zero net investment (spread) portfolio that is long in high book-to-market stocks and short in low book-to-market stocks (HML)

- The factors as well as description about constructions are available at Kenneth French's web site.
- FF (1996) have shown that their model is able to explain many of the market anomalies known back then - excluding the momentum effect.
- SMB and HML are based on six size book-to-market sorted portfolios.
- The six portfolios are constructed at the end of each June as the intersections of 2 portfolios formed on market equity and 3 portfolios formed on the ratio of book equity to market equity.
  - The size breakpoint for year t is the median NYSE market equity at the end of June of year t.
  - BE/ME for June of year t is the book equity for the last fiscal year end in t-1 divided by ME for December of t-1. The BE/ME breakpoints are the 30th and 70th NYSE percentiles.
  - SMB is the average return on the three small portfolios minus the average return on the three big portfolios.

- HML is the average return on the two value portfolios minus the average return on the two growth portfolios.

- FF (1993) argue that their factors are state variables in an ICAPM sense.
- Liew and Vassalou (2000) make a good case for the FF assertion.
- They find that the FF factors forecast GDP growth.
- But in the end, the Fama-French findings are empirically based, lack any theoretical underpinning.
- The model does poorly out of sample.
- And it does not explain many market anomalies, noted earlier, that had been discovered more recently.
- A potentially better pricing model?
  - A recent paper by Chen, Novy-Marx, and Zhang (2010) titled An "Alternative Three-Factor Models proposes a new three factor model."
  - The three factors are (i) the market portfolio, (ii) an investment factor, and (iii) a return on asset factor.
  - That model is claimed to explain short run reversal, earnings momentum, accrual, and stock valuation ratios (e.g., the book to market effect), among several other anomalies.
  - Still, it does not explain all anomalies and, moreover, the statistical tests are based on equity portfolios not individual securities.

  - Indeed, if you implement that factor model using individual stocks you will see that its performance is rather poor.
  - In particular, you will see that for all anomalies that model improves, relative to the CAPM, only for the profitability based anomaly (which is one of their factor).

1.8.2 Other Multifactor Models

- Carhart (1997) proposes a four-factor model to evaluate performance of equity mutual funds — MKT, SMB, HML, and WML, where WML is a momentum factor.
- He shows that profitability of "hot hands" based trading strategies (documented by Hendricks, Patel, and Zeckhauser (1993)) disappears when investment payoffs are adjusted by WML.
- The profitability of "smart money" based trading strategies in mutual funds (documented by Zheng (1999)) also disappears in the presence of WML.
- Pastor and Stambaugh (2003) propose adding a liquidity factor.
- So we have five major factors to explain equity returns
  1. market
  2. SMB
  3. HML
  4. WML
  5. Liquidity

- Often bond portfolios serve as common factors (see Ferson and Harvey (1999) and the references therein).
- The collection of factors explaining HF performance is different (see Avramov, Kosowski, Naik, and Teo (2009) and the references therein) to account for non linear payoffs generated by hedge funds.

1.9 Do multifactor models really perform well?

- The sad news is that SMB, HML, WML, and a liquidity factor do not really capture the size, book-to-market, past return, and liquidity effects in individual stock returns.
- For example, Avrmaov and Chordia (2006a) examine pricing abilities of conditional and unconditional models.
- They first consider a time series specification of the form

$$r_{it} = \alpha_i + \beta_i(z_{t-1}, z_{it-1})' f_t + \epsilon_{it} \tag{27}$$

  for any security in the sample, where z_t is a macro level conditioning information and z_{it} denotes firm level conditioning information such as size and book to market.
- Note that the risk adjusted return is obtained as the sum of intercept and residual

$$r^*_{i,t} = \alpha_i + \epsilon_{it} \tag{28}$$

- They then consider the cross section regression of risk adjusted returns on stock level size, book to market, past return variables, and turnover.

- If indeed the pricing model does a good job in pricing individual stocks then the slope coefficients in the cross section regressions should be statistically indistinguishable from zero.
- The slope coefficients, however, are highly significant.
- That is, firm level size, book to market, past return, and turnover are still important even when stock returns are adjusted by the SMB, HML, WML, and liquidity factors.
- We will revisit the Avramov-Chordia methodology in the section dealing with asset pricing test methodologies.
- Ferson and Harvey (1999) show that alpha in multi factor models varies with business conditions.
- Whereas the Ferson and Harvey (1999) metric is statistical, Avramov and Chordia (2006b) demonstrate that mispricing variation with macro variables is economically significant.

1.10 Quick summary

- So the empirical evidence on model pricing abilities is disappointing whether you take single or multi factor models whether you take conditional or unconditional models.
- It is still nice to develop asset pricing theories and test methodologies.
- Special focus should be paid to testing models among individual stocks rather than industry or characteristics sorted portfolios.

- Moreover, we will entertain some Bayesian methods that take the view that asset pricing models, albeit non perfect, are still useful for estimating cost of equity estimation, evaluating fund performance, and selecting optimal portfolios.
- But before we move on we dedicate a complete section to momentum which is one of the most intriguing market anomalies.
- Momentum was well known in Wall Street before the 90s but Jegadeesh and Titman (1993) were the first to address momentum in an academic paper.

2 Momentum

- Jegadeesh and Titman (1993) document profitability of momentum strategies.
- Buying stocks that performed relatively well in the past and selling stocks that performed relatively poorly in the past generate significant positive returns.
- Formation and holding periods are 3 to 12 months.

2.1 Momentum robustness

- Stock return momentum is robust
  - Fama and French (1996) show that momentum profitability is the only CAPM-related anomaly unexplained by the Fama and French (1993) three-factor model.

– In fact, regressing gross momentum payoffs on the Fama-French factors tend to strengthen, rather than explain, momentum profitability.

– Schwert (2003) demonstrates that the size and value effects in the cross section of returns, as well as the ability of the aggregate dividend yield to forecast the equity premium disappear, reverse, or attenuate following their discovery.

– Momentum is an exception. Jegadeesh and Titman (2001, 2002) document the profitability of momentum strategies in the out of sample period after its initial discovery.

– Haugen and Baker (1996) and Rouwenhorst (1998) document momentum in international markets. Not really an out of sample evidence but still interesting.

– Korajczyk and Sadka (2004) find that momentum survives trading costs, whereas Avramov, Chordia, and Goyal (2006) show that the profitability of the other past-return anomaly, namely reversal, disappears in the presence of trading costs.

– Fama and French (2008) show that momentum is among the few robust anomalies.

– Momentum Spillover

  ∗ Gebhardt, Hvidkjaer, and Swaminathan examine the interaction between momentum in the returns of equities and corporate bonds.

  ∗ They find significant evidence of a momentum spillover from equities to investment grade corporate bonds of the same firm.

    * In particular, firms earning high (low) equity returns over the previous year earn high (low) bond returns the following year.
    * The spillover results are stronger among firms with lower-grade debt and higher equity trading volume.
  - Bond Price Momentum
    * Momentum is also prominent among corporate bonds.
    * Jostova et al find significant price momentum in US corporate bonds over the 1973 to 2008 period.
    * Bond momentum profits are significant in the second half of the sample period, 1991 to 2008, and amount to 64 basis points per month.
    * Momentum strategies are only profitable among non-investment grade bonds, where they yield 190 basis points per month.

- The prominence of the momentum profitability has generated both behavioral and rational theories.
  - Behaviroal: Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), and Hon, Lim, and Stein (2000).
  - Rational: Berk, Green, and Naik (1999) and Johnson (2002).

2.2 Momentum interactions

Momentum interactions have been documented at the stock, industry, and aggregate level.

- Stock level interactions:
  - Hon, Lim, and Stein (2000) show that momentum profitability concentrates in small stocks.
  - Lee and Swaminathan (2000) show that momentum payoffs increase with trading volume.
  - Zhang (2006) finds that momentum concentrates in high information uncertainty stocks (stocks with high return volatility, cash flow volatility, or analysts? forecast dispersion) and provides behavioral interpretation.
  - Avramov, Chordia, Jostova, and Philipov (2007) document that momentum concentrates in low rated stocks. Moreover, the credit risk effect seems to dominate the information uncertainty and size effects.
- Potential industry level interactions:
  - Moskowitz and Grinblatt (1999) show that industry momentum subsumes stock level momentum.
  - Grundy and Martin (2001) find no industry effects in momentum.
- Interactions at the aggregate level:
  - Chordia and Shavikumar (2002) show that momentum is captured by business cycle variables.
  - Avramov and Chordia (2006a) demonstrate that momentum is captured by the component in model mispricing that varies with business conditions.
  - Market states:

* Cooper, Gutierrez, and Hameed (2008) show that momentum profitability heavily depends on the state of the market.
* In particular, from 1929 to 1995, the mean monthly momentum profit following positive market returns is 0.93%, whereas the mean profit following negative market return is -0.37%.
* The study is based one the market performance over three years prior to the implementation of the momentum strategy.

2.3 More on momentum

- Momentum is also prominent in cross country analysis of stock indexes.
- Momentum can also be found in trading currencies and commodities.
- But it should be noted that over the last decade momentum has been loosing its momentum in the US markets.

2.4 Rationalizing momentum interactions

- Avramov and Hore (2007) rationalize stock-level momentum interactions in a continuous time equilibrium setting.
- Here are the details.
- We employ the stochastic differential formulation of recursive utility (SDU) advocated by Duffie and Epstein (1992).

- SDU is the continuous time analog of the recursive intertemporal utility function of Epstein-Zin (1989) and Weil (1990).
- Theoretically, the Epstein-Zin-Weil preferences break the tight association between risk aversion and EIS.
- Empirically, they match the relations between asset returns and rate of consumption more closely.
- The SDU specification is

$$f(C,J) = \frac{\beta(1-\gamma)J}{1-\frac{1}{\psi}}\left[C^{1-\frac{1}{\psi}}\left((1-\gamma)J\right)^{\frac{\frac{1}{\psi}-1}{1-\gamma}} - 1\right] \quad (29)$$

  where C is current consumption, J is the continuation utility, and ψ is the EIS.
- That is, the current utility is a function of both current consumption and utility derived from future consumption.
- Perhaps to give some insight about SDU- note that the discrete time utility of Epstein-Zin (1989) and Weil (1990) is defined recursively by

$$J_t = W\left(C_t, \mathcal{M}(\sim J_{t+1}|\mathcal{F}_t)\right), \quad (30)$$

  where $\sim J_{t+1}|\mathcal{F}_t$ denotes the distribution of the utility J_{t+1} at time $t+1$ conditional on the information $\mathcal{F}$ available at time t and $\mathcal{M}$ is a certainty equivalent functional.
- In continuous time J has a SDE representation of the form

$$dJ_t = \left[-f(C_t, J_t) - \frac{1}{2}A(J_t)\sigma_J\sigma_J'\right]dt + \sigma_J dB_t \quad (31)$$

- The pair (f, A) generating J is called an aggregator.

- Roughly speaking the connection between (f, A) and $(W, \mathcal{M})$ is that f is the differential counterpart of W while A is a measure of the local risk aversion of $\mathcal{M}$.

- Note, in particular, that $A(J_t)$ is a variance multiplier applying a penalty as a multiple of the utility volatility.

- Integrating over (31) as well as zeroing out A we get

$$J_t = E_t \int_t^{\infty} f(c_s, J_s) ds. \tag{32}$$

- There is an ordinary equivalence between (f, A) and $(f, 0)$.

- Moreover, for analytical tractability we assume that $\psi = 1$, which is the point of equivalence between wealth and substitution effects.

- Specifically, if $\psi < 1$, the income effect dominates while if $\psi > 1$, the substitution effect dominates.

- The normalized aggregator is the limit of (29) taking the form

$$f(C, J) = \beta(1-\gamma)J \left[\log C - \frac{\log(1-\gamma)J}{1-\gamma} \right]. \tag{33}$$

- The joint system of the observed dividend growth and the unobserved expected dividend growth are formulated as

$$\frac{dD_t}{D_t} = X_t dt + \sigma_D dW_1, \tag{34}$$

$$dX_t = \kappa(\bar{X} - X_t)dt + \sigma_x dW_2. \tag{35}$$

- To account for leverage, we build on the novel formulation of Abel (1999).

- The consumption (specified exogenously) is a portion of dividend

$$C = D^{\lambda}, \tag{36}$$

while the remainder of the dividend stream is distributed as adjustment cost in the economy.

- The no-leverage case $\lambda = 1$ depicts an economy with no adjustment cost wherein the agent consumes the full dividend streams.

- Using Ito's lemma, the consumption growth dynamics takes the form

$$\frac{dC}{C} = \lambda\left[X_t + \frac{1}{2}(\lambda - 1)\sigma_D^2\right]dt + \lambda\sigma_D dW_1 \tag{37}$$

suggesting that expected consumption growth is slower than expected dividend growth as long as $\lambda < 1$.

- This is consistent with the US economy which demonstrates slow consumption growth along with relatively fast economic growth.

- Notice also that in a levered economy, the volatility of consumption $\lambda\sigma_D$ is smaller than the volatility of economic growth, again consistent with the US economy.

- Using the first-order condition of the Hamilton, Bellman, Jacobi equation for optimal control we are able to find an exact solution to the utility function given by

$$J(C_t, X_t) = \frac{C_t^{1-\gamma}}{1-\gamma}\exp\left(u_1 X_t + u_2\right) \tag{38}$$

where

$$u_1 = \frac{(1-\gamma)\lambda}{\kappa+\beta} \quad (39)$$

$$u_2 = \frac{(1-\gamma)\lambda}{\beta}\left[\frac{(\lambda-1-\lambda\gamma)\sigma_D^2}{2}+\frac{\kappa\bar{X}}{\kappa+\beta}+\frac{(1-\gamma)\lambda}{\kappa+\beta}\left[\frac{\sigma_x^2}{2(\kappa+\beta)}+\sigma_D\sigma_x\rho\right]\right]. \quad (40)$$

- Note that J_X is positive, J_{XX} is negative, and both κ and β impact the response of utility to expected growth rate changes.

- The equilibrium pricing kernel dynamics is given by

$$\frac{d\Lambda}{\Lambda} = -r_t^f dt - \lambda\gamma\sigma_D dW_1 + u_1\sigma_x dW_2 \quad (41)$$

where

$$r_t^f = \mu_C(X_t) + \beta - \sigma_C^2\gamma + \frac{(1-\gamma)\lambda}{\kappa+\beta}\sigma_C\sigma_x\rho. \quad (42)$$

- Note that leverage is an important determinant of both the drift and diffusion of the pricing kernel dynamics.

- The equilibrium risk-free rate has two appealing properties:
  - It varies with the expected growth rate.
  - It need not increase as γ increases, whereas in the power utility case $r_t^f = \gamma\mu_C(X_t) + \beta - \frac{\gamma(\gamma+1)\sigma_C^2}{2}$.

- The equilibrium price-dividend ratio $\frac{P_t}{D_t}$, denoted by $G(X_t)$, is

$$G(X_t) = \int_0^\infty \exp(P_1(\tau)X_t + P_2(\tau))d\tau \quad (43)$$

where P_1 and P_2 are the solutions of a system of ODEs given in our appendix.

- The excess return dynamics is given by

$$dR_t = \mu_t^R dt + \sigma_D dW_1 + \frac{G_X}{G}\sigma_x dW_2 \quad (44)$$

$$d\mu_t^R = (\cdot)dt + \left(\frac{G_X}{G}\right)_X (\lambda\gamma\sigma_D\sigma_x\rho - u_1\sigma_x^2)\sigma_x dW_2. \quad (45)$$

- The instantaneous covariance between past realized return and expected current return is given by

$$E_t\left[(dR_t - \mu_t^R dt)\left(d\mu_t^R - (\cdot)dt\right)\right] = \left(\frac{G_X}{G}\right)_X \left(\sigma_D\sigma_x\rho + \frac{G_X}{G}\sigma_X^2\right)(\lambda\gamma\sigma_D\sigma_x\rho - u_1\sigma_x^2). \quad (46)$$

- The SDEs of observed cumulative return and expected return for an investment horizon of l periods are given by

$$R_{t,t+l} = R_t + \int_t^{t+l}\mu_s^R ds + \int_t^{t+l}\sigma_D dW_1 + \int_t^{t+l}\frac{G_X}{G}\sigma_x dW_2 \quad (47)$$

$$\mu_{t,t+l}^R = \mu_t^R + \int_t^{t+l}(\cdot)ds + \int_t^{t+l}\left(\frac{G_X}{G}\right)_X (\lambda\gamma\sigma_D\sigma_x\rho - u_1\sigma_x^2)\sigma_x dW_2 \quad (48)$$

- To understand momentum effects in stock returns we compute the correlation between $R_{t,t+l}$ and $\mu_{t,t+l}^R$.
- First, the covariance between past realized return and expected current return is given by

$$\begin{aligned} Cov_t(R_{t,t+l}, \mu_{t,t+l}^R) &= E_t\left[(R_{t,t+l} - E_t R_{t,t+l})(\mu_{t,t+l}^R - E_t\mu_{t,t+l}^R)\right] \\ &= \sigma_x(\lambda\gamma\sigma_D\sigma_x\rho - u_1\sigma_x^2)\int_t^{t+l}\left(\frac{G_X}{G}\right)_X \left(\sigma_D\rho + \frac{G_X}{G}\sigma_x\right) ds. \quad (49) \end{aligned}$$

Then, the variances can also be computed as

$$V_t(\mu_{t+l}^R) = (\lambda\gamma\sigma_D\sigma_x\rho - u_1\sigma_x^2)^2\sigma_x^2\int_t^{t+l}\left(\left(\frac{G_X}{G}\right)_X\right)^2 ds \quad (50)$$

$$V_t(R_{t+l}) = \int_t^{t+l}\left(\sigma_D^2 + \left(\frac{G_X}{G}\sigma_x\right)^2 + 2\frac{G_X}{G}\sigma_D\sigma_x\rho\right) ds. \quad (51)$$

Hence, the correlation is

$$\Gamma(l) = \frac{\int_t^{t+l}\left(\frac{G_X}{G}\right)_X (\sigma_D\rho + \frac{G_X}{G}\sigma_x)ds}{\sqrt{\int_t^{t+l}\left(\left(\frac{G_X}{G}\right)_X\right)^2 ds \cdot \int_t^{t+l}\left(\sigma_D^2 + \left(\frac{G_X}{G}\sigma_x\right)^2 + 2\frac{G_X}{G}\sigma_D\sigma_x\rho\right) ds}}. \quad (52)$$

- The leading term on the numerator of $\Gamma(l)$ can be written as
$$\frac{GG_{XX}-(G_X)^2}{G^2}=\frac{1}{G^2}\left[\int_0^\infty \exp(\cdot)d\tau \int_0^\infty \exp(\cdot)P_1^2(\tau)d\tau-\left(\int_0^\infty \exp(\cdot)P_1(\tau)d\tau\right)^2\right] \quad (53)$$
which is positive from direct application of Cauchy-Schwartz inequality to functions $P_1(\tau)\sqrt{\exp(\cdot)}$ and $\sqrt{\exp(\cdot)}$, both of which are integrable in the domain.
- As we show below, the term $\frac{G_X}{G}$ is always positive for $\lambda < 1$.
- Thus, the autocorrelation $\Gamma(l)$ is positive unless two conditions are satisfied
  i. $\rho < 0$
  ii. $|\sigma_D \rho| > \frac{G_X}{G}\sigma_x$.
- The business cycle effect on the P/D ratio is instantly observable as
$$G_X=\frac{1-\lambda}{\kappa}\int_0^\infty \exp(P_1(\tau)X_t+P_2(\tau))\left(1-e^{-\kappa\tau}\right)d\tau. \quad (54)$$
- Thus the P/D ratio is increasing in the growth rate X_t as long as $\lambda < 1$. The effect is most pronounced for low λ (high leverage), it deteriorates as λ (low leverage) grows, and ultimately vanishes as λ approaches one.

2.4.1 Technical appendix for "Rationalizing Momentum Interactions"

Here are some details about deriving some of the quantities introduced in this section. Let us start with the consumption growth dynamics in (37) which is quite simple. In particular, since $C = f(D)$ by Ito's lemma
$$dC_t = f'(D)dD_t + \frac{1}{2}f''(D)(dD_t)^2 \quad (55)$$

where

$$
\begin{aligned}
f'(D) &= \lambda D_t^{\lambda-1}, & (56)\\
f''(D) &= \lambda(\lambda-1)D_t^{\lambda-2}, & (57)\\
(dD_t)^2 &= D_t^2\sigma_D^2 dt. & (58)
\end{aligned}
$$

The consumption growth dynamics $\frac{dC_t}{C_t}$ follows directly.

Deriving the utility function in (38) is more challenging. Here are the steps. The utility process J satisfies the Bellman equation with respect to equilibrium consumption

$$\mathcal{D}\mathcal{J}(C, X, t) + f(C, J) = 0 \tag{59}$$

where $\mathcal{D}\mathcal{J}$ is the differential operator applied to J with respect to $\{C, X, t\}$ with the boundary condition $J(C, x, T) = 0$. We are interested in the equilibrium as $T \to \infty$. Thus, we drop the explicit time dependence assuming that the agent is infinitely long-lived and has reached the equilibrium over time. We can formulate an exact solution to the value function as follows.

The Bellman equation in (59) can be written as

$$
\begin{aligned}
J_C dC + J_X dX + \frac{1}{2}J_{CC}(dC)^2 + \frac{1}{2}J_{XX}(dX)^2 + J_{XC}dCdX & \quad (60)\\
+\beta(1-\gamma)J\left[\log C - \frac{\log(1-\gamma)J}{1-\gamma}\right] = 0 &
\end{aligned}
$$

which leads to

$$J_C C\lambda \left[X_t + \frac{1}{2}(\lambda - 1)\sigma_D^2\right] + J_X \kappa(\bar{X} - X_t) + \frac{1}{2} J_{CC} C^2 \lambda^2 \sigma_D^2 + \frac{1}{2} J_{XX} \sigma_X^2 + J_{XC} C\lambda\sigma_D\sigma_x\rho + \beta(1-\gamma)J\left[\log C - \frac{\log(1-\gamma)J}{1-\gamma}\right] = 0 \quad (61)$$

The continuation utility J has a solution of the form

$$(1-\gamma)J = \exp(u_0 \ln C_t + u_1 X_t + u_2) \quad (62)$$

Then,

$$J_C = J\frac{u_0}{C}, \quad (63)$$
$$J_X = J u_1, \quad (64)$$
$$J_{CC} = J\frac{u_0}{C^2}(u_0 - 1), \quad (65)$$
$$J_{XX} = J u_1^2, \quad (66)$$
$$J_{CX} = J\frac{u_0 u_1}{C}. \quad (67)$$

Substituting the partial derivatives of J into (61) yields

$$J u_0 \lambda \left[X_t + \frac{1}{2}(\lambda - 1)\sigma_D^2\right] + J u_1 \kappa(\bar{X} - X_t) + \frac{1}{2} J u_0(u_0 - 1)\lambda^2\sigma_D^2 + \frac{1}{2} J u_1^2 \sigma_X^2 + J u_0 u_1 \lambda \sigma_D \sigma_x \rho + \beta(1-\gamma)J\left[\log C - \frac{u_0 \log C + u_1 X_t + u_2}{1-\gamma}\right] = 0 \quad (68)$$

Collecting terms reduces the above equation to a system of equations that can be solved recursively. In particular, focusing on $\log C$ we get

$$\log C\left(1-\frac{u_0}{1-\gamma}\right)=0, \tag{69}$$

which yields $u_0 = 1-\gamma$. Then moving to X_t it follows that

$$X_t\left(u_0\lambda - u_1\kappa - \beta u_1\right)=0, \tag{70}$$

which gives

$$u_1=\frac{u_0\lambda}{\kappa+\beta}. \tag{71}$$

Finally, u_2 follows by collecting terms not involving C_t and X_t.

We next find the riskfree rate and pricing kernel. Duffie and Epstein (1992) show that the pricing kernel for SDU is given by $\Lambda_t = \exp(\int_0^t f_J ds) f_C$, where f_J and f_C are the derivatives of $f(C,J)$ in (33) with respect to J and C. In particular,

$$f_C = \beta {C_t}^{-\gamma}\exp(u_1 X_t + u_2), \tag{72}$$

$$f_J = \beta(-1-u_1X_t-u_2). \tag{73}$$

Now let $G=f_C$ and note that

$$\frac{d\Lambda}{\Lambda}=\frac{dG}{G}+f_J dt. \tag{74}$$

The drift of $\frac{d\Lambda}{\Lambda}$ is the negative riskfree rate. Furthermore,

$$dG = G_C dC + \frac{1}{2}G_{CC}(dC)^2 + G_X dX + \frac{1}{2}G_{XX}(dX)^2 + G_{XC}dCdX \tag{75}$$

where

$$
\begin{aligned}
G_C &= -\frac{G\gamma}{C}, && (76)\\
G_{CC} &= \frac{G\gamma(\gamma+1)}{C^2}, && (77)\\
G_X &= Gu_1, && (78)\\
G_{XX} &= Gu_1^2, && (79)\\
G_{CX} &= -\frac{Gu_1\gamma}{C}, && (80)\\
dC &= C\lambda\left[X_t + \frac{1}{2}(\lambda-1)\sigma_D^2\right]dt + \lambda C\sigma_D dW_1, && (81)\\
dX &= \kappa(\bar{X} - X_t)dt + \sigma_X dW_2, && (82)\\
(dC)^2 &= \lambda^2 C^2 \sigma_D^2 dt, && (83)\\
(dX)^2 &= \sigma_X^2 dt, && (84)\\
(dXdC) &= C\lambda\sigma_D\sigma_X\rho dt. && (85)
\end{aligned}
$$

The pricing kernel dynamics including the riskfree rate follow by substituting the quantities in (75) - (85) into (74).

Next we find the price dividend ratio. The firm stock price is given by

$$
\begin{aligned}
P_t &= \frac{1}{\Lambda_t} E_t \int_t^\infty \Lambda_s D_s ds && (86)\\
&= \frac{1}{\Lambda_t} \int_t^\infty E_t \Lambda_s D_s ds && (87)\\
&= \frac{1}{\Lambda_t} \int_t^\infty E_t\left[\exp(\log\Lambda_s + \log D_s)\right] ds. && (88)
\end{aligned}
$$

Now, let $q_s = (\log\Lambda_s + \log D_s)$. Applying Ito's lemma to q_s

yields

$$dq_s = \frac{d\Lambda}{\Lambda} - \frac{1}{2\Lambda^2}d\Lambda^2 + \frac{dD}{D} - \frac{1}{2D^2}dD^2, \tag{89}$$

$$= (aX_s + b)dt + \phi_1 dW_1 + \phi_2 dW_2, \tag{90}$$

where

$$a = 1 - \lambda, \tag{91}$$

$$b = \frac{1}{2}\sigma_D^2(\gamma\lambda - 1) + u_1\kappa\bar{X} - \beta(1 + u_2), \tag{92}$$

$$\phi_1 = (1 - \gamma\lambda)\sigma_D, \tag{93}$$

$$\phi_2 = u_1\sigma_X. \tag{94}$$

Applying Feynman-Kac gives

$$E_t\left[\exp(q_s)\right] = f(q_t, X_t, t). \tag{95}$$

The process $P_t\Lambda_t$ is a martingale. Thus

$$f_q(aX_t + b) + f_X\kappa(\bar{X} - X_t) - f_\tau \tag{96}$$
$$+\frac{1}{2}f_{qq}(\phi_1 + \phi_2)^2 + \frac{1}{2}f_{XX}\sigma_X^2 + f_{qX}\sigma_X(\phi_1 + \phi_2) = 0.$$

As usual, we guess a solution of the form

$$f = \exp\left[p_0(\tau)q_t + p_1(\tau)X_t + p_2(\tau)\right]. \tag{97}$$

Then,

$$f_q = fp_0(\tau), \tag{98}$$

$$f_X = fp_1(\tau), \tag{99}$$

$$f_t = -f\left[p_0'(\tau)q_t + p_1'(\tau)X_t + p_2'(\tau)\right], \tag{100}$$

$$f_{qq} = fp_0^2(\tau), \tag{101}$$

$$f_{XX} = fp_1^2(\tau), \tag{102}$$

$$f_{Xq} = fp_0(\tau)p_1(\tau). \tag{103}$$

It immediately follows that

$$\begin{aligned}&p_0(\tau)(aX_t+b)+p_1(\tau)\kappa(\bar{X}-X_t)-p_0'(\tau)q_t-p_1'(\tau)X_t-p_2'(\tau)\\&+\frac{1}{2}p_0^2(\tau)(\phi_1+\phi_2)^2+\frac{1}{2}p_1^2(\tau)\sigma_X^2\\&+p_0(\tau)p_1(\tau)\sigma_X(\phi_1+\phi_2)=0.\end{aligned} \tag{104}$$

Also note $p_0(0) = 1$, $p_1(0) = 0$, and $p_2(0) = 0$. Collecting q_t terms we get

$$-p_0'(\tau)q_t = 0, \tag{105}$$

which gives $p_0(\tau) = 1$. Next, collecting X_t terms we get

$$(1-\lambda)-\kappa p_1(\tau)-p_1'(\tau)=0, \tag{106}$$

which gives

$$p_1(\tau)=\frac{1-\lambda}{\kappa}\left[1-\exp(-\kappa\tau)\right]. \tag{107}$$

Then p_2 is solving the ODE

$$\begin{aligned}p_2'(\tau)=u_1\kappa\bar{X}+\frac{1}{2}\sigma_D^2(\lambda\gamma-1)-\beta(u_2+1)+\\p_1(\tau)\left[\kappa\bar{X}+u_1\sigma_x^2-(\lambda\gamma-1)\sigma_D\sigma_x\rho\right]\\+\frac{1}{2}\left[(\lambda\gamma-1)^2\sigma_D^2+u_1^2\sigma_x^2-2u_1(\lambda\gamma-1)\sigma_D\sigma_x\rho+p_1^2(\tau)\sigma_x^2\right]\end{aligned}$$

Plugging in $p_1(\tau)$, $p_2(\tau)$ becomes

$$p_2(\tau)=a\tau+b(e^{-\kappa\tau}-1)+c(1-e^{-2\kappa\tau}) \tag{108}$$

where

$$a = \left(u_1 + \frac{1-\lambda}{\kappa}\right)\left(\bar{X}\kappa - (\lambda\gamma - 1)\sigma_D\sigma_x\rho + \frac{1}{2}\sigma_x^2\left(u_1 + \frac{1-\lambda}{\kappa}\right)\right)$$
$$+ \frac{1}{2}\sigma_D^2\lambda\gamma(\lambda\gamma - 1) - \beta(u_2 + 1) \tag{109}$$
$$b = \frac{1-\lambda}{\kappa^2}\left[\frac{\sigma_x^2(1-\lambda)}{\kappa} + \kappa\bar{X} + u_1\sigma_x^2 - (\lambda\gamma - 1)\sigma_D\sigma_x\rho\right] \tag{110}$$
$$c = \frac{\sigma_x^2}{4}\frac{(1-\lambda)^2}{\kappa^3} \tag{111}$$

So we have

$$P_t = \frac{1}{\Lambda_t}\int_t^\infty \exp\left(q_t + p_1(\tau)X_t + p_2(\tau)\right) ds, \tag{112}$$
$$= \frac{1}{\Lambda_t}\int_t^\infty \Lambda_t D_t \exp\left(p_1(\tau)X_t + p_2(\tau)\right) ds. \tag{113}$$

Hence,

$$\frac{P_t}{D_t} = \int_t^\infty \exp\left(p_1(\tau)X_t + p_2(\tau)\right) ds. \tag{114}$$

Next, we would like to find the return dynamics. For this, we need to track the $\frac{dP_t}{P_t}$ dynamics. Here are the steps. We can express the stock price as

$$P_t = G_t D_t \tag{115}$$

where G_t is on the right-hand-side of (114). Note also that

$$dG = G_t dt + G_X dX + \frac{1}{2}G_{XX}(dX)^2, \tag{116}$$
$$= (.)dt + G_X\sigma_x dW_2. \tag{117}$$

Then

$$dP = GdD + DdG + dDdG \tag{118}$$

and

$$\frac{dP}{P} = \frac{dD}{D} + \frac{dG}{G} + \frac{dD}{D}\frac{dG}{G}, \tag{119}$$

$$= (.)dt + \sigma_d dW_1 + \frac{G_X}{G}\sigma_x dW_2. \tag{120}$$

The drift is the expected return μ_t^R. To find the expected return dynamics we can use the relation

$$\mu_t^R - r_f = -\text{cov}\left(\frac{d\Lambda}{\Lambda}, \frac{dP}{P}\right), \tag{121}$$

where the right hand side obtains as

$$-\text{cov} = \text{cov}\left(\gamma\lambda\sigma_d dW_{1t} - u_1\sigma_x dW_{2t}, \sigma_d dW_1 + \frac{G_X}{G}\sigma_x dW_2\right), \tag{122}$$

$$= (\gamma\lambda\sigma_d^2 - \sigma_x u_1 \sigma_d \rho) + \frac{G_X}{G}(\gamma\lambda\sigma_d\sigma_x\rho - u_1\sigma_x^2). \tag{123}$$

Hence the expected return dynamics is

$$d\mu_t^R = (.)dt + \left(\frac{G_X}{G}\right)_X (\gamma\lambda\sigma_d\sigma_x\rho - u_1\sigma_x^2)\sigma_x dW_2. \tag{124}$$

3 Corporate Based Anomalies

- So far we have been dealing with asset pricing anomalies.
- The corporate finance literature has documented a host of other interesting anomalies.
- Some anomalies are suspected to be under-reaction to news some over reaction.
- Here is a list of prominent anomalies
  - Stock Split
  - Dividend initiation and omission
  - Stock repurchase
  - Spinoff
  - Merger arbitrage
- Fama (1998) is a comprehensive source of corporate anomalies.
- External Financing is something noticeable to talk about.
  - Finance research has documented negative relation between transactions of external financing and future stock returns.
  - In particular, future returns are typically low following IPOs (initial public offerings), SEOs (seasoned public offerings), debt offerings, and bank borrowings.
  - Conversely, future stock returns are typically high following stock repurchases.
  - Richardson and Sloan (2003) nicely summarize all external financing transactions in one measure.

– They show that their comprehensive measure of external financing has a stronger relation with future returns relative to measures based on individual transactions.

– The external financing measure, denoted by $\Delta XFIN$, is constructed as follows.

  $\Delta XFIN$ is the total cash received from issuance of new debt and equity offerings minus cash used for retirement of existing debt and equity. All components are normalized by the average value of total assets.

– This measure considers all sorts of equity offerings including common and preferred stocks as well as all sorts of debt offerings including straight bonds, convertible bonds, bank loans, notes, etc. Interest payments on debt as well as dividend payments on preferred stocks are not considered as retiring debt or equity. However, dividend payments on common stocks are considered as retiring equity. In essence, dividends on common stocks are treated as stock repurchases.

– The $\Delta XFIN$ measure can be decomposed as

$$\Delta XFIN = \Delta CEquity + \Delta PEquity + \Delta Debt \quad (125)$$

  where

  $\Delta CEquity$ is the common equity issuance minus common equity repurchase minus dividend
  $\Delta PEquity$ is the referred equity issuance minus retirement and repurchase of preferred stocks
  $\Delta Debt$ is the debt issuance minus debt retirement and repurchase.

4 Return Distribution and Investment Horizons

4.1 Multi period Returns and Continuous Compounding

- Let R_t be the time t rate of return including capital gain and dividend.
- The continuously compounded (cc) return is

$$r_t = \ln(1 + R_t) \tag{126}$$

- The holding period return (HPR) over an investment horizon of K periods is

$$\begin{aligned} R(t+1, t+K) &= \prod_{k=1}^{K}(1 + R_{t+k}), & (127) \\ &= \exp\left(\sum_{k=1}^{K} r_{t+k}\right). & (128) \end{aligned}$$

- Thus, the continuously compounded HPR is

$$r_{t+1,t+K} = \sum_{k=1}^{K} r_{t+k} \tag{129}$$

- We often assume that the cc return is iid normally distributed with mean μ and variance σ^2. Then

$$r_{t+1,t+K} \sim N(K\mu, K\sigma^2) \tag{130}$$

- What is the distribution of HPR?
- Lognormal with mean and variance given by

$$\begin{aligned} E[HPR] &= \exp\left\{K(\mu + \sigma^2/2)\right\}, & (131) \\ V[HPR] &= \exp\left\{2K(\mu + \sigma^2/2)\right\}[\exp(K\sigma^2) - 1] & (132) \end{aligned}$$

- Note that the lognormal distribution is not symmetric. The median of HPR is

$$\text{median}[HPR] = \exp\{K\mu\}, \tag{133}$$

suggesting that the mean to median ratio is $\exp\left(\frac{K}{2}\sigma^2\right)$.

- This, in turn, implies that the pdf of the HPR becomes more positively skewed the longer the investment horizon.

4.2 Partitioning the return interval

- Let r_y denote the cc return for period y - say one year. Partition a year into H intervals, and let $r_{y,h}$ $h = 1, \ldots, H$ denote the cc return over interval h in period y.

- Note that

$$r_y = \sum_{h=1}^{H} r_{y,h}. \tag{134}$$

- If r_y has mean and variance μ and σ^2, the moments of $r_{y,h}$ are

$$E\{r_{y,h}\} = \mu_H = \frac{\mu}{H}, \tag{135}$$

$$var\{r_{y,h}\} = \sigma_H^2 = \frac{\sigma^2}{H} \tag{136}$$

- Assume we have a sample of Y years so that there are $n = YH$ observations of the short interval returns.

- Define the sample moments by $\bar{r}$ and s^2 - then the estimators of μ and σ are

$$\hat{\mu} = H\bar{r}, \tag{137}$$

$$\hat{\sigma}^2 = Hs^2. \tag{138}$$

- Clearly these estimators are unbiased. Moreover,

$$
\begin{aligned}
var\{H\bar{r}\} &= H^2 var(\bar{r}), && (139)\\
&= H^2\frac{\sigma_H^2}{n}, && (140)\\
&= \frac{\sigma^2}{Y}, && (141)
\end{aligned}
$$

which is the same as if the annual returns were used.

- That is, partitioning the return interval does not improve the estimator of μ.
- However, it does improve the estimator for σ^2.
- In particular,

$$
\begin{aligned}
var\{Hs^2\} &= H^2\frac{2}{n-1}\sigma_H^4, && (142)\\
&= H^2\frac{2}{HY-1}\frac{\sigma^4}{H^2}, && (143)\\
&= \frac{2\sigma^4}{HY-1}. && (144)
\end{aligned}
$$

- Notice that the variance approaches 0 as H becomes large.
- Andersen et al (2001) exploit this property to estimate the realized variance.

4.3 Variance and investment horizon when returns are predictable

- Let r_t be the cc return in time t and r_{t+1} be the cc return in time $t+1$.
- Assume that $var(r_t) = var(r_{t+1})$.

- Then the cumulative two period return is $R(t-1,t+1) = r_t + r_{t+1}$.
- The question of interest: is $var(R(t-1,t+1))$ greater than equal to or smaller than $2 * var(r_t)$.
- Of course if stock returns are iid the variance growths linearly with the horizon as noted earlier.
- However, let us assume that stock returns can be predictable by the dividend yield.
- That is, we consider the predictive system

$$r_t = \alpha + \beta div_{t-1} + \epsilon_t \quad (145)$$

$$div_t = \phi + \delta div_{t-1} + \eta_t \quad (146)$$

  where $var(\epsilon_t) = \sigma_1^2$, $var(\eta_t) = \sigma_2^2$, and $cov(\epsilon_t, \eta_t) = \sigma_{12}$ and the residuals are uncorrelated in all leads and lags.
- It follows that

$$var(r_t + r_{t+1}) = 2\sigma_1^2 + \beta^2\sigma_2^2 + 2\beta\sigma_{12} \quad (147)$$

- Thus, if $\beta^2\sigma_2^2 + 2\beta\sigma_{12} < 0$ the conditional variance of two period return is less than the twice conditional variance of one period return, which is indeed the case based on empirical evidence.
- That is, the longer the investment horizon the less risky equities appear if one is making investment decisions based on the current value of the dividend yield. Thus one is willing to invest more in equities over longer investment horizons.

5 Beta Pricing versus the Discount Factor Representation

5.1 The First Order Conditions

- The absence of arbitrage in a dynamic economy guarantees the existence of a strictly positive discount factor that prices all traded assets [see Harisson and Kreps (1979)].
- Asset prices are set by the investors' first order condition:

$$\mathrm{E}\left[\xi_{t+1} R_{i,t+1} | \mathcal{I}_t\right] = 1, \tag{148}$$

  where $\mathrm{E}[\bullet|\mathcal{I}_t]$ is the expectation operator conditioned on $\mathcal{I}_t$, the full set of information available to investors at time t.
- The fundamental pricing equation (148) holds for any asset either stock, bond, option, or real investment opportunity.
- It holds for any two subsequent periods t and $t+1$ of a multi-period model.
- It does not assume complete markets
- It does not assume the existence of a representative investor
- It does not assume equilibrium in financial markets.
- It imposes no distributional assumptions about asset returns nor any particular class of preferences.
- Let us now replace the consumption-based expression for marginal utility growth with a linear model obeying the form

$$\xi_{t+1} = a_t + b_t' f_{t+1}. \tag{149}$$

- Notation: a_t and b_t are fixed or time-varying parameters and f_{t+1} denotes $K \times 1$ vector of fundamental factors that are proxies for marginal utility growth.
- Theoretically, the pricing kernel representation is equivalent to the beta pricing specification.
- See equations (14) and (15) in Avramov (2004) and the references therein.
- The CAPM, for one, says that

$$\xi_{t+1} = a_t + b_t r_{w,t+1}, \tag{150}$$

where $r_{w,t+1}$ is the time $t+1$ return on a claim to total wealth.

5.2 Is the pricing Kernel linear or nonlinear in the factors?

- In a single-period economy the pricing kernel is given by

$$\xi_{t+1} = \frac{U'(W_{t+1})}{U'(W_t)}. \tag{151}$$

- The Taylor's series expansion of the pricing kernel around $U'(W_t)$ is

$$\begin{aligned} \xi_{t+1} &= 1 + \frac{W_t U''(W_t)}{U'(W_t)} r_{w,t+1} + o(W_t), & (152) \\ &= a + b r_{w,t+1}, & (153) \end{aligned}$$

where $a = 1+o(W_t)$ and b is the negative relative risk aversion coefficient.

- This first order approximation results in the traditional CAPM.

- The second order approximation is given by

$$\begin{aligned}\xi_{t+1} &= 1 + \frac{W_t U''(W_t)}{U'(W_t)} r_{w,t+1} + \frac{W_t^2 U'''(W_t)}{2U'(W_t)} r_{w,t+1}^2 + o(W_t), \\ &= a + b r_{w,t+1} + c r_{w,t+1}^2. \end{aligned} \tag{154}$$

- This additional factor is related to co skewness in asset returns.
- Harvey and Siddique (2000) exhibit the relevance of this factor in explaining the cross sectional variation in expected returns.

5.3 Are pricing kernel parameters fixed or time-varying?

- Let us start with preferences represented by $U(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}$.
- Take Taylor's series expansion of the pricing kernel $\rho \frac{U'(c_{t+1})}{U'(c_t)}$ around $U'(c_t)$ and obtain

$$\xi_{t+1} = 1 - \gamma \Delta c_{t+1} + o(c_t), \tag{155}$$

$$= a + b \Delta c_{t+1}. \tag{156}$$

- Under the power preferences, pricing kernel parameters are time invariate (time varying) if γ is time invariate (time varying).
- Next, consider the habit-formation economy of Campbell and Cochrane (1999).
- The utility function under habit formation

$$U(c_t, x_t) = \frac{(c_t - x_t)^{1-\gamma}}{1-\gamma}, \tag{157}$$

where x_t is the external consumption habit.

- Then the pricing kernel parameters are time-varying even when the risk aversion parameter is constant.
- See, e.g., Lettau and Ludvigson (2000).
- Modeling time variation:
  - Assume that a_t and b_t are linear functions of z_t in a conditional single-factor model:

$$\begin{aligned} \xi_{t+1} &= a(z_t) + b(z_t) f_{t+1}, && (158) \\ a_t &= a_0 + a_1 z_t, && (159) \\ b_t &= b_0 + b_1 z_t. && (160) \end{aligned}$$

- Then a conditional single-factor model becomes an unconditional multifactor model

$$\xi_{t+1} = a_0 + a_1 z_t + b_0 f_{t+1} + b_1 f_{t+1} z_t. \tag{161}$$

- The set of factors is $= [z_t, f_{t+1}, f_{t+1} z_t]'$.
- The multi-factor representation of (161) is

$$\xi_{t+1} = a_0 + a_1' z_t + b_0' f_{t+1} + b_1' [f_{t+1} \otimes z_t]. \tag{162}$$

- Later we will formulate asset pricing tests based on the pricing kernel representation.
- But we first focus on the beta pricing representation.
- Even when both methods are equivalent theoretically - the empirical tests and their statistical properties are quite different.
- Below we present test statistics based on both methods.

6 Estimating and Evaluating Asset Pricing Models

We first use Regression-Based Tests of Linear Factor Models

6.1 Time-Series Regressions

- The simplest case thoroughly analyzed by Campbell, Lo, and MacKinlay (1997, Chapter 5) is the single factor model

$$r_{i,t} = \alpha_i + \beta_i f_t + \epsilon_{i,t}. \tag{163}$$

- The regression intercepts reflect the pricing errors as long as the factors are return spreads.
- As noted earlier, an asset pricing model says that the regression intercepts must be zero.
- To test whether all the pricing errors are jointly equal to zero we assume that the regression residuals are IID Normal.
- Later we will deal with GMM based methods that relax the IID Normal assumptions.

6.1.1 An asymptotic test statistic

- Let us derive the test statistic using concepts of maximum likelihood estimation:

  1. Rewriting equation (163) in its multivariate form yields

  $$r_t = \alpha + \beta f_t + \epsilon_t. \tag{164}$$

  2. Let Θ represent the set of parameters: $\Theta = (\alpha', \beta', vech(\Sigma)')'$.

3. Under normality, the likelihood function for ϵ_t is

$$\mathcal{L}(\epsilon_t|\Theta) = (2\pi)^{-\frac{N}{2}}\Sigma^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(r_t - \alpha - \beta f_t)'\Sigma^{-1}(r_t - \alpha - \beta f_t)\right]. \tag{165}$$

Moreover, the IID assumption suggests that

$$\begin{aligned} \mathcal{L}(\epsilon_1, \epsilon_2, \ldots, \epsilon_N|\Theta) &= (2\pi)^{-\frac{TN}{2}}\Sigma^{-\frac{T}{2}} \qquad (166) \\ &\times \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(r_t - \alpha - \beta f_t)'\Sigma^{-1}(r_t - \alpha - \beta f_t)\right]. \end{aligned}$$

4. The estimates of α, β, and Σ are obtained by maximizing the likelihood function (166).

5. In particular, let us denote the set of estimates by $\hat{\Theta}$.

6. Asymptotically

$$\Theta - \hat{\Theta} \sim N(0, I(\Theta)^{-1}), \tag{167}$$

where $I(\Theta)$ is the information matrix, the matrix of second derivatives of the likelihood with respect to the parameters.

7. Let us estimate the parameters

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \Sigma^{-1}\left[\sum_{t=1}^{T}(r_t - \alpha - \beta f_t)\right], \tag{168}$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = \Sigma^{-1}\left[\sum_{t=1}^{T}(r_t - \alpha - \beta f_t)f_t\right], \tag{169}$$

$$\frac{\partial \mathcal{L}}{\partial \Sigma} = -\frac{T}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}\left[\sum_{t=1}^{T}\epsilon_t\epsilon_t'\right]. \tag{170}$$

8. So the MLEs are given by

$$\hat{\alpha} = \hat{\mu} - \hat{\beta}\hat{E}(f), \tag{171}$$

$$\hat{\beta} = \frac{\sum_{t=1}^{T}(r_t - \hat{\mu})(f_t - \hat{E}(f))}{\sum_{t=1}^{T}(f_t - \hat{E}(f))^2}, \tag{172}$$

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t{}', \tag{173}$$

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t. \tag{174}$$

9. Let us now find the standard errors of $\hat{\alpha}$ and $\hat{\beta}$.

10. Note that the information matrix is given by

$$I(\Theta) = \begin{bmatrix} -T\Sigma^{-1} & -\Sigma^{-1}E(f) \\ -\Sigma^{-1}E(f) & -\Sigma^{-1}\left(E(f)^2 + \sigma(f)^2\right) \end{bmatrix}. \tag{175}$$

we disregard the third row and third column since Σ is uncorrelated with α and β.

11. Thus, it follows that

$$\hat{\alpha} \sim N\left(\alpha, \frac{1}{T}\left[1 + \frac{\hat{E}(f)^2}{\hat{\sigma}(f)^2}\right]\Sigma\right). \tag{176}$$

12. We can form a Wald test statistic of the null hypothesis

$$\mathcal{H}_0 : \ \alpha = 0 \tag{177}$$

against the alternative hypothesis

$$\mathcal{H}_1 : \ \alpha \neq 0. \tag{178}$$

13. The Wald statistic is

$$J_1 = \hat{\alpha}'[\mathrm{Var}(\hat{\alpha})]^{-1}\hat{\alpha}. \tag{179}$$

14. Bottom line:

$$J_1 = T \left[1 + \left(\frac{\hat{E}(f)}{\hat{\sigma}(f)} \right)^2 \right]^{-1} \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}. \tag{180}$$

- Next, let us take the test statistic into the data.
- The algorithm for implementing the statistic is as follows:
  1. Run separate regressions for the test assets on the common factor.
  2. Retain the estimated regression intercepts $\hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_N]'$ and the estimated residuals $\epsilon_t = [\epsilon_t^1, \epsilon_t^2, \ldots, \epsilon_t^N]'$ for $t = [1, 2, \ldots, T]$.
  3. Compute the residual covariance matrix $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t \epsilon_t'$.
  4. Compute the sample mean and the sample standard deviation of the factor $\hat{E}(f)$ and $\hat{\sigma}(f)$, respectively.
  5. Compute J_1.
- The asymptotic distribution of J_1 is χ^2_N.

6.1.2 A finite sample test

- A finite sample test statistic for the joint equality of the pricing errors to zero has been developed by Gibbons, Ross, and Shanken – GRS (1989).
- The GRS test is

$$J_2 = \frac{T - N - 1}{N} \left[1 + \left(\frac{\hat{E}(f)}{\hat{\sigma}(f)} \right)^2 \right]^{-1} \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}. \tag{181}$$

- The statistic J_2 has the finite-sample distribution $F_{N,T-N-1}$.
- What about the economics of the time series test statistics?
- GRS show that their test statistic can be expressed in terms of how far inside the ex post frontier the reference return is

$$J_2 = \frac{T-N-1}{N}\left[\frac{(\mu_q/\sigma_q)^2 - \left(\hat{E}(f)/(\hat{\sigma}(f)\right)^2}{1+\left(\hat{E}(f)/(\hat{\sigma}(f)\right)^2}\right], \quad (182)$$

where $(\mu_q/\sigma_q)^2$ is the Sharpe ratio of the ex post tangency portfolio formed from the tangency portfolio.

- So, testing the joint equality of the pricing errors to zero amounts to testing whether the portfolio-based factor is ex-ante mean variance efficient — whether it is on the mean-variance efficient frontier based on *population* moments.
- Of course, by using sample moments, which involve estimation errors, you should not expect the factor to lie on the ex post frontier even if it is on the true ex ante frontier.
- To this point, we have dealt with a single portfolio-based factor.
- What if there are multiple factors such as the Fama and French (1993) three-factor model?
- The regression specification is

$$r_{i,t} = \alpha_i + \beta_i' f_t + \epsilon_{i,t}. \quad (183)$$

- The pricing restrictions are the same: the pricing errors reflected by the regression intercepts should be equal to zero.

- The GRS statistic is then generalized to

$$J_3 = \frac{T-N-K}{N}\left[1+\hat{E}(f)'\hat{\Omega}^{-1}\hat{E}(f)\right]^{-1}\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} \sim F_{N,T-N-K}, \tag{184}$$

  where K is the number of factors and $\hat{\Omega}$ is the estimate of the variance covariance matrix of the factors.

- In sum, the central question in the time-series framework is whether the factor is ex ante mean-variance efficient (or, in multi-factor models, whether some optimal combination of the factors is on the ex ante efficient frontier).

- Time series based test statistics indicate how far away is the tested portfolio (or a combination of portfolios in a multi-factor model) from the efficient frontier.

- There are several other time series tests described in CLM Chapter 5. One of these is a GMM test which we describe below in the GMM section.

6.2 Cross Sectional Regressions

- The time series procedures are designed primarily to test asset pricing models based on factors that are asset returns.

- The cross-sectional technique can be implemented whether or not the factor is a return spread.

- Consumption growth is a good example of a non portfolio based factor.

- The central question in the cross section framework is why average returns vary across assets.

- For instance, the CAPM says that average excess returns should be proportional to betas.
- So plot the sample average excess returns on the estimated betas.
- But even if the model is correct, this plot will not work out perfectly well because of sampling errors.
- The idea is to run a cross-sectional regression to fit a line through the scatterplot of average returns on estimated betas.
- Then examine the deviations from a linear relation.
- In the cross section approach you can also examine whether a factor is indeed priced.

- Let us formalize the concepts.
- Two regression steps are at the heart of the cross-sectional approach:
  - First, estimate betas from the time-series regression of excess returns on some pre-specified factors
  $$r_{i,t} = \alpha_i + \beta_i' f_t + \epsilon_{i,t}. \tag{185}$$
  - Then run the cross-section regression of average returns on the betas
  $$\bar{r}_i = \beta_i' \lambda + \nu_i. \tag{186}$$
- Notation: λ – the regression coefficient – is the risk premium, and ν_i — the regression disturbance – is the pricing error.
- Assume for analytic tractability that there is a single factor, let $\bar{r} = [\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_N]'$, and let $\beta = [\beta_1, \beta_2, \ldots, \beta_N]'$.

- The OLS cross-sectional estimates are

$$\hat{\lambda} = (\beta'\beta)^{-1}\beta'\bar{r}, \quad (187)$$

$$\hat{\nu} = \bar{r} - \hat{\lambda}\beta. \quad (188)$$

- Furthermore, let Σ be the covariance matrix of asset returns, then it follows that

$$\sigma^2(\hat{\lambda}) = \frac{1}{T}(\beta'\beta)^{-1}\beta'\Sigma\beta(\beta'\beta)^{-1}, \quad (189)$$

$$\text{cov}(\hat{\nu}) = \frac{1}{T}\left(I - \beta(\beta'\beta)^{-1}\beta'\right)\Sigma\left(I - \beta(\beta'\beta)^{-1}\beta'\right) \quad (190)$$

- We could test whether all pricing errors are zero with the statistic

$$\hat{\nu}'cov(\hat{\nu})^{-1}\hat{\nu} \sim \chi^2_{N-1}. \quad (191)$$

- We could also test whether a factor is priced

$$\frac{\hat{\lambda}}{\sigma(\hat{\lambda})} \sim t_{N-1} \quad (192)$$

- Thus far, we assume that β in (186) is known.
- However, β is estimated in the time-series regression and therefore is not a known parameter.
- So we have the EIV problem.
- Shanken (1992) corrects the cross-sectional estimates to account for the errors in estimating betas
- Shanken assumes homoscedasticity in the variance of asset returns conditional upon the realization of factors.

- Under this assumption he shows that the standard errors based on the cross sectional procedure overstate the precision of the estimated parameters.

- The EIV corrected estimates are

$$\sigma^2_{eiv}(\hat{\lambda}) = \sigma^2(\hat{\lambda})\Upsilon + \frac{1}{T}\Omega_f \tag{193}$$

$$\text{cov}_{eiv}(\hat{\nu}) = \text{cov}(\hat{\nu})\Upsilon, \tag{194}$$

  where Ω_f is the variance-covariance matrix of the factors and $\Upsilon = 1 + \lambda'\Omega_f^{-1}\lambda$.

- Of course, if factors are return spreads then $\lambda'\Omega_f^{-1}\lambda$ is the squared Sharpe ratio attributable to a mean-variance efficient investment in the factors.

6.3 Fama and MacBeth (FM) Procedure

- FM (1973) propose an alternative procedure for running cross-sectional regressions, and for producing standard errors and test statistics.
- The FM approach involves two steps as well.
- The first step is identical to the one described above. Specifically, estimate beta from a time series regression.
- The second step is different.
- In particular, instead of estimating a single cross-sectional regression with the sample averages on the estimated betas, FM run a cross-sectional regression at *each time period*

$$r_{i,t} = \delta_{0,t} + \beta_i'\delta_{1,t} + \epsilon_{i,t}. \tag{195}$$

- Let $r_t = [r_{1,t}, r_{2,t}, \ldots, r_{N,t}]'$, let $\delta_t = [\delta_{0,t}, \delta_{1,t}]'$, let $X_i = [1, \beta_i]'$, and let $X = [X_1, X_2, \ldots, X_N]'$ then the cross sectional estimates for δ_t and $\epsilon_{i,t}$ are given by

$$\hat{\delta}_t = (X'X)^{-1}X'r_t, \tag{196}$$

$$\hat{\epsilon}_{i,t} = r_{i,t} - X_i'\hat{\delta}_t. \tag{197}$$

- FM suggest that we estimate δ and ϵ_i as the averages of the cross-sectional estimates

$$\hat{\delta} = \frac{1}{T}\sum_{t=1}^{T}\hat{\delta}_t, \tag{198}$$

$$\hat{\epsilon}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{i,t}. \tag{199}$$

- They suggest that we use the cross-sectional regression estimates to generate the sampling error for these estimates

$$\sigma^2(\hat{\delta}) = \frac{1}{T^2}\sum_{t=1}^{T}\left(\hat{\delta}_t - \hat{\delta}\right)^2, \tag{200}$$

$$\sigma^2(\hat{\epsilon}_i) = \frac{1}{T^2}\sum_{t=1}^{T}\left(\hat{\epsilon}_{i,t} - \hat{\epsilon}_i\right)^2. \tag{201}$$

- It is natural to use this sampling theory to test whether all the pricing errors are jointly zero as we have before.

- In particular, let $\hat{\epsilon} = [\hat{\epsilon}_1, \hat{\epsilon}_2, \ldots, \hat{\epsilon}_N]'$, then the variance-covariance matrix of the sample pricing errors is

$$cov(\hat{\epsilon}) = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{\epsilon}_t - \hat{\epsilon})(\hat{\epsilon}_t - \hat{\epsilon})'. \tag{202}$$

- Then we can test whether all pricing errors are zero using the test statistic (191).

6.4 Another useful cross sectional approach

- Below we describe an asset pricing test that applies to both portfolios as well as single securities.
- The ability to test pricing models using single stocks is sound.
- For one it avoids the data-snooping biases that are inherent in portfolio based approaches, as noted by Lo and MacKinlay (1990).
- It is also robust to the sensitivity of asset pricing tests to the portfolio grouping procedure.
- Brennan, Chordia, and Subrahmanyam (BCS) (1998) develop the test.
- Avramov and Chordia (2006a) extend the BCS methodology to test pricing models with potential time varying factor loadings.
- The first stage is a time series regression of excess returns on asset pricing factors with time varying factor loadings.
- The second stage is a regression of risk adjusted returns on equity characteristics.
- Under the null of exact pricing, equity characteristics should be statistically insignificant in the cross-section.
- The practice of using risk adjusted returns, rather than gross or excess returns, is intended to address the finite sample bias attributable to errors in estimating factor loadings in the first-pass time series regressions.
- Let us formalize the Avramov-Chordia test.

- Assume that returns are generated by a conditional version of a K-factor model

$$R_{jt} = E_{t-1}(R_{jt}) + \sum_{k=1}^{K} \beta_{jkt-1} f_{kt} + e_{jt}, \tag{203}$$

  where E_{t-1} is the conditional expectations operator, R_{jt} is the return on security j at time t, f_{kt} is the unanticipated (with respect to information available at $t-1$) time t return on the k'th factor, and β_{jkt-1} is the conditional beta.

- $E_{t-1}(R_{jt})$ is modeled using the exact pricing specification

$$E_{t-1}(R_{jt}) - R_{Ft} = \sum_{k=1}^{K} \lambda_{kt-1} \beta_{jkt-1}, \tag{204}$$

  where R_{Ft} is the riskfree rate and λ_{kt} is the risk premium for factor k at time t.

- The estimated risk-adjusted return on each security for month t is then calculated as:

$$R^*_{jt} \equiv R_{jt} - R_{Ft} - \sum_{k=1}^{K} \hat{\beta}_{jkt-1} F_{kt}, \tag{205}$$

  where $F_{kt} \equiv f_{kt} + \lambda_{kt-1}$ is the sum of the factor innovation and its corresponding risk premium and $\hat{\beta}_{jkt}$ is the conditional beta estimated by a first-pass time-series regression over the entire sample period as per the specification given below.

- Our risk adjustment procedure assumes that the conditional zero-beta return equals the conditional risk-free rate, and that the factor premium is equal to the excess return on the factor, as is the case when factors are return spreads.

- Next, we run the cross-sectional regression

$$R^*_{jt} = c_{0t} + \sum_{m=1}^{M} c_{mt} Z_{mjt-1} + e_{jt}, \qquad (206)$$

  where Z_{mjt-1} is the value of characteristic m for security j at time $t-1$, and M is the total number of characteristics.

- Under exact pricing, equity characteristics do not explain risk-adjusted return, and are thus insignificant in the specification (206).

- To examine significance, we estimate the vector of characteristics rewards each month as

$$\hat{c}_t = (Z'_{t-1} Z_{t-1})^{-1} Z'_{t-1} R^*_t, \qquad (207)$$

  where Z_{t-1} is a matrix including the M firm characteristics for N_t test assets and R^*_t is the vector of risk-adjusted returns on all test assets.

- To formalize the conditional beta framework developed here let us rewrite the specification in (206) using the generic form

$$R_{jt} - [R_{Ft} + \beta(\theta, z_{t-1}, X_{jt-1})' F_t] = c_{0t} + c_t Z_{jt-1} + e_{jt}, \qquad (208)$$

  where X_{jt-1} and Z_{jt-1} are vectors of firm characteristics, z_{t-1} denotes a vector of macroeconomic variables, and θ represents the parameters that capture the dependence of β on the macroeconomic variables and the firm characteristics.

- Ultimately, the null to test is $c_t = 0$.

- While we have checked the robustness of our results for the general case where $X_{jt-1} = Z_{jt-1}$, the paper focuses on the case where the factor loadings depend upon firm-level size, book-to-market, and business-cycle variables.

- That is, the vector X_{jt-1} stands for size and book-to-market and the vector Z_{jt-1} stands for size, book-to-market, turnover, and various lagged return variables.

- The dependence on size and book-to-market is motivated by the general equilibrium model of Gomes, Kogan, and Zhang (2003), which justifies separate roles for size and book-to-market as determinants of beta.

- In particular, firm size captures the component of a firm's systematic risk attributable to its growth option, and the book-to-market ratio serves as a proxy for risk of existing projects.

- Incorporating business-cycle variables follows the extensive evidence on time series predictability (see, e.g., Keim and Stambaugh (1986), Fama and French (1989), and Chen (1991)).

- In the first pass, the conditional beta of security j is modeled as

$$\beta_{jt-1} = \beta_{j1} + \beta_{j2} z_{t-1} + (\beta_{j3} + \beta_{j4} z_{t-1}) Size_{jt-1} + (\beta_{j5} + \beta_{j6} z_{t-1}) BM_{jt-1}, \quad (209)$$

  where $Size_{jt-1}$ and BM_{jt-1} are the market capitalization and the book-to-market ratio at time $t-1$.

- The first pass time series regression for the very last specification is

$$\begin{aligned} r_{jt} &= \alpha_j + \beta_{j1} r_{mt} + \beta_{j2} z_{t-1} r_{mt} + \beta_{j3} Size_{jt-1} r_{mt} \\ &+ \beta_{j4} z_{t-1} Size_{jt-1} r_{mt} \quad (210) \\ &+ \beta_{j5} BM_{jt-1} r_{mt} + \beta_{j6} z_{t-1} BM_{jt-1} r_{mt} + u_{jt}, \end{aligned}$$

  where $r_{jt} = R_{jt} - R_{Ft}$ and r_{mt} is excess return on the value-weighted market index.

- Then, R^*_{jt} in (206), the dependent variable in the cross-section regression, is given by $\alpha_j + u_{jt}$.
- The time series regression (210) is run over the entire sample.
- While this entails the use of future data in calculating the factor loadings, Fama and French (1992) have shown that this forward looking does not impact any of the results.
- For perspective, it is useful to compare our approach to earlier studies
  - Fama and French (1992) estimate beta by assigning the firm to one of 100 size-beta sorted portfolios. Firm's beta (proxied by the portfolio's beta) is allowed to evolve over time when the firm changes its portfolio classification.
  - Fama and French (1993) focus on 25 size and book-to-market sorted portfolios, which allow firms' beta to change over time as they move between portfolios.
  - Brennan, Chordia, and Subrahmanyam (1998) estimate beta each year in a first-pass regression using 60 months of past returns. They do not explicitly model how beta changes as a function of size and book-to-market, as we do, but their rolling regressions do allow beta to evolve over time.
- We should also distinguish our beta-scaling procedure from those proposed by Shanken (1990) and Ferson and Harvey (1999) as well as Lettau and Ludvigson (2001).
- Shanken and Ferson and Harvey use predetermined variables to scale factor loadings in asset pricing tests.

- Lettau and Ludvigson use information variables to scale the pricing kernel parameters. In both procedures, a one-factor conditional CAPM can be interpreted as an unconditional multifactor model.
- Our conditional beta factor model does not have that unconditional multifactor interpretation since the firm-level $Size_j$ and BM_j are not common across all test assets.

6.4.1 Asset Pricing Models using Individual Stocks

- Follow up to Avramov and Chordia (2006a) - several recent paper advocate the use of individual stocks for asset pricing tests
- See in particular, Lewellen, Nagel, and Shanken (2010), Ang, Liu, and Schwarz (2008), and Daniel and Titman (2011).

6.5 GMM

6.5.1 Overview

- We can test finance theories by the GMM of Hansen (1982).
- Let us describe the basic concepts of GMM and propose some applications.
- Let Θ be an $m \times 1$ vector of parameters to be estimated from a sample of observations $x_1, x_2, \ldots, x_T$.
- A popular classical estimator for Θ is $\hat{\Theta}$ that maximizes the sample likelihood.

- One drawback in the maximum likelihood principle is that it requires specifying the joint density of the observations.
- The ML principle is indeed a parametric one.
- ML typically makes the IID, Normal, and homoskedastic assumptions.
- All these assumptions can be relaxed in the GMM framework.
- The GMM only requires specification of certain moment conditions (often referred as orthogonality conditions) rather than the full density.
- It is therefore considered a nonparametric approach.
- Do not get it wrong: The GMM is not necessarily perfect.
- First, it may not make efficient use of all the information in the sample.
- Second, non parametric approaches typically have low power in out of sample tests possibly due to over-fitting.
- Also the GMM is asymptotic and can have poor, even measurable, final sample properties.
- As always, you win some you lose some.
- Let $f_t(\Theta)$ be an $r \times 1$ vector of moment conditions.
- Note that f_t is not necessarily linear in the data or the parameters, and it can be heteroskedastic and serially correlated.
- If $r = m$, i.e., if there is the same number of parameters as there are moments, then the system is exactly identified.

- In this case one could find the GMM estimate $\hat{\Theta}$, which satisfies

$$E\left(f_t(\hat{\Theta})\right) = 0. \tag{211}$$

- However, in testing economic theories, there should be more moment conditions than there are parameters.
- In this case, one cannot set *all* the moment conditions to be equal to zero, as I show below.
- Let us analyze both cases of exact identification ($r = m$) and over identification ($r > m$).
- To implement the GMM first compute the sample average of $E[f_t(\Theta)]$ given by

$$g_T(\Theta) = \frac{1}{T}\sum_{t=1}^{T} f_t(\Theta). \tag{212}$$

- If $r = m$, then the GMM estimator $\hat{\Theta}$ solves

$$g_T(\hat{\Theta}) = \frac{1}{T}\sum_{t=1}^{T} f_t(\hat{\Theta}) = 0. \tag{213}$$

- Otherwise, the GMM estimator minimizes the quadratic form

$$J_T(\Theta) = g_T(\Theta)' W_T g_T(\Theta), \tag{214}$$

  where W_T is some $r \times r$ weighing matrix to be discussed later.
- Differentiating (214) with respect to Θ yields

$$D_T(\Theta)' W_T g_T(\Theta), \tag{215}$$

  where

$$D_T(\Theta) = \frac{\partial g_T(\Theta)}{\partial \Theta'}. \tag{216}$$

- The GMM estimator $\hat{\Theta}$ solves

$$D_T(\hat{\Theta})'W_T g_T(\hat{\Theta}) = 0. \tag{217}$$

- Observe from (217) that the left (and obviously the right) hand side is an $m \times 1$ vector. Therefore, as $r > m$ only a linear combination of the moments, given by $D_T(\hat{\Theta})'W_T$, is set to zero.

- Hansen (1982, theorem 3.1) tells us the asymptotic distribution of the GMM estimate is

$$\sqrt{T}(\hat{\Theta} - \Theta) \sim N(0, V), \tag{218}$$

where

$$\begin{aligned} V &= \left(D_0'WD_0\right)^{-1} D_0'WSWD_0(D_0'WD_0)^{-1}, & (219)\\ S &= \lim_{T\to\infty} Var\left[\sqrt{T} g_T(\Theta)\right], & (220)\\ &= \sum_{j=-\infty}^{\infty} E\left[f_t(\Theta) f_{t-j}(\Theta)'\right], & (221)\\ D_0 &= E\left[\frac{\partial g_T(\Theta)}{\partial \Theta'}\right] = \frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial f_t(\Theta)}{\partial \Theta'}\right]. & (222) \end{aligned}$$

- To implement the GMM one would like to replace S with its sample estimate.

- If the moment conditions are serially uncorrelated then

$$S_T = \frac{1}{T}\sum_{t=1}^{T} f_t(\hat{\Theta}) f_t(\hat{\Theta})'. \tag{223}$$

- We have not yet addressed the issue of how to choose the optimal weighting matrix.

- Hansen shows that optimally $W = S^{-1}$.
- The optimal V matrix is therefore

$$V^* = \left(D_0' S^{-1} D_0\right)^{-1}. \tag{224}$$

- Moreover, if $W = S^{-1}$, i.e., if the weighting matrix is chosen optimally, then an overidentifying test statistic is given by

$$TJ_T(\hat{\Theta}) \sim \chi^2_{r-m}. \tag{225}$$

- This statistic is quite intuitive.
- In particular, note from (220) that $S_T = TVar\left[g_T(\hat{\Theta})\right]$.
- Thus, the test statistic in (225) can be expressed as the minimized value of the model errors (in asset pricng context pricing errors) weighted by their covariance matrix

$$g_T(\hat{\Theta})' \left\{var\left[g_T(\hat{\Theta})\right]\right\}^{-1} g_T(\hat{\Theta}) \sim \chi^2_{r-m}. \tag{226}$$

- Below we display several applications of the GMM.
- The work of Hansen and Singleton (1982, 1983) is, to my knowledge, the first to apply the GMM in general and in the context of asset pricing in particular.

6.5.2 Application# 1: Estimating the mean of a time series

- You observe $x_1, x_2, \ldots, x_T$ and want to estimate the sample mean.
- In this case there is a single parameter $\Theta = \mu$ and a single moment condition

$$f_t(\Theta) = (x_t - \mu). \tag{227}$$

The system is exactly identified.

- Notice that

$$g_T(\Theta) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \mu). \tag{228}$$

- Setting

$$g_T(\hat{\Theta}) = 0. \tag{229}$$

- Then the GMM estimate for μ is the sample mean

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} x_t. \tag{230}$$

- Moreover, if x_t's are uncorrelated then

$$S = E\left[f_t(\Theta) f_t(\Theta)'\right], \tag{231}$$

estimated using

$$S_T = \frac{1}{T}\sum_{t=1}^{T} f_t(\hat{\Theta}) f_t(\hat{\Theta})' = \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{\mu})^2. \tag{232}$$

- To compute the variance of the estimate we need to find D_0:

$$D_0 = E\left[\frac{\partial g_T(\Theta')}{\partial \Theta}\right] = \frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial f_t(\Theta)}{\partial \Theta'}\right] = -1. \tag{233}$$

- The optimal V matrix is then given by

$$V = (D_0' S^{-1} D_0)^{-1} = S. \tag{234}$$

Use S_T as a consistent estimator.

- Asymptotically

$$\hat{\mu} \sim N\left(\mu, \frac{1}{T} V\right).$$

6.5.3 Application # 2: Estimating the market model coefficients when the residuals are heteroskedastic and serially uncorrelated

- In this application we will focus on a single security.
- The next application expands the analysis to accommodate multiple assets.
- Here is the market model for security i

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \epsilon_{i,t}. \tag{235}$$

- There are two parameters: $\Theta = [\alpha_i, \beta_i]'$.
- There are also two moment conditions

$$f_t(\Theta) = \begin{bmatrix} r_{i,t} - \alpha_i - \beta_i r_{m,t} \\ (r_{i,t} - \alpha_i - \beta_i r_{m,t}) r_{m,t} \end{bmatrix}. \tag{236}$$

- Let us rewrite the moment conditions compactly using the following form

$$f_t(\Theta) = x_t(r_t - x_t'\beta), \tag{237}$$

where

$$\begin{aligned} r_t &= r_{i,t}, & (238) \\ x_t &= [1, r_{m,t}]', & (239) \\ \epsilon_t &= \epsilon_{i,t}, & (240) \\ \beta &= [\alpha_i, \beta_i]'. & (241) \end{aligned}$$

- Let us now compute the sample moment

$$g_T(\Theta) = \frac{1}{T}\sum_{t=1}^{T} x_t(r_t - x_t'\beta). \tag{242}$$

- Since the system is exactly identified setting $g_T(\hat{\Theta}) = 0$ yields the GMM estimate

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\sum_{t=1}^{T} x_t r_t\right), \tag{243}$$

$$= (X'X)^{-1} X'R, \tag{244}$$

where $X = [x_1, x_2, \ldots, x_T]'$ and $R = [r_1, r_2, \ldots, r_T]'$.

- The GMM estimator for β is the usual OLS estimator.

- To find V first compute

$$\frac{\partial f_t(\Theta)}{\partial \Theta'} = -x_t x_t', \tag{245}$$

$$\frac{\partial g_T(\Theta)}{\partial \Theta'} = D_T(\Theta) = -\frac{1}{T}\sum_{t=1}^{T} x_t x_t' = -\frac{X'X}{T}. \tag{246}$$

- Moreover, if ϵ's are serially uncorrelated then

$$S_T = \frac{1}{T}\sum_{t=1}^{T} f_t(\hat{\Theta}) f_t(\hat{\Theta})', \tag{247}$$

$$= \frac{1}{T}\sum_{t=1}^{T} x_t \hat{\epsilon}_t \hat{\epsilon}_t' x_t', \tag{248}$$

$$= \frac{1}{T}\sum_{t=1}^{T} x_t x_t' \hat{\epsilon}_t^2. \tag{249}$$

- Using the optimal weighting matrix, it follows that

$$Var(\hat{\beta}) = \frac{1}{T}(D_0' S^{-1} D_0)^{-1}, \tag{250}$$

estimated by

$$\hat{Var}(\hat{\beta}) = \frac{1}{T}\left(D_T' S_T^{-1} D_T\right)^{-1}, \quad (251)$$

$$= (X'X)^{-1}\left(\sum_{t=1}^{T} x_t x_t' \hat{\epsilon}_t^2\right)(X'X)^{-1}. \quad (252)$$

6.5.4 Application # 3: Testing the CAPM

- Here, we derive the CAPM test described on pages 208-210 in Campbell, Lo, and MacKinlay.
- The specification that we have is

$$r_t^e = \alpha + \beta r_{mt}^e + \epsilon_t. \quad (253)$$

- The CAPM says $\alpha = 0$.
- The $2N \times 1$ parameter vector in the CAPM model is described by $\Theta = [\alpha', \beta']'$.
- In the following I will give a recipe for implementing the GMM in estimating and testing the CAPM.

  1. Start with identifying the $2N$ moment conditions:

  $$f_t(\Theta) = x_t \otimes \epsilon_t = \begin{bmatrix} 1 \\ r_{m,t}^e \end{bmatrix} \otimes \epsilon_t = \begin{bmatrix} \epsilon_t \\ r_{m,t}^e \epsilon_t \end{bmatrix}, \quad (254)$$

  where $\epsilon_t = r_t^e - (x_t' \otimes I_N)\Theta$.

  2. Compute D_0.

  $$\frac{\partial f_t(\Theta)}{\partial \Theta'} = x_t \otimes -\left(x_t' \otimes I_N\right), \quad (255)$$

  $$= -\begin{bmatrix} 1 & r_{m,t}^e \\ r_{m,t}^e & r_{m,t}^{e2} \end{bmatrix} \otimes I_N. \quad (256)$$

Moreover,

$$D_0 = E\left[\frac{\partial g_T(\Theta)}{\partial \Theta'}\right] \quad (257)$$

$$= E\left[\frac{\partial f_t(\Theta)}{\partial \Theta'}\right] \quad (258)$$

$$= -\begin{bmatrix} 1 & \mu_m \\ \mu_m & \mu_m^2 + \sigma_m^2 \end{bmatrix} \otimes I_N, \quad (259)$$

where $\mu_m = E(r_{m,t})$ and $\sigma_m^2 = var(r_{m,t})$.

3. In implementing the GMM, D_0 will be replaced by its sample estimate, which amounts to replacing the population moments μ_m and σ_m^2 by their sample analogs $\hat{\mu}_m$ and $\hat{\sigma}_m^2$.
4. That is,

$$D_T = -\begin{bmatrix} 1 & \hat{\mu}_m \\ \hat{\mu}_m & \hat{\mu}_m^2 + \hat{\sigma}_m^2 \end{bmatrix} \otimes I_N, \quad (260)$$

5. There are as many moments conditions as there are parameters.
6. Still, you can test the CAPM since you only focus on the model restriction $\alpha = 0$.
7. In particular, compute $g_T(\Theta)$ and find the GMM estimator

$$g_T(\Theta) = \frac{1}{T}\sum_{t=1}^{T} f_t(\Theta), \quad (261)$$

$$= \frac{1}{T}\sum_{t=1}^{T}(x_t \otimes \epsilon_t). \quad (262)$$

The GMM estimator $\hat{\Theta}$ satisfies $g_T(\hat{\Theta}) = 0$

$$\frac{1}{T}\sum_{t=1}^{T}(x_t \otimes \hat{\epsilon}_t) = 0 \tag{263}$$

$$\frac{1}{T}\sum_{t=1}^{T}\left(x_t \otimes \left[r_t - (x_t' \otimes I_N)\hat{\Theta}\right]\right) = 0 \tag{264}$$

$$\frac{1}{T}\sum_{t=1}^{T}(x_t \otimes r_t) = \frac{1}{T}\sum_{t=1}^{T}(x_t \otimes x_t' \otimes I_N)\hat{\Theta} \tag{265}$$

The GMM estimator is thus given by

$$\begin{aligned}
\hat{\Theta} &= \left[\begin{pmatrix} 1 & \hat{\mu}_m \\ \hat{\mu}_m & \hat{\mu}_m^2 + \hat{\sigma}_m^2 \end{pmatrix}^{-1} \otimes I_N\right]\begin{pmatrix} \hat{\mu} \\ \hat{\sigma}_{rm} + \hat{\mu}\hat{\mu}_m \end{pmatrix}, \\
&= \left[\frac{1}{\hat{\sigma}_m^2}\begin{pmatrix} \hat{\mu}_m^2 + \hat{\sigma}_m^2 & -\hat{\mu}_m \\ -\hat{\mu}_m & 1 \end{pmatrix} \otimes I_N\right]\begin{pmatrix} \hat{\mu} \\ \hat{\sigma}_{rm} + \hat{\mu}\hat{\mu}_m \end{pmatrix}, \\
&= \begin{bmatrix} \hat{\mu} - \frac{\hat{\sigma}_{rm}}{\hat{\sigma}_m^2}\hat{\mu}_m \\ \frac{\hat{\sigma}_{rm}}{\hat{\sigma}_m^2} \end{bmatrix} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}.
\end{aligned} \tag{266}$$

These are the OLS estimators for the CAPM parameters.

8. Estimate S assuming that the moment conditions are serially uncorrelated

$$\begin{aligned}
S_T &= \frac{1}{T}\sum_{t=1}^{T} f_t(\hat{\Theta})f_t(\hat{\Theta})', && (267) \\
&= \frac{1}{T}\sum_{t=1}^{T}(x_t x_t' \otimes \hat{\epsilon}_t\hat{\epsilon}_t') && (268)
\end{aligned}$$

9. Given S_T and D_T compute V_T the sample estimate of the optimal variance matrix

$$V_T = \left(D_T' S_T^{-1} D_T\right)^{-1} \tag{269}$$

10. The asymptotic distribution of $\hat{\Theta}$ is given by

$$\hat{\Theta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \frac{1}{T}V \right), \tag{270}$$

so you should substitute V_T for V.

11. Now, we can derive the test statistic. In particular, let $\alpha = R\Theta$ where $R = [I_N, 0_N]$ and note that under the null hypothesis $\mathcal{H}_0 : \alpha = 0$ the asymptotic distribution of $R\hat{\Theta}$ is given by

$$R\hat{\Theta} \sim N\left(0, R(\frac{1}{T}V)R'\right). \tag{271}$$

The statistic J_7 in CLM is derived using the Wald statistic

$$J_7 = \hat{\Theta}'R'\left(R(\frac{1}{T}V_T)R'\right)^{-1} R\hat{\Theta}, \tag{272}$$

$$= T\hat{\alpha}'\left(R(D_T'S_T^{-1}D_T)^{-1}R'\right)\hat{\alpha}. \tag{273}$$

12. Under the null hypothesis $J_7 \sim \chi^2_N$.

- It should be noted that if the regression errors are both serially uncorrelated and homoscedastic then the matrix S_T in (268) is

$$\frac{1}{T}(X'X) \otimes \Sigma = \begin{bmatrix} 1 & \hat{\mu}_m \\ \hat{\mu}_m & \hat{\mu}_m^2 + \hat{\sigma}_m^2 \end{bmatrix} \otimes \Sigma \tag{274}$$

- Thus J_7 in (273) becomes the traditional Wald statistic in (180).

- So, the GMM test statistic is a generalized version of Wald correcting for heteroscedasticity.

- You can also relax the non serial correlation assumption.

6.5.5 Application # 4: Asset pricing tests based on over-identification

- Harvey (1989) nicely implements the GMM to test conditional asset pricing models.
- The conditional CAPM, wlog, implies that

$$\begin{aligned} \mathbb{E}(r_t|z_{t-1}) &= cov(r_t, r_{mt}|z_{t-1})\lambda_t, \quad (275) \\ &= \mathbb{E}\left[(r_t - \mathbb{E}[r_t|z_{t-1}])(r_{mt} - \mathbb{E}[r_{mt}|z_{t-1}])|z_{t-1}\right]\lambda_t, \end{aligned}$$

where z_t denotes a set of M instruments observed at time t.

- Let us assume that λ_t is constant, that is $\lambda_t = \lambda$ for all t.
- Let $x_t = [1, z_t']'$.
- Moreover,

$$\mathbb{E}[r_t|z_{t-1}] = \delta_r x_{t-1}, \quad (276)$$

$$\mathbb{E}[(r_{mt}|z_{t-1}] = \delta_m x_{t-1}. \quad (277)$$

- Then, let us define several residuals

$$u_{rt} = r_t - \delta_r x_{t-1}, \quad (278)$$

$$u_{mt} = r_{mt} - \delta_m x_{t-1}, \quad (279)$$

$$\begin{aligned} e_t &= r_t - \mathbb{E}\left[(r_t - \mathbb{E}[r_t|z_{t-1}])(r_{mt} - \mathbb{E}[r_{mt}|z_{t-1}])|z_{t-1}\right]\lambda \\ &= r_t - (r_t - \delta_r x_{t-1})(r_{mt} - \delta_m x_{t-1})\lambda. \quad (280) \end{aligned}$$

- Collecting the residuals into one vector yields

$$f_t(\Theta) = [u_{rt}', u_{mt}, e_t']', \quad (281)$$

where $\Theta = [vec(\delta_r)', \delta_m', \lambda]'$.

- That is, there are $M(N+1)+1$ parameters.

- How many moment conditions do we have? More than you think!
- Note that

$$\mathbb{E}[f_t(\Theta)|z_{t-1}] = 0, \tag{282}$$

which means that we have the following $2N + 1$ moment conditions

$$\mathbb{E}[f_t(\Theta)] = 0, \tag{283}$$

as well as $M(2N+1)$ additional moment conditions involving the instruments

$$\mathbb{E}[f_t(\Theta) \otimes z_{t-1}] = 0. \tag{284}$$

- Overall, there are $(2N + 1)(M + 1)$ moment conditions.
- You have more moment conditions than parameters.
- Hence, you can test the model using the χ^2 over identifying test.
- Harvey considers several other generalizations.

6.5.6 Application # 5: Testing return predictability over long horizon

- Here is a nice GMM application in the context of return predictability over long return horizon.
- The model estimated is

$$R_{t+k} = \alpha + \beta' z_t + \epsilon_{t+k} \tag{285}$$

where

$$R_{t+k} = \sum_{i=1}^{k} r_{t+i} \tag{286}$$

with r_{t+i} being log return at time $t + i$.

- Fama and French (1989) observe a dramatic increase in the sample R^2 as the return horizon grows from one month to four years.

- Kirby (1997) challenges the Fama-French findings using a GMM framework that accounts for serial correlation in the residuals.

- Let us formalize his test statistic.

- There are $M+1$ parameters $\Theta = (\alpha, \beta')'$, where β is a vector of dimension M.

- From Hansen (1982)

$$\sqrt{T}(\Theta - \hat{\Theta}) \sim N(0, V), \tag{287}$$

where

$$V = (D_0' S^{-1} D_0)^{-1}. \tag{288}$$

- There are $M+1$ moment conditions:

$$f_t(\Theta) = \begin{bmatrix} R_{t+k} - \alpha - \beta' z_t \\ (R_{t+k} - \alpha - \beta' z_t) z_t \end{bmatrix}. \tag{289}$$

- Compute D_0:

$$D_0 = \mathbb{E}\left[\frac{\partial f_t(\Theta)}{\partial \Theta'}\right] = \begin{bmatrix} -1 & -\mu_z \\ -\mu_z' & -(\Sigma_z + \mu_z \mu_z') \end{bmatrix}. \tag{290}$$

- It would be useful to take the inverse of D_0

$$D_0^{-1} = \begin{bmatrix} -(1 + \mu_z' \Sigma_z^{-1} \mu_z) & \mu_z' \Sigma_z^{-1} \\ \Sigma_z^{-1} \mu_z & -\Sigma_z^{-1} \end{bmatrix}. \tag{291}$$

- And the matrix S is given by

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E}\begin{bmatrix} \epsilon_{t+k}\epsilon_{t+k-j} & \epsilon_{t+k}\epsilon_{t+k-j}z'_{t-j} \\ \epsilon_{t+k}\epsilon_{t+k-j}z_t & \epsilon_{t+k}\epsilon_{t+k-j}z_t z'_{t-j} \end{bmatrix}. \quad (292)$$

- In estimating predictive regressions we scrutinize the slope coefficients only.

- And we know that

$$\sqrt{T}(\beta - \hat{\beta}) \sim N(0, \tilde{V}), \quad (293)$$

where $\tilde{V}$ is the $M \times M$ lower-right submatrix of $V = D_0^{-1} S D_0^{-1\prime}$.

- It follows that

$$\begin{aligned} \tilde{V} &= \sum_{j=-\infty}^{\infty} \mathbb{E}\begin{bmatrix} \mu_z' \Sigma_z^{-1} \\ -\Sigma_z^{-1} \end{bmatrix}' \begin{bmatrix} \epsilon_{t+k}\epsilon_{t+k-j} & \epsilon_{t+k}\epsilon_{t+k-j}z'_{t-j} \\ \epsilon_{t+k}\epsilon_{t+k-j}z_t & \epsilon_{t+k}\epsilon_{t+k-j}z_t z'_{t-j} \end{bmatrix} \begin{bmatrix} \mu_z' \Sigma_z^{-1} \\ -\Sigma_z^{-1} \end{bmatrix}, \\ &= \Sigma_z^{-1} \left[\sum_{j=-\infty}^{\infty} \mathbb{E}\left(\epsilon_{t+k}\epsilon_{t+k-j}\right)(z_t - \mu_z)(z_{t-j} - \mu_z)' \right] \Sigma_z^{-1}, \\ &= \Sigma_z^{-1} \left[\sum_{j=-\infty}^{\infty} \mathbb{E}\left(\delta_{t+k}\delta'_{t+k-j}\right) \right] \Sigma_z^{-1}, \end{aligned} \quad (294)$$

where

$$\delta_{t+k} = \epsilon_{t+k} z_t, \quad (295)$$
$$\delta_{t+k-j} = \epsilon_{t+k-j} z_{t-j}. \quad (296)$$

- What if you are willing to assume that there is no autocorrelation in the residuals?

- Then,

$$\tilde{V} = \Sigma_z^{-1} \left[\mathbb{E}\left(\delta_{t+k}\delta'_{t+k}\right)\right] \Sigma_z^{-1}, \quad (297)$$

which can be estimated by

$$\tilde{V} = \hat{\Sigma}_z^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\delta_{t+k}\delta_{t+k}'\right]\hat{\Sigma}_z^{-1}. \tag{298}$$

The estimator for $\tilde{V}$ is identical to the heteroskedasticity-consistent covariance matrix estimator of White (1980).

- What if you are willing to assume that there is no autocorrelation in the residuals and there is no heteroskedasticity?

$$\begin{aligned}\tilde{V} &= \Sigma_z^{-1}\left[\mathbb{E}\left(\epsilon_{t+k}\epsilon_{t+k-j}\right)\mathbb{E}(z_t-\mu_z)(z_{t-j}-\mu_z)'\right]\Sigma_z^{-1}, \\ &= \sigma_\epsilon^2\Sigma_z^{-1}. \end{aligned} \tag{299}$$

- In that case

$$\sqrt{T}(\hat{\beta}-\beta) \sim N(0,\sigma_\epsilon^2\Sigma_z^{-1}). \tag{300}$$

- Under the null hypothesis that $\beta = 0$

$$\sqrt{T}\hat{\beta} \sim N(0,\sigma_\epsilon^2\Sigma_z^{-1}). \tag{301}$$

- Note also that under the null $\sigma_\epsilon^2 = \sigma_r^2$, where σ_r^2 is the variance of the cumulative log return.

- Using properties of the χ^2 distribution, it follows that

$$T\frac{\hat{\beta}'\hat{\Sigma}_z\hat{\beta}}{\hat{\sigma}_r^2} \sim \chi^2(M), \tag{302}$$

suggesting that

$$TR^2 \sim \chi^2(M). \tag{303}$$

- So we are able to derive a limiting distribution for the regression R^2.

- Kirby considers cases with heteroskedasticity and serial correlation.
- Then the distribution of the regression slope coefficient and the R^2 are much more complex.
- His conclusion: the R^2 in a predictive regression does not increase with the investment horizon.

6.5.7 Incorporating serial correlation

- In most applications in financial economics there is no a priori reason to believe that the regression residuals are serially uncorrelated.
- Consequently, a suitable scheme is required in order to obtain a consistent positive definite estimator of S.
- Notice that we cannot estimate the infinite sum in

$$S = \sum_{j=-\infty}^{\infty} E(u_t u_{t-j}). \tag{304}$$

- Therefore, we must limit the number of terms.
- More terms means more ability to pick up autocorrelation if there is any.
- But this comes at the cost of losing efficiency in finite samples.
- Newey and West (1987) propose a popular weighting scheme

$$\hat{S} = \sum_{j=-k}^{k} \frac{|k-j|}{k} \left(\frac{1}{T} \sum_{t=j+1}^{T} (u_t - \bar{u})(u_{t-j} - \bar{u})' \right), \tag{305}$$

where $\bar{u}$ denotes the sample mean and $k-1$ is the maximum lag length that receives a nonzero weight.

- The value k should increase with the sample size but not too rapidly.
- This estimator guarantees positive definiteness by downweighting higher order autocovariances, and it is consistent because the downweighting disappears asymptotically.
- The Newey and West (1987) weighting scheme is the most commonly used.
- Andrews (1991) develops a more complex, albeit useful, estimator.
- Another way to estimate S is by first running a VAR(1) for u_t

$$u_t = Au_{t-1} + v_t. \tag{306}$$

- With this structure you can estimate S via

$$\hat{S} = (I - A)^{-1} E(v_t v_t')(I - A)^{-1\prime}. \tag{307}$$

6.6 The Hansen Jagannathan — HJ — (1997) Distance Measure

The HJ measure is used for comparing and testing asset pricing models.

6.6.1 Motivation

- Suppose you want to compare the performance of competing, not necessarily nested, asset pricing models.

- If there is only one asset, then you can compare the pricing error, i.e., the difference between the market price of an asset and the hypothetical price implied by a particular SDF.
- However, when there are many assets, it is rather difficult to compare the pricing errors across the different candidate SDFs unless pricing errors of one SDF are always smaller across all assets.
- One simple idea would be to examine the pricing error on the portfolio (there are infinitely many such portfolios) that is most mispriced by a given model.
- Then, the superior model is the one with the smallest pricing error.
- However, there is a practical problem in implementing this simple idea.
- Suppose there are at least two assets which do not have the same pricing error for a given candidate SDF.
- Let R_{1t} and R_{2t} denote the corresponding gross returns.
- Suppose that (i) the date $t-1$ market prices of these payoffs are both unity, and (ii) the model assigns prices of $1+\psi_i$, i.e., the pricing errors are ψ_1 and ψ_2.
- Consider now forming a zero-investment portfolio by going long one dollar in security 1 and short one dollar in security 2.
- The pricing error of this zero-cost position is $\psi_1 - \psi_2$.

- That is, as long as the difference is not zero the pricing error of any portfolio of the two assets can be arbitrarily large by adding a scale multiple of this zero-investment portfolio.

6.6.2 The HJ idea

- HJ propose a way of normalization.
- They suggest examining the portfolio which has the maximum pricing errors among all portfolio payoffs that have the unit second moments.
- Let us demonstrate.
- Suppose that the SDF is modeled as

$$\begin{aligned} \xi_t(\Theta) &= \Theta_0 + \Theta_{vw} R_t^{vw} + \Theta_{prem} R_{t-1}^{prem} + \Theta_{labor} R_t^{labor}, \\ &= \Theta' Y_t, \end{aligned} \tag{308}$$

where

$$\Theta = [\Theta_0, \Theta_{vw}, \Theta_{prem}, \Theta_{labor}]', \tag{309}$$
$$Y_t = [1, R_t^{vw}, R_{t-1}^{prem}, R_t^{labor}]'. \tag{310}$$

- Moreover, let $R_t = [R_{1t}, R_{2t}, \ldots, R_{Nt}]'$, and let

$$f_t(\Theta) = R_t \xi_t(\Theta) - \iota_N = R_t Y_t' \Theta - \iota_N. \tag{311}$$

- Observe that $E[f_t(\Theta)]$ is the vector of pricing errors.
- In unconditional models, the number of moment conditions is equal to N, the number of test assets.
- HJ show that the maximum pricing error per unity norm of any portfolio of these N assets is given by

$$\delta = \sqrt{E[f_t(\Theta)'][E(R_t R_t')]^{-1} E[f_t(\Theta)]}. \tag{312}$$

- This is the HJ distance measure (which is not the HJ bound).
- Since the vector Θ is unknown, a natural way to estimate the system is to choose those values that minimize (312).
- We can then assess the specification error of a given stochastic discount factor by examining the maximum pricing error δ.
- Next, compute some sample moments

$$\begin{aligned} D_T &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t(\Theta)}{\partial \Theta} = \frac{1}{T}\sum_{t=1}^{T} R_t Y_t' = \frac{1}{T}R'Y, && (313)\\ g_T(\Theta) &= \frac{1}{T}\sum_{t=1}^{T} f_t(\Theta) = D_T\Theta - \iota_N, && (314)\\ G_T &= \frac{1}{T}\sum_{t=1}^{T} R_t R_t' = \frac{1}{T}R'R, && (315) \end{aligned}$$

where

$$\begin{aligned} R &= [R_1, R_2, \ldots, R_T]', && (316)\\ Y &= [Y_1, Y_2, \ldots, Y_T]. && (317) \end{aligned}$$

- The sample analog of the HJ distance is thus

$$\delta_T = \sqrt{\min_{\Theta} g_T(\Theta)' G_T^{-1} g_T(\Theta)}. \quad (318)$$

- The first order condition of the minimization problem

$$\min_{\Theta} g_T(\Theta) G_T^{-1} g_T(\Theta), \quad (319)$$

is given by

$$D_T' G_T^{-1} g_T(\Theta) = 0, \quad (320)$$

which gives an analytic expression for the sample minimizer

$$\hat{\Theta} = \left(D_T' G_T^{-1} D_T\right)^{-1} D_T' G_T^{-1} \iota_N, \quad (321)$$

$$= T\left(Y'R(R'R)^{-1}R'Y\right)^{-1} Y'R(R'R)^{-1}\iota_N. \quad (322)$$

- It follows that

$$g_T(\hat{\Theta}) = R'Y\left(Y'R(R'R)^{-1}R'Y\right)^{-1} Y'R(R'R)^{-1}\iota_N - \iota_N. \quad (323)$$

- From Hansen (1982) the asymptotic variance of $\hat{\Theta}$ is given by

$$var(\hat{\Theta}) = \frac{1}{T}(D_T' G_T^{-1} D_T)^{-1} D_T' G_T^{-1} S_T G_T^{-1} D_T (D_T' G_T^{-1} D_T)^{-1}, \quad (324)$$

where, if the data is serially uncorrelated, the estimate of the variance matrix of pricing errors is given by

$$S_T = \frac{1}{T}\sum_{t=1}^{T} f_t(\hat{\Theta}) f_t(\hat{\Theta})'. \quad (325)$$

- That is, the estimator $\hat{\Theta}$ is equivalent to a GMM estimator defined by Hansen (1982) with the moment condition $E[f(\Theta)] = 0$ and the weighting matrix G^{-1}.

- If the weighing matrix is optimal in the sense of Hansen (1982), then $T\delta_T^2$ is asymptotically a random variable of χ^2 distribution with $N - m$ dof, where m is the dimension of Θ.

- Moreover, the optimal variance of $\hat{\Theta}$ becomes

$$var(\hat{\Theta}) = \frac{1}{T}(D_T' S_T^{-1} D_T)^{-1}. \quad (326)$$

- However, G is generally not optimal, and thus the distribution of $T\delta_T^2$ is not χ^2_{N-m}.

- Instead, the limiting distribution of this statistic is given by

$$u = \sum_{j=1}^{N-m} \lambda_j \nu_j, \tag{327}$$

where $\nu_1, \nu_2, \dots \nu_{N-m}$ are independent $\chi^2(1)$ random variables, and $\lambda_1, \lambda_2, \dots \lambda_{N-m}$ are $N-m$ nonzero eigenvalues of the matrix A given by

$$A = S^{0.5} G'^{-0.5} (I_N - (G^{-0.5}) D [D' G^{-1} D]^{-1} D' G'^{-0.5})(G^{-0.5})(S^{0.5})', \tag{328}$$

and where $S^{0.5}$ and $G^{0.5}$ are the upper-triangle matrices from the Cholesky decomposition of S and G.

- As long as we have a consistent estimate S_T of the matrix S, we can estimate the matrix A by replacing S and G by S_T and G_T, respectively.

- One can generate a large number of draws from the nonstandard distribution (327) to determine the p-value of the HJ distance measure, or whether or not it is equal to zero.

- You can follow the below-described algorithm to compute the empirical p-value:

  1. Use results from (315) and (323) to compute $T\delta_T^2 = T g_T(\hat{\Theta}) G_T^1 g_T(\hat{\Theta})$.
  2. Obtain the $N-m$ largest eigenvalues of $\hat{A}$, a consistent estimate of A.
  3. Generate $N-m$ independent draws from $\chi^2(1)$. For example, using the Matlab command $g = chi2rnd(\nu, 1000, 1)$ generates 1000 independent draws from $\chi^2(\nu)$

4. Based on these independent draws, compute the statistic u_i in (327)
5. If $u_i > T\delta_T^2$ set $I_i = 1$. Otherwise set $I_i = 0$
6. Repeat steps 3-5 100,000 times
7. The empirical p-value is given by $\frac{1}{100{,}000}\sum_{i=1}^{100{,}000} I_i$.

6.6.3 Considering conditional pricing models

- Let us now demonstrate the implementation of the HJ measure when the pricing kernel takes the form

$$\xi_{t+1} = (\Theta_0' \boldsymbol{X_t}) + (\Theta_1' \boldsymbol{X_t}) f_{t+1}^1 + \ldots + (\Theta_K' \boldsymbol{X_t}) f_{t+1}^K, \quad (329)$$

where $\boldsymbol{X_t}' = [1, \boldsymbol{Z}_t']$ and $\boldsymbol{Z}_t$ is an $M \times 1$ vector of information variables and f_{t+1}^k $(k = [1, 2, \ldots, K])$ denotes a proxy for marginal utility growth, or a macroeconomy factor.

- As noted earlier the first order condition implies that

$$E\left[R_{t+1}\left((\Theta_0' \boldsymbol{X_t}) + (\Theta_1' \boldsymbol{X_t}) f_{t+1}^1 + \ldots + (\Theta_K' \boldsymbol{X_t}) f_{t+1}^K\right) | \boldsymbol{Z_t}\right] = \iota_N. \quad (330)$$

- We collect the vector of errors,

$$\begin{aligned} \boldsymbol{f}_{t+1} &= R_{t+1}\left((\Theta_0' \boldsymbol{X_t}) + (\Theta_1' \boldsymbol{X_t}) f_{t+1}^1 + \ldots + (\Theta_K' \boldsymbol{X_t}) f_{t+1}^K\right) - \iota_N, \\ &= R_{t+1} Y_{t+1}' \boldsymbol{\Theta} - \iota_N, \end{aligned} \quad (331)$$

where

$$\Theta' = [\Theta_0', \Theta_1', \ldots, \Theta_K'], \quad (332)$$

$$Y_{t+1}' = [\boldsymbol{X_t}', \boldsymbol{X_t}' f_{t+1}^1, \ldots, \boldsymbol{X_t}' f_{t+1}^K]. \quad (333)$$

- Overall there are $(K+1)(M+1)$ parameters to estimate.

- Equation (330) implies

$$E[\boldsymbol{f}_{t+1}|\boldsymbol{Z_t}] = 0, \tag{334}$$

and therefore

$$E\left[(\boldsymbol{f}_{t+1} \otimes \boldsymbol{Z_t})\,|\boldsymbol{Z_t}\right] = 0. \tag{335}$$

- This forms a set of $N \times (M+1)$ moment conditions given by the compact notation

$$g_T(\Theta) = \frac{1}{T}\sum_{t=0}^{T-1}\left[\boldsymbol{f_{t+1}} \otimes \boldsymbol{X_t}\right]. \tag{336}$$

- To estimate and test the model we minimize the quadratic form

$$\delta^2 = g_T(\Theta)' G_T^{-1} g_T(\Theta), \tag{337}$$

with D and G being estimated by

$$D_T = \frac{1}{T}\sum_{t=0}^{T-1}\left[(R_{t+1} \otimes \boldsymbol{X_t})\,Y'_{t+1}\right], \tag{338}$$

$$G_T = \frac{1}{T}\sum_{t=0}^{T-1}\left[(\boldsymbol{R_{t+1}} \otimes \boldsymbol{X_t})(\boldsymbol{R_{t+1}} \otimes \boldsymbol{X_t})'\right]. \tag{339}$$

- Finally, the sample analog of (336) is

$$g_T(\hat{\Theta}) = D_T\hat{\Theta} - \left(\iota_N \otimes \bar{\boldsymbol{X}}\right), \tag{340}$$

where

$$\hat{\Theta} = (D_T' G_T^{-1} D_T)^{-1} D_T' G_T^{-1}\left(\iota_N \otimes \bar{\boldsymbol{X}}\right), \tag{341}$$

$$\bar{\boldsymbol{X}} = \frac{1}{T}\sum_{t=0}^{T-1}\boldsymbol{X_t}. \tag{342}$$

6.6.4 HJ distance measure vs. the standard GMM

Both the GMM and HJ distance measure are cross sectional tests of asset pricing models.

1. Since the distance measure is formed using a weighting matrix that is invariate across competing SDF candidates it can be used to compare the performance of nested and nonnested asset pricing models.
2. In the standard GMM the optimal weighting matrix S^{-1} varies across competing specifications.
3. Therefore, the standard GMM cannot be used for comparing misspecification across competing models.
4. The HJ distance measure avoids the pitfall embedded in the standard GMM of favoring pricing models that produce volatile pricing errors.
5. In the HJ distance measure the weighting matrix is not a function of the parameters, which may result in a more stable estimation procedure.
6. On the other hand, the optimal GMM provides the most efficient estimate among estimates that use linear combinations of pricing errors as moments, in the sense that the estimated parameters have the smallest asy. covariance.

6.6.5 Some additional insights about the HJ distance measure

- It would be useful to provide details on the derivation of (312) to understand the intuition behind the HJ model misspecification measure.

- In particular, let $\mathcal{M}$ denote the set of all admissible discount factors, and let ξ to be the candidate discount factor entertained by the researcher.
- HJ note that when the pricing model is false there is a strictly positive distance between $\mathcal{M}$ and ξ.
- HJ attempt to minimize the distance $||\xi - m||$ where $m \in \mathcal{M}$ subject to the first order condition $E(mR) = \iota_N$.
- Notice that the distance $||x||$ is the usual norm $\sqrt{E(x^2)}$. So if $E(x) = 0$ then $||x||$ is simply the standard deviation of x.
- The Lagrangian of this minimization programming is given by

$$\mathcal{L} = \min_{m \in \mathcal{M}} \left\{ E(\xi - m)^2 + 2\lambda'[E(mR) - \iota_N] \right\}. \tag{343}$$

Let $\tilde{m}$ and $\tilde{\lambda}$ be the solution to (343). HJ solve (343) to find

$$\xi - \tilde{m} = \tilde{\lambda}' R, \tag{344}$$

where

$$\tilde{\lambda} = E(R'R)^{-1} E(\xi R - \iota_N). \tag{345}$$

- Thus, the HJ-distance is

$$\delta = \min||\xi - m|| = ||\xi - \tilde{m}|| = ||\tilde{\lambda}' R|| = \left[\tilde{\lambda}' E(R'R) \tilde{\lambda} \right]^{1/2}. \tag{346}$$

- Substituting (345) into (346) gives the HJ distance displayed in (312)

$$\sqrt{E(\xi R - \iota_N)' \left[E(R'R) \right]^{-1} E(\xi R - \iota_N)}, \tag{347}$$

and the result follows.

- Another issue we have not yet resolved is why the HJ distance measure is the maximum pricing error for a set of portfolios with the norm of the portfolio return equal to one.
- To show that this is indeed the case, let us assume that

$$E(\xi - \tilde{m}) = E(\tilde{\lambda}'R) = 0. \tag{348}$$

  That is to say that both the true and approximate models give the same riskfree rate, i.e., $\frac{1}{E(\xi)} = \frac{1}{E(\tilde{m})}$.
- So the focus is on the models' ability to correct for risk.
- Under this assumption the
- HJ distance measure is nothing else than

$$\delta = \sigma(\xi - \tilde{m}). \tag{349}$$

- Focusing on the true pricing kernel the $N \times 1$ vector of expected return is given by

$$E(R) = R_f - R_f cov(\tilde{m}, R). \tag{350}$$

- And the expected return on a portfolio $x'R$ is given by

$$E(x'R) = x'R_f - R_f cov(\tilde{m}, x'R). \tag{351}$$

- Similarly, based on the approximate pricing kernel, the expected return on a portfolio $x'R$ is given by

$$E^{\xi}(x'R) = x'R_f - R_f cov(\xi, x'R). \tag{352}$$

- The mispricing in expected return is given by

$$|E(x'R) - E^{\xi}(x'R)| = |R_f cov(\xi - \tilde{m}, x'R)|. \tag{353}$$

- Dividing both sides of (353) by $\sigma(x'R)$ yields the mispricing per unity of standard deviation

$$\left|\frac{E(x'R) - E^{\xi}(x'R)}{\sigma(x'R)}\right| = \left|\frac{R_f cov(\xi - \tilde{m}, x'R)}{\sigma(x'R)}\right|. \quad (354)$$

- One can immediately recognize that

$$\left|\frac{E(x'R) - E^{\xi}(x'R)}{\sigma(x'R)}\right| \leq R_f \sigma(\xi - \tilde{m}), \quad (355)$$

or the maximum possible pricing error expressed as expected return error per unit of standard deviation is given by the product of the riskfree rate and the HJ distance measure.

- The maximally mispriced portfolio, $x^{*\prime}R$, has a unit norm and it is the one for which

$$corr(x^{*\prime}R, \xi - \tilde{m}) = 1. \quad (356)$$

- That is, the worst priced portfolio is perfectly correlated with the difference between the true and the false discount factor.
- We conjecture that this portfolio is given by

$$x^* = \tilde{\lambda}/\delta. \quad (357)$$

- Check: First notice that $x^{*\prime}R = \tilde{\lambda}'R/\delta$ and that $\xi - \tilde{m} = \tilde{\lambda}'R$. The unit correlation therefore follows.
- Moreover, the norm of the portfolio is given by

$$1/\delta\sqrt{\tilde{\lambda}'E(RR')\tilde{\lambda}} = 1. \quad (358)$$

- Substituting x^* into the left hand side of (354) and recognizing that $E(\tilde{\lambda}'R)$=0 (see (348)) give

$$\frac{E^{\xi}(\tilde{\lambda}'R)}{\sigma(\tilde{\lambda}'R)} = R_f \delta. \quad (359)$$

- So the maximum absolute pricing error per unit of std. is equal to the riskfree rate times the HJ distance measure.
- As pointed out by Campbell and Cochrane (2000) the advantage and disadvantage of the HJ measure is that it depends only on the model, not on the set of test portfolios.
- Let us demonstrate the advantage.
- Several studies (e.g., Roll and Ross (1994) and Kandel and Stambaugh (1995)) have shown that pricing errors of an approximate model can depend dramatically on the test portfolio examined.
- The HJ procedure eliminates this dependence by evaluating the pricing error of the worst possible portfolio. The search for that portfolio extends over all possible contingent claims.

7 Stock Return Predictability

7.1 Predictability based on observable macro variables

- A question of long-standing interest to both academics and practitioners is whether returns on risky assets are predictable.
- Fama and Schwert (1977), Keim and Stambaugh (1986), Fama and French (1989), and Lettau and Ludvigson (2001), among others, identify ex ante observable variables that predict future returns on stocks and bonds.
- The evidence on predictability is typically based upon the *joint* system

$$r_t = a + \beta' z_{t-1} + u_t, \tag{360}$$
$$z_t = c + \rho z_{t-1} + v_t. \tag{361}$$

- Statistically, predictability means that at least one of the β coefficients is significant at conventional levels.
- Economically, predictability means that you can properly time the market, switching between an equity fund and a money market fund.

7.2 Is the evidence on predictability robust?

- Predictability based on macro variables is still a research controversy:
  - Asset pricing theories do not specify which ex ante variables predict asset returns;
  - Recent work by Menzly, Santos, and Veronesi (2006) does make some progress but the variables idnetified are those we already know from empirical research.
  - Statistical biases in slope coefficients of a predictive regression;
  - Potential data mining in choosing the macro variables;
  - Poor out-of-sample performance of predictive regressions;
- Schwert (2003) shows that the value and size effects as well as time series predictability tend to attenuate and even disappear after their discovery.
- Indeed, the power of macro variables to predict the equity premium substantially deteriorates during the post-discovery period.

7.2.1 Finite sample bias in the slope coefficient

- Recall, we work with (assuming one predictor for simplicity)

$$r_t = a + \beta z_{t-1} + u_t, \tag{362}$$
$$z_t = c + \rho z_{t-1} + v_t. \tag{363}$$

- Now, let σ_v^2 denote the variance of v_t, and let σ_{uv} denote the covariance between u_t and v_t.

- We know from Kendall (1954) that the OLS estimate of the persistence parameter ρ is biased, and that the bias is $-1(1+3\rho)/T$.

- Stambaugh (1999) shows that under the normality assumption, the finite sample bias in $\hat{\beta}$, the slope coefficient in a predictive regression, is

$$Bias = \mathbb{E}\left(\hat{\beta} - \beta\right) = -\frac{\sigma_{uv}}{\sigma_v^2}\left(\frac{1+3\rho}{T}\right).$$

- The bias can easily be derived.

- Note that the OLS estimates of β and ρ are

$$\hat{\beta} = (X'X)^{-1}X'R = \beta + (X'X)^{-1}X'U, \quad (364)$$

$$\hat{\rho} = (X'X)^{-1}X'Z = \rho + (X'X)^{-1}X'V, \quad (365)$$

where $R = [r_1, r_2, \ldots, r_T]'$, $Z = [z_1, z_2, \ldots, z_T]'$, $U = [u_1, u_2, \ldots, u_T]'$, $V = [v_1, v_2, \ldots, v_T]'$, , $X = [\iota_T, Z_{-1}]$, ι_T is a T-dimension vector of ones, and $Z_{-1} = [z_0, z_1, \ldots, z_{T-1}]'$.

- Also note that u_t can be decomposed into two orthogonal components

$$u_t = \frac{\sigma_{uv}}{\sigma_v^2}v_t + e_t, \quad (366)$$

where e_t is uncorrelated with z_{t-1} and v_t.

- Hence, the predictive regression slope can be rewritten as

$$\hat{\beta} = \beta + \frac{\sigma_{uv}}{\sigma_v^2}(\hat{\rho} - \rho) + (X'X)^{-1}X'E, \quad (367)$$

where $E = [e_1, e_2, \ldots, e_T]'$.

- Amihud and Hurvich (2004, 2009) advocate a nice approach to deal with statistical inference in the presence of a small sample bias.

7.2.2 Potential data mining

- Repeated visits of the same database lead to a problem that statisticians refer to as data mining (also model overfitting or data snooping).
- It reflects the tendency to discover spurious relationships by applying tests inspired by evidence coming up from prior visits to the same database.
- Merton (1987) and Lo and MacKinlay (1990), among others, discuss the problems of overfitting data in tests of financial models.
- In the context of predictability, data mining has been investigated by Foster, Smith, and Whaley (FSW - 1997).

7.2.3 Foster, Smith, and Whaley (1997)

- FSW adjust the test size for potential overfitting, using theoretical approximations as well as simulation studies.
- They assume that
  1. M potential regressors are available.
  2. All possible regression combinations are tried.
  3. Only $m < M$ regressors with the highest R^2 are reported.

  Their evidence shows:

1. Inference about predictability could be erroneous when potential specification search is not accounted for in the test statistic.
2. Using other industry, size, or country data as a control to guard against variable-selection biases can be misleading.

7.2.4 The poor out-of-sample performance of predictive regressions

- A good reference here is Bossaerts and Hillion (1999) and a follow up work by Goyal and Welch (2006).
- The adjustment of the test size merely help in correctly rejecting the null of no relationship.
- It would, however, provide little information if the empiricist is asked to discriminate between competing models under the alternative of existing relation.
- FSW propose examining predictability using model selection criteria.
  - Suppose there are M potential predictors.
  - There are 2^M competing specifications.
  - Select one winning specification based on the adjusted R^2, the AIC, the SIC, and other criteria.
- The selected model (regardless of the criterion used) always retains predictors – not a big surprise – indicating in-sample predictability.
- Implicitly you assign a $\frac{1}{2^M}$ probability that the IID model is correct - so you are biassed in favor of detecting predictability.

- The out-of-sample performance of the selected model is always a disaster.
- The Bayesian approach of model averaging improves out-of-sample performance. Coming soon!
- We will also study asset allocation with predictability.
- Before we move on let us note at least two papers responding to the apparently nonexistent out of sample predictability of the equity premium
  - Cochrane (2008) points out that we should detect predictability either in returns or dividends.
  - Cambpell and Thompson (2008) document predictability after restricting the equity premium to be positive.
  - Rapach, Strauss, and Zhou (2010) combine model forecast, similar to the Bayesian Model Averaging concept, but using equal weights.

8 Bayesian econometrics and asset pricing

8.1 Overview

- I would highly recommend using the textbook of Zellner (1971) along with the notes below.
- It all starts with the well known Bayes Theorem that deals with conditional probabilities

$$P(A|B) = \frac{P(A,B)}{P(B)}, \tag{368}$$

$$= \frac{P(B|A)P(A)}{P(B)} \tag{369}$$

  where A and B are events and $P(A|B)$ is a conditional probability that event A occurs based on the realization of B.

- Let us take Bayes Theorem into finance.
- In particular, consider the time series predictive regression

$$r_t = a + bz_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \tag{370}$$

  where r_t is excess stock return realized at time t and z_{t-1} is a variable observed at time $t-1$.

- Let us denote the set of parameters by Θ where $\Theta = [a, b, \sigma^2]'$.
- For a frequentist econometrician Θ is fixed yet unknown.
- The frequentist estimates Θ using parametric (ML) or non parametric (GMM) methods.
- Then the ML/GMM estimate $\hat{\Theta}$ is asymptotically normally distributed, as discussed earlier.

- The Bayesian philosophy is quite different.
- Θ is stochastic!
- You can learn about Θ based on its posterior distribution.
- That is, given prior views about Θ as well as evidence coming up from data - the posterior summarizes everything we know about Θ.
- For simplicity, let us assume the predictor is nonstochastic and let $r = [r_1, r_2, \ldots, r_T]'$ and $x = [x_0, x_1, \ldots x_{T-1}]'$ such that $x_t = [1, z_t']'$.
- Then the Bayes Theorem becomes

$$P(\Theta|r) = \frac{P(r|\Theta)p(\Theta)}{P(r)} \tag{371}$$

where $P(\Theta|r)$ is the posterior density of Θ - the density of interest, $P(r|\Theta)$ is the likelihood function, and $P(\Theta)$ is the prior distribution of Θ.

- Notice that $P(r)$ does not tell us anything about Θ. So we can write

$$P(\Theta|r) \propto P(r|\Theta)p(\Theta). \tag{372}$$

- Of course, since a probability distribution function integrates to unity it follows that

$$P(\Theta|r) = \frac{P(r|\Theta)p(\Theta)}{\int_\Theta P(r|\Theta)p(\Theta)d\Theta}. \tag{373}$$

- The posterior density could have an analytic reduced form expression if prior beliefs are of the conjugate form or if the prior density is diffuse or improper, as illustrated below.

- But there are many cases in which the exact form of the posterior is unknown.
- Then we can often employ very powerful MCMC techniques to simulate the posterior density.
- That is one major advantage of Bayesian econometrics.
- Indeed, there are many applications in finance (mostly in asset pricing but also in corporate) that use MCMC methods including Gibbs Sampling, Metropolis Hastings, and Importance Sampling.
- Other advantages to using Bayesian econometrics include
  - Eliciting economically motivated prior beliefs
  - Incorporating model uncertainty
  - Accounting for estimation risk
- We will discuss all these issues in the subsections that follow.

8.2 The Case of Conjugate Priors

- Let's start with the tractable case of conjugate priors.
- Consider the multivariate form of the predictive regression (370)

$$R = XB + U, \tag{374}$$

where

$$vec(U) \sim N(0, \Sigma \otimes I_T). \tag{375}$$

Moreover, R is a $T \times N$ matrix of excess returns, X is a $T \times (m+1)$ matrix with m being the number of predictors, and U is a $T \times N$ matrix of residuals.

- The priors for B and Σ are the normal inverted Wishart

$$P(b|\Sigma) \propto |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(b-b_0)'\left[\Sigma^{-1}\otimes\Psi_0\right](b-b_0)\right), \quad (376)$$

$$P(\Sigma) \propto |\Sigma|^{-\frac{\nu_0+N+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[S_0\Sigma^{-1}\right]\right), \quad (377)$$

where

$$b = vec(B) \quad (378)$$

and b_0, Ψ_0, and S_0 are prior parameters to be specified by the researcher.

- The likelihood function of normally distributed data constituting the actual sample obeys the form

$$P(R|B,\Sigma,X) \propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[(R-XB)'(R-XB)\right]\Sigma^{-1}\right), \quad (379)$$

which can be rewritten in a very convenient form

$$P(R|B,\Sigma,X) \propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[\hat{S}+(B-\hat{B})'X'X(B-\hat{B})\right]\Sigma^{-1}\right) \quad (380)$$

where

$$\hat{S} = (R-X\hat{B})'(R-X\hat{B}), \quad (381)$$

$$\hat{B} = (X'X)^{-1}X'R. \quad (382)$$

- An equivalent representation for the likelihood function (380) is given by

$$P(R|b,\Sigma,X) \propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}(b-\hat{b})'\left[\Sigma^{-1}\otimes(X'X)\right](b-\hat{b})\right), \times \exp\left(-\frac{1}{2}\mathrm{tr}\left[\hat{S}\Sigma^{-1}\right]\right), \quad (383)$$

where

$$\hat{b} = \mathrm{vec}(\hat{B}). \quad (384)$$

- Combining the likelihood (383) with the prior and completing the square on b yield

$$P(b|\Sigma,R,X) \propto |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(b-\tilde{b})'\left[\Sigma^{-1}\otimes F\right](b-\tilde{b})\right), \quad (385)$$

$$P(\Sigma|R,X) \propto |\Sigma|^{-\frac{\nu+N+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[\tilde{S}\Sigma^{-1}\right]\right), \quad (386)$$

where

$$
\begin{aligned}
F &= \Psi_0 + X'X, && (387)\\
\tilde{b} &= \text{vec}(\tilde{B}), && (388)\\
\tilde{B} &= F^{-1}X'X\hat{B} + F^{-1}\Psi_0 B_0, && (389)\\
\tilde{S} &= \hat{S} + S_0 + \hat{B}'X'X\hat{B} + B_0'\Psi_0 B_0 \\
&- \tilde{B}'F\tilde{B}, && (390)\\
\nu &= T_0 + T. && (391)
\end{aligned}
$$

- So the posterior for B is normal and for Σ is inverted Wishart.
- That is the conjugate prior idea - the prior and posterior have the same distributions but with different parameters.
- Interestingly, $\tilde{B}$ is a weighted average of B_0 and $\hat{B}$. That is,

$$\tilde{B} = WB_0 + (I - W)\hat{B}, \tag{392}$$

where $W = I - F^{-1}X'X$.

8.3 What if the posterior does not obey a reduced form expression?

- Let us demonstrate the Gibbs Sampling - an appealing Bayesian numerical method - in the context of measuring pricing errors.
- The analysis below is based on "Measuring the Pricing Error of the Arbitrage Pricing Theory" by Geweke and Zhou (RFS 1996) who test the pricing abilities of the APT (Ross (1976)).
- The basic APT model assumes that returns on N risky portfolios are related to K pervasive unknown factors

- The relation is described by the K factor model:

$$r_t = \mu + \beta f_t + \epsilon_t, \tag{393}$$

where r_t is returns (not excess returns) on N assets and f_t is a set of K factor innovations (factors are not pre-specified).

- Specifically,

$$\begin{aligned} E\{f_t\} &= 0, \\ E\{f_t f_t'\} &= I_K, \\ E\{\epsilon_t | f_t\} &= 0, \\ E\{\epsilon_t \epsilon_t' | f_t\} &= \Sigma = diag(\sigma_1^2, \ldots, \sigma_N^2), \\ \beta &= [\beta_1, \ldots, \beta_K]. \end{aligned}$$

- Moreover, under exact APT, the μ vector satisfies the restriction

$$\mu = \lambda_0 + \beta_1 \lambda_1 + \ldots, + \beta_K \lambda_K. \tag{394}$$

- The original APT model is about an approximated relation.

- The objective throughout is to explore a measure that summarizes the deviation from exact pricing.

- That measure is denoted by $\mathcal{Q}^2$ and is given by

$$\mathcal{Q}^2 = \frac{1}{N} \mu' \left[I_N - \beta^* (\beta'^* \beta^*)^{-1} \beta^{*\prime} \right] \mu, \tag{395}$$

where

$$\beta^* = [I_N, \beta]. \tag{396}$$

- Recovering the sampling distribution of $\mathcal{Q}^2$ is hopeless.

- However, using the Gibbs sampling technique, we can simulate the posterior distribution of $\mathcal{Q}^2$.

- Specifically, we assume that observed returns and latent factors are jointly normally distributed.

- That is, we have

$$\begin{bmatrix} f_t \\ r_t \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ \mu \end{pmatrix}, \begin{pmatrix} I_K & \beta' \\ \beta & \beta\beta' + \Sigma \end{pmatrix}\right]. \tag{397}$$

- Here are some additional notations:
  - Data: $R = \left[r_1', \ldots, r_T'\right]'$
  - Parameters: $\Theta = \left[\mu', vec(\beta)', vech(\Sigma)'\right]'$
  - Latent variables: $f = \left[f_1', \ldots, f_T'\right]'$

- To evaluate the pricing error we need draws from the posterior distribution $P(\Theta|R)$.

- We draw from the joint posterior in a slightly different manner than that suggested in the paper.

- First, we employ a multivariate regression setting. Moreover, the well-known identification problem is not accounted for to simplify the analysis.

- The prior is improper

$$P_0(\Theta) \propto |\Sigma|^{-\frac{1}{2}} = (\sigma_1 \ldots \sigma_N)^{-1}. \tag{398}$$

- Reexpressing the arbitrage pricing equation, we obtain:

$$r_t' = F_t'B' + \epsilon_t', \tag{399}$$

where

$$F_t' = [1, f_t'], \tag{400}$$

$$B = [\mu, \beta]. \tag{401}$$

- Rewriting the system in a matrix notation, we get

$$R = FB' + E. \tag{402}$$

- Why do we need to use the Gibbs sampling technique?
- Because the likelihood function $P(R|\Theta)$ (and therefore the posterior density) cannot be expressed analytically.
- However, $P(R|\Theta, F)$ does obey an analytical form:

$$P(R|\Theta, F) \propto |\Sigma|^{-\frac{T}{2}} \exp\Big[-\frac{1}{2}\mathrm{tr}\left\{[R - FB']'[R - FB'\Big]\Sigma^{-1}\right\}. \tag{403}$$

- Therefore, we can compute the full conditional posterior densities:

$$P(B|\Sigma, F, R), \tag{404}$$
$$P(\Sigma|B, F, R), \tag{405}$$
$$P(F|B, \Sigma, R). \tag{406}$$

- The Gibbs sampling chain is formed as follows:
  1. Specify starting values $B^{(0)}$, $\Sigma^{(0)}$, and $F^{(0)}$ and set $i = 1$.
  2. Draw from the full conditional distributions:
     - Draw $B^{(i)}$ from $P(B|\Sigma^{(i-1)}, F^{(i-1)}, R)$
     - Draw $\Sigma^{(i)}$ from $P(\Sigma|B^{(i)}, F^{(i-1)}, R)$
     - Draw $F^{(i)}$ from $P(\Sigma|B^{(i)}, \Sigma^{(i)}, R)$
  3. Set $i = i + 1$ and go to step 2.
- After m iterations the sample $B^{(m)}, \Sigma^{(m)}, F^{(m)}$ is obtained.

- Under mild regularity conditions (see, for example, Tierney ,1994), $(B^{(m)}, \Sigma^{(m)}, F^{(m)})$ converges in distribution to the relevant marginal and joint distributions:

$$\begin{aligned} P(B^{(m)}|R) &\rightarrow P(B|R), & (407)\\ P(\Sigma^{(m)}|R) &\rightarrow P(\Sigma|R), & (408)\\ P(F^{(m)}|R) &\rightarrow P(F|R), & (409)\\ P(B^{(m)}, \Sigma^{(m)}, F^{(m)}|R) &\rightarrow P(B, \Sigma, F|R). & (410) \end{aligned}$$

- For m large enough, the G values

$$(B^{(g)}, \Sigma^{(g)}, F^{(g)})_{g=m+1}^{m+G}$$

are a sample from the joint posterior.

- A remaining task is to find the full conditional posterior densities.

- Note:

$$P(B|\Sigma, F, R) \propto \exp\left\{-\frac{1}{2}\text{tr}[R - FB']'[R - FB']\Sigma^{-1}\right\}. \tag{411}$$

- Let $b = \text{vec}(B')$. Then

$$P(b|\Sigma, F, R) \propto \exp\left\{-\frac{1}{2}[b - \hat{b}]'\left(\Sigma^{-1} \otimes (F'F)\right)[b - \hat{b}]\right\}, \tag{412}$$

where

$$\hat{b} = \text{vec}[(F'F)^{-1}F'R]. \tag{413}$$

- Therefore,

$$b|\Sigma, F, R \sim N\left(\hat{b}, \Sigma \otimes (F'F)^{-1}\right). \tag{414}$$

Also note:

$$P(\sigma_i|B,F,R) \propto \sigma_i^{-(T+1)} \exp\left(-\frac{TS_i^2}{2\sigma_i^2}\right), \tag{415}$$

where TS_i^2 is the i-th diagonal element of the $N \times N$ matrix

$$[R - FB']'[R - FB'], \tag{416}$$

suggesting that:

$$\frac{TS_i^2}{\sigma_i^2} \sim \chi^2(T). \tag{417}$$

- Finally,

$$f_t|\mu,\beta,\Sigma,r_t \sim N(M_t,H_t), \tag{418}$$

where

$$M_t = \beta'(\beta\beta' + \Sigma)^{-1}(r_t - \mu), \tag{419}$$

$$H_t = I_K - \beta'(\beta\beta' + \Sigma)^{-1}\beta. \tag{420}$$

8.4 Incorporating model uncertainty - Avramov (2002) - Bayesian Model Averaging

- We noted earlier that the sample evidence about return predictability was quite disappointing.
- So should we dismiss predictability altogether due to the absence of out of sample predictability? Not so fast!
- Avramov (2002) concludes that Bossaerts and Hillion fail to discover out-of-sample predictability because they ignore model uncertainty.
- Goyal and Welch (2006) do not account for model uncertainty either.

- Indeed, in the heart of the model selection approach, one uses a specific criterion to select a single model and then operates as if the model is correct with a unit probability.
- What about the other $2^M - 1$ competing specifications?
- Avramov proposes examining the sample evidence on predictability using Bayesian Model Averaging (BMA) that considers *all* models.
- Specifically, BMA computes posterior probabilities for all 2^M models and uses the probabilities as weights on the individual models to obtain a composite weighted model.
- Statistical and economic inferences on predictability are made based on the weighted model.
- We can also use the "Bayesian t ratio" to examine predictability
- Let $p(\mathcal{M}_j)$ be the posterior probability of model j, let β_j be the vector (of size M) of the slope coefficients, and let $Var(\beta_j)$ be the $(M \times M)$ variance covariance matrix of the errors in estimating β.
- Recall, the 'usual' t ratio is obtained by dividing β_j by its standard error.
- Let us now compute the t-ratio when model uncertainty is accounted for.
- The numerator would be

$$\beta = \sum_{j=1}^{2^M} p(\mathcal{M}_j)\beta_j.$$

- The denominator would be the square root of the diagonal elements in

$$Var(\beta) = \sum_{j=1}^{2^M} p(\mathcal{M}_j) \left[Var(\beta_j) + (\beta_j - \beta)(\beta_j - \beta)' \right].$$

- Another metric: cumulative probability
- The idea here is to compute the cumulative probability for every predictor.
- Let A be a $2^M \times M$ matrix representing all forecasting models by zeros (exclusion of predictors) and ones (inclusion).
- Let $\mathcal{P}$ be the $2^M \times 1$ vector of posterior probabilities.
- The cumulative probability of each predictor is given by $A'\mathcal{P}$.

The empirical evidence: Table 2 in Avramov (2002)

	Predictive Variables													
	Div	BM	EY	WML	Def	Tbill	RET	DEF	TERM	Jan	Inf	SMB	HML	Term
Portfolio:														
SL	0.20	0.08	0.38	0.02	0.14	0.28	0.48	0.04	0.12	0.21	0.31	0.16	0.05	0.08
	0	0	**1**	0	0	0	0	0	0	0	**1**	0	0	0
SM	0.12	0.06	0.16	0.02	0.10	0.09	0.40	0.06	0.54	0.77	0.15	0.12	0.02	0.19
	0	0	0	0	0	0	0	0	**1**	**1**	0	0	0	0
SH	0.06	0.05	0.07	0.03	0.06	0.04	0.49	0.03	0.35	1.00	0.06	0.08	0.02	0.22
	0	0	0	0	0	0	**1**	0	0	**1**	0	0	0	0
BL	0.12	0.05	0.14	0.04	0.14	0.13	0.05	0.07	0.20	0.03	0.69	0.05	0.05	0.15
	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0
BM	0.15	0.06	0.15	0.03	0.20	0.34	0.03	0.09	0.54	0.07	0.23	0.04	0.03	0.25
	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0
BH	0.07	0.06	0.06	0.03	0.09	0.09	0.02	0.03	0.17	0.92	0.21	0.02	0.02	0.47
	0	0	0	0	0	0	0	0	0	**1**	0	0	0	**1**

Out-of-sample performance: Table 6 in Avramov (2002)

	$T_0 = 50$	$T_0 = 100$	$T_0 = 25$	All	iid	Adj R^2	AIC	SIC	FIC	PIC
					The Rolling Scheme – Monthly Sample					
MPE	0.0006	0.0007	0.0003	-0.0006	0.0007	-0.0002	0.0000	-0.0023	0.0001	-0.0003
t-statistic	0.4225	0.4944	0.2368	-0.3874	0.5126	-0.1551	0.0176	-1.5588	0.0365	-0.2117
Efficiency	-0.0563	-0.0287	-0.2335	-0.7874	-0.4371	-0.7642	-0.7919	-0.9454	-0.8709	-0.7926
t-statistic	-0.1788	-0.0863	-0.8557	-7.8065	-1.3761	-7.4691	-6.9512	-7.9715	-7.4193	-7.2763
Serial Correlation	0.0397	0.0499	0.0323	-0.0284	0.0684	-0.0043	-0.0051	0.0274	-0.0185	-0.0269
t-statistic	0.6676	0.8326	0.5494	-0.4856	1.1288	-0.0738	-0.0895	0.5024	-0.3167	-0.4659
MSE	0.2137	0.2141	0.2139	0.2333	0.2155	0.2309	0.2298	0.2319	0.2339	0.2312
					The Recursive Scheme – Monthly Sample					
MPE	-0.0003	-0.0004	-0.0003	0.0005	-0.0010	0.0007	0.0012	0.0013	0.0028	0.0020
t-statistic	-0.1049	-0.1659	-0.1421	0.1847	-0.3924	0.3018	0.5103	0.5099	1.1455	0.8152
Efficiency	-0.2357	-0.1047	-0.4018	-0.6708	-0.4500	-0.5959	-0.5804	-0.7953	-0.7319	-0.7300
t-statistic	-0.6675	-0.2572	-1.3455	-3.0407	-0.9504	-2.8158	-2.7292	-3.7994	-3.0175	-2.9242
Serial Correlation	0.0401	0.0489	0.0372	0.0036	0.0706	0.0144	0.0143	0.0417	0.0144	0.0120
t-statistic	0.6728	0.8079	0.6320	0.0655	1.1414	0.2597	0.2572	0.7386	0.2663	0.2194
MSE	0.2133	0.2133	0.2143	0.2231	0.2155	0.2197	0.2189	0.2260	0.2237	0.2239
					MSE's for the Quarterly Sample					
Rolling	0.7546	0.7577	0.7651	0.9333	0.7777	0.9041	0.8629	0.8570	0.9286	0.9347
Recursive	0.7757	0.7678	0.7930	0.8312	0.7781	0.8163	0.8233	0.8952	0.8170	0.8337

9 Bayesian Asset Allocation

- This section is based upon a review paper of Avramov and Zhou (2010) prepared for the Annual Review of Financial Studies.
- We first study asset allocation when stock returns are assumed to be unpredictable or IID.
- We then incorporate potential return predictability.

9.1 Asset allocation when returns are IID

9.1.1 The Mean Variance Framework and Estimation Risk

- Assume there are $N+1$ assets, one of which is riskless and others are risky.
- Denote by r_{ft} and r_t the rates of returns on the riskless asset and the risky assets at time t, respectively.
- Then
$$R_t \equiv r_t - r_{ft} 1_N \tag{421}$$
are excess returns on the N risky assets.
- Assume that the joint distribution of R_t is iid over time, with mean μ and covariance matrix V.
- In the static mean-variance framework an investor at time T chooses his portfolio weights w, so as to maximize his quadratic objective function
$$U(w) = E[R_p] - \frac{\gamma}{2} \mathrm{Var}[R_p] = w'\mu - \frac{\gamma}{2} w'Vw, \tag{422}$$

where $R_p = w'R_{T+1}$ is the future uncertain portfolio return at time $T+1$ and γ is the coefficient of relative risk aversion.

- It is well-known that, when both μ and V are assumed known, the optimal portfolio weights are

$$w^* = \frac{1}{\gamma} V^{-1}\mu, \tag{423}$$

and the maximized expected utility is

$$U(w^*) = \frac{1}{2\gamma}\mu' V^{-1}\mu = \frac{\theta^2}{2\gamma}, \tag{424}$$

where $\theta^2 = \mu' V^{-1}\mu$ is the squared Sharpe ratio of the *ex ante* tangency portfolio of the risky assets.

- This is the well known mean-variance theory pioneered by Markowitz (1952).

- In practice, the problem is that w^* is not computable because μ and V are unknown. As a result, the above mean-variance theory is usually applied in two steps.

- In the first step, the mean and covariance matrix of the asset returns are estimated based on the observed data.

- Given a sample size of T, the standard maximum likelihood estimators are

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t, \tag{425}$$

$$\hat{V} = \frac{1}{T}\sum_{t=1}^{T} (R_t - \hat{\mu})(R_t - \hat{\mu})'. \tag{426}$$

- Then, in the second step, these sample estimates are treated as if they were the true parameters, and are simply plugged into (423) to compute the estimated optimal portfolio weights,

$$\hat{w}^{\text{ML}} = \frac{1}{\gamma}\hat{V}^{-1}\hat{\mu}. \tag{427}$$

- The two-step procedure gives rise to a parameter uncertainty problem because it is the estimated parameters, not the true ones, that are used to compute the optimal portfolio weights. Consequently, the utility associated with the plug-in portfolio weights can be substantially different from $U(w^*)$.

- Denote by θ a vector of all the parameters (both μ and V). Mathematically, the two-step procedure maximizes the expected utility conditional on the estimated parameters, denoted by $\hat{\theta}$, being equal to the true ones,

$$\max_w [U(w) \,|\, \theta = \hat{\theta}], \tag{428}$$

  and the uncertainty or estimation errors are ignored.

- Intuitively, two risky assets with the same estimated expected returns of 10% (assume all other true moments are identical) can be entirely different if the first one has an estimation error of 20%, while the second has 1%. However, both are treated as the same in the above two-step procedure.

- Under the normality assumption, and under the standard diffuse prior,

$$p_0(\mu, V) \propto |V|^{-\frac{N+1}{2}}, \tag{429}$$

  the posterior distribution is (see, e.g., Zellner (1971)),

$$p(\mu, V|\mathbf{\Phi}_T) = p(\mu \,|\, V, \mathbf{\Phi}_T) \times p(V \,|\, \mathbf{\Phi}_T) \tag{430}$$

with

$$p(\mu \,|\, V, \mathbf{\Phi}_T) \propto |V|^{-1/2} \exp\{-\frac{1}{2}\mathrm{tr}[T(\mu - \hat{\mu})(\mu - \hat{\mu})' V^{-1}]\}, \tag{431}$$

$$P(V) \propto |V|^{-\frac{\nu}{2}} \exp\{-\frac{1}{2}\mathrm{tr}\, V^{-1}(T\hat{V})\}, \tag{432}$$

where 'tr' denote the trace of a matrix, R is a $T \times N$ matrix formed by the returns, and $\nu = T + N$.

- The predictive distribution is

$$p(R_{T+1}|\mathbf{\Phi}_T) \propto |V + (R_{T+1} - \hat{\mu})(R_{T+1} - \hat{\mu})'/(T+1)|^{-T/2}, \tag{433}$$

which is a multivariate t-distribution with $T - N$ degrees of freedom.

- While the problem of estimation error is recognized by Markowitz (1952), it is only in the 70s that this problem receives serious attention. Winkler (1973) and Winkle and Barry (1975) are earlier examples of Bayesian studies on predicting prices and portfolio choice.

- Brown (1976, 1978) and Klein and Bawa (1976) lay out independently and clearly the Bayesian predictive density approach, especially Brown (1976) who explains thoroughly the estimation error problem and the associated Bayesian approach. Later, Bawa, Brown, and Klein (1979) provide an excellent review of the early literature.

- Under the diffuse prior, (429), it is known that the Bayesian optimal portfolio weights are

$$\hat{w}^{\mathrm{Bayes}} = \frac{1}{\gamma}\left(\frac{T - N - 2}{T + 1}\right)\hat{V}^{-1}\hat{\mu}. \tag{434}$$

- In contrast with the classical weights $\hat{w}^{\mathrm{ML}}$, the Bayesian holds the portfolio of proportion to $\hat{V}^{-1}\hat{\mu}$, but the proportional coefficient is $(T-N-2)/(T+1)$ instead of 1.
- The coefficient can be substantially smaller when N is large relative to T. Intuitively, the assets are riskier in the Bayesian framework since parameter uncertainty is an additional source of risk and this risk is now factored into the portfolio decision.
- As a result, the position in the risky assets are generally smaller than before.
- However, in the classical framework, $\hat{w}^{\mathrm{ML}}$ is a biased estimator of the true weights since, under the normality assumption,

$$E\hat{w}^{\mathrm{ML}} = \frac{T-N-2}{T}w^* \neq w^*. \tag{435}$$

- Let

$$\tilde{V}^{-1} = \frac{T}{T-N-2}\hat{V}^{-1}, \tag{436}$$

then $\tilde{V}^{-1}$ is an unbiased estimator of V^{-1}.

- The unbias estimator of w^* is

$$\bar{w} = \frac{1}{\gamma}\frac{T-N-2}{T}\hat{V}^{-1}\hat{\mu}, \tag{437}$$

which is a scalar adjustment of $\hat{w}^{\mathrm{ML}}$.

- In contrast with the Bayesian weights, the unbiased one differs from it by a scalar $T/(T+1)$. The difference is independent of N, and is negligible for all practical sample sizes T.
- Hence, parameter uncertainty makes little difference between Bayesian and classical approaches if the diffuse prior is used.

- Therefore, to provide new insights, it is important for a Bayesian to use informative priors, which is a decisive advantage of the Bayesian approach that can incorporate useful information easily into portfolio analysis.

9.1.2 Conjugate Prior

- The conjugate prior, which retains the same class of distributions, is a natural and common information prior on any problem.

- This prior in our context assumes a normal prior for the mean and inverted prior for V,

$$\mu \,|\, V \;\sim\; N(\mu_0, \frac{1}{\tau}V), \tag{438}$$

$$V \;\sim\; IW(V_0, \nu_0), \tag{439}$$

  where μ_0 is the prior mean, and τ is a prior parameter reflecting the prior precision of μ_0, and ν_0 is a similar parameter on V.

- Under this prior, the posterior distribution of μ and V are of the same form as the case for the diffuse prior, except that now the posterior mean of μ is given by

$$\tilde{\mu} = \frac{\tau}{T+\tau}\mu_0 + \frac{T}{T+\tau}\hat{\mu}. \tag{440}$$

- This says that the posterior mean is simply a weighted average of the prior and sample.

- Similarly, V_0 can be updated by

$$\tilde{V} = \frac{T+1}{T(\nu_0+N-1)}\left(V_0 + T\hat{V} + \frac{T\tau}{T+\tau}(\mu_0-\hat{\mu})(\mu_0-\hat{\mu})'\right), \tag{441}$$

which is a weighted average of the prior, sample and deviations of $\hat{\mu}$ from μ_0.

- Frost and Savarino (1986) provide an interesting case of the conjugate prior by assuming the assets are with identical means, variances, and patterned covariances, a priori.
- They find that such a prior improves performance.
- This prior is related the well known $1/N$ rule that invests equally across N assets.

9.1.3 Hyperparameter Prior

- By introducing hyperparameters η and λ, Jorion (1986) uses the following prior on μ,

$$p_0(\mu \,|\, \eta, \lambda) \propto |V|^{-1} \exp\{-\frac{1}{2}(\mu - \eta 1_N)'(\lambda V)^{-1}(\mu - \eta 1_N)\}, \tag{442}$$

in which η and λ govern the prior distribution of μ.

- Using diffuse priors on both η and λ, and integrating them out from a suitable distribution, the predictive distribution of the future asset return can be obtained as usual.
- Based on this, Jorion (1986) obtains

$$w^{\mathrm{PJ}} = \frac{1}{\gamma}(\hat{V}^{\mathrm{PJ}})^{-1}\hat{\mu}^{\mathrm{PJ}}, \tag{443}$$

where

$$\hat{\mu}^{\mathrm{PJ}} = (1 - \hat{v})\hat{\mu} + \hat{v}\hat{\mu}_g 1_N, \tag{444}$$

$$\hat{V}^{\mathrm{PJ}} = \left(1 + \frac{1}{T + \hat{\lambda}}\right)\bar{V} + \frac{\hat{\lambda}}{T(T + 1 + \hat{\lambda})}\frac{1_N 1_N'}{1_N'\bar{V}^{-1}1_N}, \tag{445}$$

$$\hat{v} = \frac{N+2}{(N+2) + T(\hat{\mu} - \hat{\mu}_g 1_N)'\bar{V}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N)}, \quad (446)$$

$$\hat{\lambda} = (N+2)/[(\hat{\mu} - \hat{\mu}_g 1_N)'\bar{V}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N)] \quad (447)$$

with $\bar{V} = T\hat{V}/(T - N - 2)$ an adjusted sample covariance matrix, and $\hat{\mu}_g = 1_N'\bar{V}^{-1}\hat{\mu}/1_N'\bar{V}^{-1}1_N$ the average excess return on the sample global minimum-variance portfolio.

- An alternative motivation of Jorion (1986) portfolio rule is from a shrinkage perspective by considering the Bayes-Stein estimator of the expected return, μ, with

$$\hat{\mu}^{\mathrm{BS}} = (1 - v)\hat{\mu} + v\mu_g 1_N, \quad (448)$$

where $\mu_g 1_N$ is the shrinkage target with $\mu_g = 1_N'V^{-1}\mu/1_N'V^{-1}1_N$, and v is the weight given to the target.

- Jorion (1986) as well as subsequent studies find that w^{PJ} improves w^{ML} substantially, implying that it does so also for the Bayesian strategy under the diffuse prior.

9.1.4 The Black-Litterman Model

- One of the major challenges in using the Markowitz's portfolio rule $\hat{w}^{\mathrm{ML}}$ is that it often implies unusually large long and short positions when no portfolio constraints are imposed, and it takes many zero positions when no-short-sell constraints are imposed.

- Black and Litterman (1992) provide a novel solution to this problem.

- They assume that the investor starts with views of the market, then updates them with his own views via the Bayesian rule.

- In a typical case, the market views are taken as the equilibrium expected returns implied by the CAPM.

- The implied portfolio is the value-weighted index, and this is a reasonable portfolio to start with.

- Then, if the investor has no views different from the market, he will hold the market.

- If he has different views, Black and Litterman (1992) propose a way to combine the investor's view with the market.

- The equilibrium expected excess returns (or the equilibrium risk premium, as often referred in the literature) are the ones implied by an asset pricing model, the CAPM here,

$$\mu^e = \gamma V w_e, \tag{449}$$

  where w_e is the equilibrium portfolio weights or weights of the value-weighted stock index, and γ is the market risk-aversion coefficient (γ) when the utility is interpreted as that of the market representative investor

- Assume that the true expected excess return μ is normally distributed with mean μ^e,

$$\mu = \mu^e + \epsilon^e, \quad \epsilon^e \sim N(0, \tau V), \tag{450}$$

  where ϵ^e is the deviation of μ from μ^e that is normally distributed with zero mean and covariance matrix τV, and τ is a scalar indicating the degree of belief in how close μ is to the equilibrium value.

- In the absence of any views on future stock returns, and in the special case of $\tau = 0$, the investor's portfolio weights must be equal to w_e, weights of the value-weighted index.

- Black and Litterman (1992) also consider views on the relative performance of the stocks that can be represented mathematically by a single vector equation,

$$P\mu = \mu^v + \epsilon^v, \quad \epsilon^v \sim N(0, \Omega), \tag{451}$$

  where P is a $K \times N$ matrix summarizing K views, μ^v is a K-vector summarizing the prior means of the view portfolios, and ϵ^v is the residual vector. The covariance matrix of the residuals, Ω, measures the degree of confidence the investor has in his views.

- Applying the Bayesian rule to the beliefs in both equilibrium relationship and the investor's views, Equations (450) and (451), Black and Litterman [1992] obtain the Bayesian updated expected returns and risks as

$$\bar{\mu}^{\mathrm{BL}} = [(\tau V)^{-1} + P'\Omega^{-1}P]^{-1}[(\tau V)^{-1}\mu^e + P'\Omega^{-1}\mu^v], \tag{452}$$

$$\bar{V}^{\mathrm{BL}} = V + [(\tau V)^{-1} + P'\Omega^{-1}P]^{-1}. \tag{453}$$

- Replacing V by $\hat{V}$ and plugging these two updated estimates into the (427), one obtains the Black and Litterman solution to the portfolio choice problem.

- This is widely known in practice as the Black and Litterman model.

- Note that the Black-Litterman expected return, $\bar{\mu}^{\mathrm{BL}}$, is a weighted average of the equilibrium expected return and the investor's views.

- Intuitively, the less confident the investor is in his views, the closer $\bar{\mu}^{\mathrm{BL}}$ is to the equilibrium value, and so the closer the Black-Litterman portfolio is to w_e.
- This is indeed the case as shown mathematically by He and Litterman (1999).
- Hence, the Black and Litterman model tilts the investor's optimal portfolio away from the market portfolio according to the strength of his views.
- Since the market portfolio is a reasonable one without any extreme positions, any suitably controlled tilt should also yield a portfolio without any extreme positions.
- This is perhaps one of the major reasons that make the BLM popular in practice.
- However, Black and Litterman model is referred as a Bayesian approach by various related studies, but the model is not entirely Bayesian.
- This is because the data-generating process is not spelled out explicitly, nor is the Bayesian predictive density used anywhere.
- Zhou (2009) treats the investors' view as yet another layer of priors, and combines this and the equilibrium prior both with the data-generating process, resulting a formal Bayesian treatment and an extension of the famous Black and Litterman model.
- Let us illustrate how one can use the BLM in practice. The illustrations are based on class notes written by Robert Stambaugh.

- Recall we can rely somehow on the sample estimates of volatilities and correlations but we are uncomfortable about average returns.
- Also recall that if p is a tangency portfolio, than the expected return for each asset in the tangency portfolio must obey

$$
\begin{aligned}
E_i &= R_f + (E_p - R_f)\frac{\sigma_{ip}}{\sigma_p^2} && (454)\\
&= R_f + S_p\frac{\sigma_{ip}}{\sigma_p}. && (455)
\end{aligned}
$$

- The problem is that we don?t know what the tangency portfolio and Sharpe ratio are.
- Let us define some normal portfolio, say x.
- We can use sample estimates to measure the Sharpe ratio of the normal portfolio, call it $\bar{S}$.
- Then we can compute the neutral expected returns

$$E_i^* = R_f + \bar{S}\frac{\sigma_{ip}}{\sigma_p}. \tag{456}$$

9.1.5 Computing Neutral Expected Returns: An Example

- Assume there are two risky assets with sample estimates given by $\hat{\mu_1} = 0.928\%$, $\hat{\mu_2} = 1.132\%, \sigma_1 = 0.0529, \sigma_2 = 0.0668$, and $\rho = 0.9493$.
- Using these sample estimates the sample based tangency portfolio would be -0.0125 and 1.125.
- This is an extreme portfolio.

- The BLM implies something different.
- Let us assume normal weights given by 0.8 and 0.2
- Then the Sharpe ratio of the normal portfolio is 0.12.
- The normal expected returns are given by $\bar{E}_1 = 0.966\%$ and $\bar{E}_2 = 1.111\%$.
- Those normal expected returns are used in the portfolio optimization.

9.1.6 Asset Pricing Prior

- Pástor (2000) and Pástor and Stambaugh (2000) introduce interesting priors that reflect an investor's degree of belief in an asset pricing model.
- To see how this class of priors is formed, assume $R_t = (y_t, x_t)$, where y_t contains the excess returns of m non-benchmark positions and x_t contains the excess returns of K $(= N - m)$ benchmark positions.
- Consider a factor model multivariate regression

$$y_t = \alpha + Bx_t + u_t, \tag{457}$$

where u_t is an $m \times 1$ vector of residuals with zero means and a non-singular covariance matrix $\Sigma = V_{11} - BV_{22}B'$, and α and B are related to μ and V through

$$\alpha = \mu_1 - B\mu_2, \qquad B = V_{12}V_{22}^{-1}, \tag{458}$$

where μ_i and V_{ij} $(i, j = 1, 2)$ are the corresponding partition of μ and V,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \; V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}. \tag{459}$$

- For a factor-based asset pricing model, such as the three-factor model of Fama and French (1993), the restriction is $\alpha = 0$.
- To allow for mispricing uncertainty, Pástor (2000), and Pástor and Stambaugh (2000) specify the prior distribution of α as a normal distribution conditional on Σ,

$$\alpha|\Sigma \sim N\left[0, \sigma_\alpha^2\left(\frac{1}{s_\Sigma^2}\Sigma\right)\right], \tag{460}$$

where s_Σ^2 is a suitable prior estimate for the average diagonal elements of Σ. The above alpha-Sigma link is also explored by MacKinlay and Pástor (2000) in the classical framework.
- The magnitude of σ_α represents an investor's level of uncertainty about the pricing ability of a given model.
- When $\sigma_\alpha = 0$, the investor believes dogmatically in the model and there is no mispricing uncertainty.
- On the other hand, when $\sigma_\alpha = \infty$, the investor believes that the pricing model is entirely useless.
- Although the above prior is motivated from asset pricing consideration, it also has a shrinkage interpretation similar to Equation (448).
- The prior on α implies a prior mean on μ, say μ_0.
- It can be shown that the predictive mean is

$$\mu_p = \tau\mu_0 + (1-\tau)\hat{\mu}, \tag{461}$$

where τ depends upon the sample size and the confidence of belief in the pricing model.

- Pástor (2000), and Pástor and Stambaugh (2000) find that the asset pricing priors make a substantial CER difference in portfolio decisions.

9.1.7 Objective Prior

- Previous priors are placed on μ and V, not on the solution to the problem, the portfolio weights.
- In many applications, supposedly innocuous diffuse priors on some basic model parameters can actually imply rather strong prior convictions about particular economic dimensions of the problem.
- For examples, in the context of testing portfolio efficiency, Kandel, McCulloch, and Stambaugh (1995) find that the diffuse prior in fact implies a strong prior on inefficiency of a given portfolio.
- Tu and Zhou (2009) show that the diffuse prior implies a large prior differences in portfolios weights across assets.
- In short, diffuse priors can be unreasonable in an economic sense in some applications.
- As a result, it is important to use informative priors on the model parameters that can imply reasonable priors on functions of interest.
- Tu and Zhou (2009) advocate a method of constructing priors based on a prior on the solution of an economic objective.

- In maximizing an economic objective, even before the Bayesian investor observes any data, he is likely to have some idea about the range of the solution.
- This will allow him to form a prior on the solution, from which the prior on the parameters can be backed out.
- In other words, to maximizing the mean-variance utility here, an investor may have a prior on the portfolio weights, such as an equal- or value-weighted portfolios of the underlying assets.
- This prior can then be transformed into a prior on μ and V.
- This prior on μ and V implies a reasonable prior on the portfolio weights by construction.
- Because of the way such priors on the primitive parameters are motivated, they are called objective-based priors.
- Formally, for the current portfolio choice problem, the objective-based prior starts from a prior on w,

$$w \sim N(w_0, V_0 V^{-1}/\gamma). \tag{462}$$

where w_0 and V_0 are suitable prior constants with known values, and then back out a prior on μ,

$$\mu \sim N\left[\gamma V w_0, \sigma_\rho^2 \left(\frac{1}{s^2} V\right)\right], \tag{463}$$

where s^2 is the average of the diagonal elements of V.

- The prior on V can be taken as the usual inverted Wishart distribution.

- Using monthly returns on the Fama-French 25 size and book-to-market portfolios and three factors from January 1965 to December 2004, Tu and Zhou (2009) find that the investment performances under the objective-based priors can be significantly different from those under diffuse and asset pricing priors, with differences in terms of annual certainty-equivalent returns greater than 10% in many cases.

9.1.8 Investing in Mutual Funds

- Baks, Metrik, and Wachter (2001) (henceforth BMW) and Pastor and Stambaugh (2002a b) have explored the role of prior information about fund performance in making investment decisions.
- BMW consider a mean variance optimizing investor who is quite skeptical about the ability of a fund manager to pick stocks and time the market.
- They find that even with a high degree of skepticism about fund performance the investor would allocate considerable amounts to actively managed funds.
- Pastor and Stambaugh nicely extend the BMW methodology.
- In BMW as well as past papers studying mutual fund performance, performance is typically defined as the intercept in the regression of the fund?s excess returns on excess return of one or more benchmark assets.
- Pastor and Stambaugh recognize the possibility that the intercept in such regressions could be a mix of fund performance as well as model mispricing.

- In particular, consider the case wherein benchmark assets used to define fund performance are unable to explain the cross section dispersion of passive assets, that is, the sample alpha in the regression of non benchmark passive assets on benchmarks assets is nonzero.
- Then model mispricing emerges in the performance regression.
- Thus, Pastor and Stambaugh formulate prior beliefs on both performance and mispricing.
- Geczy, Stambaugh, and Levin (2005) apply the Pastor Stambaugh methodology to study the cost of investing in socially responsible mutual funds.
- Comparing portfolios of these funds to those constructed from the broader fund universe reveals the cost of imposing the socially responsible investment (SRI) constraint on investors seeking the highest Sharpe ratio.
- This SRI cost depends crucially on the investor's views about asset pricing models and stock-picking skill by fund managers.
- BMW and Pastor and Stambaugh assume that the prior on alpha is independent across funds.
- Jones and Shanken (2002) show that under such an independence assumption, the maximum posterior mean alpha increases without bound as the number of funds increases and "extremely large" estimates could randomly be generated.
- This is true even when fund managers have no skill.
- Thus they propose incorporating prior dependence across funds.

- Then, investors aggregate information across funds to form a general belief about the potential for abnormal performance.
- Each fund's alpha estimate is shrunk toward the aggregate estimate, mitigating extreme views.

9.2 Asset allocation with predictability

Let us extend the asset allocation framework to the case where future returns are potentially predictable by macro-wide variables such as the dividend yield.

9.3 One-period Models

- Consider a one-period optimizing investor who must allocate at time T funds between the value-weighted NYSE index and one-month Treasury bills.
- The investor makes portfolio decisions based on estimating the predictive system

$$r_t = a + b' z_{t-1} + u_t, \tag{464}$$

$$z_t = \theta + \rho z_{t-1} + v_t, \tag{465}$$

  where r_t is the continuously compounded NYSE return in month t in excess of the continuously compounded T-bill rate for that month, z_{t-1} is a vector of M predictive variables observed at the end of month $t-1$, b is a vector of slope coefficients, and u_t is the regression disturbance in month t.
- The evolution of the predictive variables is essentially stochastic.

- Typically a first order vector autoregression is employed to model that evolution.
- The residuals in equations (464) and (465) are assumed to obey the normal distribution.
- In particular, let $\eta_t = [u_t, v_t']'$ then $\eta_t \sim N(0, \Sigma)$ where

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{vu} & \Sigma_v \end{bmatrix}. \tag{466}$$

- The distribution of r_{T+1}, the time $T+1$ NYSE excess return, conditional on data and model parameters is $N(a + b'z_T, \sigma_u^2)$.
- Assuming the inverted Wishart prior distribution for Σ and multivariate normal prior for the intercept and slope coefficients in the predictive system, the Bayesian predictive distribution $P\left(r_{T+1}|\Phi_T\right)$ obeys the Student t density.
- Then, considering a power utility investor with parameter of relative risk aversion denoted by γ the optimization formulation is

$$\omega^* = \arg\max_{\omega} \int_{r_{T+1}} \frac{\left[(1-\omega)\exp(r_f) + \omega\exp(r_f + r_{T+1})\right]^{1-\gamma}}{1-\gamma} P\left(r_{T+1}|\Phi_T\right) dr_{T+1}, \tag{467}$$

  subject to ω being nonnegative.

- It is infeasible to have analytic solution for the optimal portfiolio.
- However, it can easily be solved numerically. In particular, given G independent draws for R_{T+1} from the predictive distribution, the optimal portfolio is found by implementing a constrained optimization code to maximize the quantity

$$\frac{1}{G}\sum_{g=1}^{G} \frac{\left\{(1-\omega)\exp(r_f) + \omega\exp(r_f) + R_{T+1}^{(g)})\right\}^{1-\gamma}}{1-\gamma} \tag{468}$$

subject to ω being nonnegative.

- Kandel and Stambaugh (1996) show that even when the statistical evidence on predictability, as reflected through the R^2 is the regression (464), is weak, the current values of the predictive variables, z_T, can exert a substantial influence on the optimal portfolio.

9.4 Multi-period Models

- Whereas Kandel and Stambaugh (1996) study asset allocation in a single-period framework, Barberis (2000) analyzes multi-period investment decisions, considering both a buy-and-hold investor as well as an investor who dynamically rebalances the optimal stock-bond allocation.

- Implementing long horizon asset allocation in a buy-and-hold setup is quite straightforward.

- In particular, let K denote the investment horizon, and let $R_{T+K} = \sum_{k=1}^{K} r_{T+k}$ be the cumulative (continuously compounded) return over the investment horizon.

- Avramov (2002) shows that the distribution for R_{T+K} conditional on the data (denoted Φ_T) and set of parameters (denoted Θ) is given by

$$R_{T+K}|, \Theta, \Phi_T \sim N(\lambda, \Upsilon), \tag{469}$$

where

$$\lambda = Ka + b' \left[(\rho^K - I_M)(\rho - I_M)^{-1}\right] z_T \tag{470}$$
$$+ \; b' \left[\rho \left(\rho^{K-1} - I_M\right)(\rho - I_M)^{-1} - (K-1) I_M\right](\rho - I_M)^{-1}\theta,$$

$$\Upsilon = K\sigma_u^2 + \sum_{k=1}^{K} \delta(k)\Sigma_v\delta(k)' + \sum_{k=1}^{K} \sigma_{uv}\delta(k)' + \sum_{k=1}^{K} \delta(k)\sigma_{vu}, \tag{471}$$

$$\delta(k) = b' \left[\left(\rho^{k-1} - I_M\right)(\rho - I_M)^{-1}\right]. \tag{472}$$

- Drawing from the Bayesian predictive distribution is done in two steps.
- First, draw the model parameters Θ from their posterior distribution.
- Second, conditional on model parameters, draw R_{T+K} from the normal distribution formulated in (469) - (472).
- The optimal portfolio can then be found using (468) where R_{T+K} replaces R_{T+1} and Kr_f replaces r_f.
- It should be noted that incorporating dynamic rebalancing, intermediate consumption, and learning could establish a non trivial challenge for recovering the optimal portfolio choice.
- Brandt, Goyal, Santa Clara, and Stroud (2005) address the challenge using a tractable simulation based method.
- Notice that the iid set-up corresponds to $b = 0$ in the predictive regression (464), which yields $\lambda_{iid} = Ka$ and $\Upsilon_{iid} = K\sigma_u^2$ in (470) and (471).
- The conditional mean and variance in an iid world increase linearly with the investment horizon.

- Thus, there is no horizon effect when (i) returns are iid and (ii) estimation risk is not accounted for, as indeed shown by Samuelson (1969) and Merton (1969) in an equilibrium framework.
- Incorporating estimation risk, Barberis (2000) shows that the allocation to equity diminishes with the investment horizon, as stocks appear to be riskier in longer horizons.
- Incorporating return predictability and estimation risk, Barberis (2000) shows that investors allocate considerably more heavily to equity the longer their horizon.
- An essential question is what are the benefits of using the Bayesian approach in studying asset allocation with predictability?
- There are at least four important features of the Bayesian approach which we describe below including the ability to account for estimation risk and model uncertainty, the feasibility of powerful and tractable simulation methods, and the ability to elicit economically based prior beliefs.
- Indeed, unlike in the single period case wherein estimation risk plays virtually no role, estimation risk does play an important role in long horizon investments with predictability.
- Barberis shows that a long horizon investor who ignores it may overallocate to stocks by a sizeable amount.
- Further, advances in computational Bayesian methods facilitates tractable solutions of fairly complex portfolio choice problems.

- To illustrate, even when the predictors evolve stochastically, both Kandel and Stambaugh (1996) and Barberis (2000) assume that the initial value of the predictive variables z_0 is non-stochastic.
- With stochastic initial value the distribution of future returns conditioned on model parameters does not longer obey a well known distributional form.
- Nevertheless, Stambaugh (1999) easily gets around this problem by implementing the Metropolis Hastings (MH) algorithm, a Markov Chain Monte Carlo procedure introduced by Metropolis et al (1953) and generalized by Hastings (1970).
- There are other several powerful numerical Bayesian algorithms such as the Gibbs Sampler and data augmentation (see a review paper by Chib and Greenberg (1996)) which make the Bayesian approach broadly applicable.

9.5 Model Uncertainly and Asset Allocation

- Financial economists have been identified variables that predict future stock returns, as noted earlier.
- However, the "correct" predictive regression specification has remained an open issue for several reasons.
- For one, existing equilibrium pricing theories are not explicit about which variables should enter the predictive regression.
- This aspect is undesirable, as it renders the empirical evidence subject to data overfitting concerns.

- Indeed, Bossaerts and Hillion (1999) confirm in-sample return predictability, but fail to demonstrate out-of-sample predictability.

- Moreover, the multiplicity of potential predictors also makes the empirical evidence difficult to interpret.

- For example, one may find an economic variable statistically significant based on a particular collection of explanatory variables, but often not based on a competing specification.

- Given that the true set of predictive variables is virtually unknown, the Bayesian methodology of model averaging, described below, is attractive, as it explicitly incorporates model uncertainty in asset allocation decisions.

- The Bayesian weighted predictive distribution of cumulative excess continuously compounded returns averages over the model space, and integrates over the posterior distribution that summarizes the within-model uncertainty about Θ_j where j is the model identifier. It is given by

$$P\left(R_{T+K}|\Phi_T\right) = \sum_{j=1}^{2^M} P\left(\mathcal{M}_j|\Phi_T\right) \int_{\Theta_j} P\left(\Theta_j|\mathcal{M}_j, \Phi_T\right) P\left(R_{T+K}|\mathcal{M}_j, \Theta_j, \Phi_T\right) d\Theta_j, \quad (473)$$

  where $P\left(\mathcal{M}_j|\Phi_T\right)$ is the posterior probability that model $\mathcal{M}_j$ is the correct one.

- Drawing from the weighted predictive distribution is done in three steps.

- First draw the correct model from the distribution of models.

- Then conditional upon the model implement the two steps, noted above, of drawing future stock returns from the model specific Bayesian predictive distribution.

9.5.1 Stock return predictability and asset pricing models - informative priors

- The Bayesian approach facilitates incorporating economically motivated priors also in the presence of return predictability.
- In particular, the classical approach has examined whether return predictability is explained by rational pricing or whether it is due to asset pricing misspecification [see, e.g., Campbell (1987), Ferson and Korajczyk (1995), and Kirby (1998)].
- Studies such as these approach finance theory by focusing on two polar viewpoints: rejecting or not rejecting a pricing model based on hypothesis tests.
- The Bayesian approach incorporates pricing restrictions on predictive regression parameters as a reference point for a hypothetical investor's prior belief.
- The investor uses the sample evidence about the extent of predictability to update various degrees of belief in a pricing model and then allocates funds across cash and stocks.
- Pricing models are expected to exert stronger influence on asset allocation when the prior confidence in their validity is stronger and when they explain much of the sample evidence on predictability.
- In particular, Avramov (2004) models excess returns on N

investable assets as

$$r_t = \alpha(z_{t-1}) + \beta f_t + u_{rt}, \tag{474}$$

$$\alpha(z_{t-1}) = \alpha_0 + \alpha_1 z_{t-1}, \tag{475}$$

$$f_t = \lambda(z_{t-1}) + u_{ft}, \tag{476}$$

$$\lambda(z_{t-1}) = \lambda_0 + \lambda_1 z_{t-1}, \tag{477}$$

where f_t is a set of K monthly excess returns on portfolio based factors, α_0 stands for an N-vector of the fixed component of asset mispricing, α_1 is an $N \times M$ matrix of the time varying component, and β is an $N \times K$ matrix of factor loadings.

- Now, a conditional version of an asset pricing model (with fixed beta) implies the relation

$$\mathbb{E}(r_t \mid z_{t-1}) = \beta\lambda(z_{t-1}) \tag{478}$$

for all t, where $\mathbb{E}$ stands for the expected value operator.

- The model (478) imposes restrictions on parameters and goodness of fit in the multivariate predictive regression

$$r_t = \mu_0 + \mu_1 z_{t-1} + v_t, \tag{479}$$

where μ_0 is an N-vector and μ_1 is an $N \times M$ matrix of slope coefficients.

- In particular, note that by adding to the right hand side of (479) the quantity $\beta\left(f_t - \lambda_0 - \lambda_1 z_{t-1}\right)$, subtracting the (same) quantity βu_{ft}, and decomposing the residual in (479) into two orthogonal components $v_t = \beta u_{ft} + u_{rt}$, we reparameterize the return-generating process (479) as

$$r_t = (\mu_0 - \beta\lambda_0) + (\mu_1 - \beta\lambda_1)z_{t-1} + \beta f_t + u_{rt}. \tag{480}$$

- Matching the right-hand side coefficients in (480) with those in (474) yields

$$\mu_0 = \alpha_0 + \beta\lambda_0, \quad (481)$$
$$\mu_1 = \alpha_1 + \beta\lambda_1. \quad (482)$$

- That is, under pricing model restrictions where $\alpha_0 = \alpha_1 = 0$ it follows that:

$$\mu_0 = \beta\lambda_0, \quad (483)$$
$$\mu_1 = \beta\lambda_1. \quad (484)$$

- This means that stock returns are predictable ($\mu_1 \neq 0$) iff factors are predictable.
- Makes sense; after all, under pricing restrictions the systematic component of returns on N stocks is captured by K common factors.
- Of course, if we relax the fixed beta assumption - time varying beta could also be a source of predictability. More later!
- Is return predictability explained by asset pricing models? Probably not!
- Kirby (1998) shows that returns are too predictable to be explained by asset pricing models.
- Ferson and Harvey (1999) show that $\alpha_1 \neq 0$.
- Avramov and Chordia (2006b) show that strategies that invest in individual stocks conditioning on time varying alpha perform extremely well. More later!
- So, should we disregard asset pricing restrictions? Not necessarily!

- The notion of rejecting or not rejecting pricing restrictions on predictability reflects extreme polar views.
- What if you are a Bayesian investor who believes pricing models could be useful albeit not perfect?
- As discussed earlier, such an approach has been formalized by Black and Litterman (1992) and Pastor (2000) in the context of iid returns and by Avramov (2004) who accounts for predictability.
- The idea is to mix the model and data.
- This is shrinkage approach to asset allocation.
- Let μ_d and Σ_d (μ_m and Σ_m) be the expected return vector and variance covariance matrix based on the data (model).
- Simplistically speaking, moments used for asset allocation are

$$\mu = \omega\mu_d + (1-\omega)\mu_m, \tag{485}$$
$$\Sigma = \omega\Sigma_d + (1-\omega)\Sigma_m, \tag{486}$$

  where ω is the shrinkage factor.
- The shrinkage of Σ is quite meaningless in this context.
- There are other quite useful shrinkage methods of Σ - see, for example, Jagannathan and Ma (2005).
- In particular,
  - If you completely believe in the model you set $\omega = 0$.
  - If you completely disregard the model you set $\omega = 1$.
  - Going with the shrinkage approach means that $0 < \omega < 1$.

- Avramov (2004) derives asset allocation under the pricing restrictions alone, the data alone, and pricing restrictions and data combined.
- He shows that
  - Optimal portfolios based on the pricing restrictions deliver the lowest Sharpe ratios.
  - Completely disregarding pricing restrictions results in the second lowest Sharpe ratios.
  - *Much* higher Sharpe ratios are obtained when asset allocation is based on the shrinkage approach.

9.6 Could you exploit predictability to design outperforming trading strategies?

- Avramov and Chordia (2006b), Avramov and Wermers (2006), and Avramov, Kosowski, Naik, and Teo (2009) are good references here.
- Let us start with Avramov and Chordia (2006b).
- They study predictability through the out of sample performance of trading strategies that invest in individual stocks conditioning on macro variables.
- Focus on the largest NYSE-AMEX firms by excluding the smallest quartile of firms from the sample.
- Ultimately we capture 3123 such firms during the July 1972 through November 2003 investment period.
- The investment universe contains 973 stocks, on average, per month.

9.6.1 The evolution of stock returns

– The underlying statistical models for excess stock returns, the market premium, and macro variables are

$$r_t = \alpha(z_{t-1}) + \beta(z_{t-1})mkt_t + v_t, \tag{487}$$
$$\alpha(z_{t-1}) = \alpha_0 + \alpha_1 z_{t-1}, \tag{488}$$
$$\beta(z_{t-1}) = \beta_0 + \beta_1 z_{t-1}, \tag{489}$$
$$mkt_t = a + b' z_{t-1} + \eta_t, \tag{490}$$
$$z_t = c + d z_{t-1} + e_t. \tag{491}$$

– Stock level predictability could come up from:

1. Model mispricing that varies with changing economic conditions ($\alpha_1 \neq 0$);
2. Factor sensitivities are predictable ($\beta_1 \neq 0$);
3. The equity premium is predictable ($b \neq 0$).

– In the end, time varying model alpha is the major source of predictability and investment profitability focusing on individual stocks, portfolios, mutual funds, and hedge funds.

– Time varying alpha is also the major source of predictability in mutual fund and hedge fund investing.

– In the mutual fund and hedge fund context alpha reflects skill. Be careful here! Alpha reflects skill only if the benchamrks used to measure perfomrance are able to price all passive payoffs.

9.6.2 The proposed trading strategy

- Form optimal portfolios from the universe of 3123 AMEX-NYSE stocks over the period 1972 through 2003 with monthly rebalancing on the basis of various models for stock returns.

- For instance, when predictability in alpha, beta, and the equity premium is permissible, the mean and variance used to form optimal portfolios are

$$\mu_{t-1} = \hat{\alpha}_0 + \hat{\alpha}_1 z_{t-1} + \hat{\beta}(z_{t-1}) \left[\hat{a} + \hat{b} z_{t-1}\right], \quad (492)$$

$$\begin{aligned} \Sigma_{t-1} &= \hat{\beta}(z_{t-1})\hat{\beta}(z_{t-1})'\hat{\sigma}^2_{mkt} + \hat{\Psi} \quad (493) \\ &+ \underbrace{\delta_1 \hat{\beta}(z_{t-1})\hat{\beta}(z_{t-1})'\hat{\sigma}^2_{mkt} + \delta_2 \hat{\Psi}}_{Estimation\ risk}. \end{aligned}$$

- The trading strategy is obtained by maximizing

$$w_t = \arg\max_{w_t} \left\{ w_t'\mu_t - \frac{1}{2(1/\gamma_t - r_{ft})} w_t' \left[\Sigma_t + \mu_t \mu_t'\right] w_t \right\}, \quad (494)$$

  where γ_t is the risk aversion level.

- We do not permit short selling of stocks but we do allow buying on margin.

9.6.3 Performance evaluation

- We implement a recursive scheme:
  - The first optimal portfolio is based on the first 120 months of data on excess returns, market premium, and predictors. (That is, the first estimation window is July 1962 through June 1972.)
  - The second optimal portfolio is based on the first 121 months of data.
- Altogether, we form 377 optimal portfolios on a monthly basis for each model under consideration.
- We record the realized excess return on any strategy

$$r_{p,t+1} = \omega_t' r_{t+1}. \tag{495}$$

- We evaluate the ex-post out-of-sample performance of the trading strategies based on the realized returns.
- Ultimately, we are able to assess the (quite large) economic value of predictability as well as show that our strategies successfully rotate across the size, value, and momentum styles during changing business conditions.

Table II: Performance measures of optimal portfolio strategies

	IID		TV-β		TV-RP		TV-α	TV-All		MKT
	UN	RE	UN	RE	UN	RE	UN	UN	RE	
Panel A: 08/72-12/03										
μ	1.90	0.64	1.33	0.51	3.16	1.65	2.90	3.16	3.04	0.98
$SE\ (\mu)$	0.76	0.22	0.72	0.21	0.80	0.61	0.87	0.84	0.77	0.28
SR	0.13	0.15	0.10	0.13	0.20	0.14	0.17	0.19	0.20	0.18
SE (SR)	0.05	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05
β	2.51	0.84	2.40	0.78	2.23	1.27	2.10	2.17	1.82	0.69
α_{cpm}	0.70	0.24	0.18	0.14	2.10	1.04	1.90	2.12	2.17	0.65
$SE\ (\alpha_{cpm})$	0.46	0.09	0.43	0.08	0.60	0.53	0.71	0.66	0.64	0.22
IR	0.08	0.13	0.02	0.09	0.18	0.10	0.14	0.17	0.18	0.15
$\tilde{\alpha}_{cpm}$	0.71	0.16	0.28	0.10	1.38	-0.08	2.18	2.24	1.55	0.07
$SE\ (\tilde{\alpha}_{cpm})$	0.45	0.09	0.42	0.08	0.56	0.40	0.72	0.67	0.61	0.12
$\Delta\ Wealth$	18.10	1.73	2.90	1.06	1247.29	13.32	188.02	716.69	802.99	4.48
Equities	200.00	94.34	200.00	84.36	195.36	74.97	199.91	200.00	179.42	64.43
Panel B: 08/72-12/87										
μ	0.92	0.45	0.00	0.33	3.87	2.05	1.98	2.72	3.36	1.39
$SE\ (\mu)$	1.29	0.35	1.19	0.31	1.38	1.05	1.36	1.35	1.08	0.44
SR	0.05	0.10	0.00	0.08	0.21	0.14	0.11	0.15	0.23	0.23
SE (SR)	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07	0.07	0.07
β	2.92	0.87	2.72	0.79	2.50	1.39	2.25	2.33	1.66	0.73
α_{cpm}	0.08	0.20	-0.78	0.10	3.15	1.65	1.33	2.05	2.88	1.18
$SE\ (\alpha_{cpm})$	0.68	0.12	0.60	0.10	1.02	0.91	1.07	1.03	0.89	0.35
IR	0.01	0.13	-0.10	0.07	0.23	0.13	0.09	0.15	0.24	0.25
$\tilde{\alpha}_{cpm}$	0.15	0.19	-0.48	0.13	1.39	-0.50	1.97	2.51	2.17	0.12
$SE\ (\tilde{\alpha}_{cpm})$	0.70	0.11	0.62	0.10	0.98	0.73	1.12	1.07	0.88	0.21
$\Delta\ Wealth$	0.58	1.32	0.15	1.05	88.14	7.13	2.58	10.56	73.93	6.16
Equities	200.00	90.11	200.00	79.27	190.54	80.73	199.82	200.00	170.10	74.32
Panel C: 01/88-12/03										
μ	2.85	0.83	2.61	0.69	2.48	1.27	3.79	3.58	2.74	0.57
$SE\ (\mu)$	0.83	0.28	0.82	0.27	0.84	0.66	1.09	1.01	1.11	0.33
SR	0.25	0.21	0.23	0.19	0.21	0.14	0.25	0.25	0.18	0.12
SE (SR)	0.08	0.08	0.08	0.08	0.08	0.07	0.07	0.08	0.07	0.08
β	1.95	0.79	1.95	0.76	1.87	1.13	1.87	1.94	2.04	0.65
α_{cpm}	1.56	0.30	1.32	0.19	1.24	0.52	2.55	2.30	1.39	0.15
$SE\ (\alpha_{cpm})$	0.57	0.15	0.57	0.13	0.62	0.56	0.94	0.83	0.92	0.27
IR	0.20	0.15	0.17	0.11	0.15	0.07	0.20	0.20	0.11	0.04
$\tilde{\alpha}_{cpm}$	1.77	0.08	1.51	0.04	1.34	0.41	2.58	2.15	1.15	0.17
$SE\ (\tilde{\alpha}_{cpm})$	0.59	0.13	0.59	0.13	0.63	0.35	0.98	0.86	0.84	0.11
$\Delta\ Wealth$	31.07	1.31	19.45	1.01	14.15	1.87	72.89	67.87	10.86	0.73
Equities	200.00	98.41	200.00	89.26	200.00	69.42	200.00	200.00	188.40	54.89

Table III: Exploring size, value, and momentum effects in trading strategies

	IID		**TV-β**		**TV-RP**		**TV-α**	**TV-All**	
	UN	RE	UN	RE	UN	RE	UN	UN	RE
Panel A: 08/72-12/03									
α_{ff}	0.75	-0.08	0.32	-0.11	1.73	0.23	1.57	1.88	1.52
$SE\ (\alpha_{ff})$	0.46	0.07	0.42	0.07	0.61	0.52	0.71	0.65	0.64
α_{ffwml}	0.48	0.05	0.12	0.02	1.62	0.93	1.01	1.32	2.08
$SE\ (\alpha_{ffwml})$	0.47	0.06	0.44	0.06	0.62	0.51	0.72	0.66	0.65
$\tilde{\alpha}_{ff}$	0.70	-0.03	0.34	-0.06	1.24	-0.35	1.77	1.89	1.03
$SE\ (\tilde{\alpha}_{ff})$	0.43	0.05	0.40	0.05	0.56	0.36	0.71	0.65	0.60
$\tilde{\alpha}_{ffwml}$	0.41	0.09	0.03	0.07	0.71	-0.08	1.15	1.17	1.13
$SE\ (\tilde{\alpha}_{ffwml})$	0.44	0.05	0.41	0.05	0.57	0.35	0.73	0.66	0.61
CS	0.71	0.01	0.11	-0.04	1.30	0.31	1.17	1.14	1.36
$SE(CS)$	0.44	0.13	0.43	0.11	0.53	0.22	0.71	0.63	0.59
Panel B: 08/72-12/87									
α_{ff}	0.39	-0.06	-0.17	-0.09	2.43	0.30	1.49	2.34	2.11
$SE\ (\alpha_{ff})$	0.69	0.09	0.59	0.09	1.04	0.88	1.11	1.05	0.90
α_{ffwml}	-0.16	0.03	-0.55	-0.02	1.72	0.88	0.56	1.63	1.85
$SE\ (\alpha_{ffwml})$	0.70	0.09	0.60	0.09	1.06	0.90	1.12	1.07	0.93
$\tilde{\alpha}_{ff}$	0.74	0.03	-0.26	0.01	0.84	-0.91	1.32	1.87	1.22
$SE\ (\tilde{\alpha}_{ff})$	0.68	0.07	0.62	0.07	1.01	0.67	1.19	1.12	0.93
$\tilde{\alpha}_{ffwml}$	0.19	0.10	-0.29	0.09	0.31	-0.59	0.52	1.21	0.79
$SE\ (\tilde{\alpha}_{ffwml})$	0.66	0.07	0.63	0.07	0.99	0.68	1.22	1.14	0.95
CS	0.30	0.04	-0.65	0.00	1.50	0.29	0.47	0.60	1.77
$SE(CS)$	0.63	0.10	0.60	0.09	0.83	0.32	1.04	0.95	0.76
Panel C: 01/88-12/03									
α_{ff}	1.58	-0.05	1.23	-0.10	1.17	0.07	1.80	1.65	0.65
$SE\ (\alpha_{ff})$	0.57	0.09	0.56	0.09	0.62	0.55	0.87	0.76	0.89
α_{ffwml}	1.42	0.11	1.25	0.07	1.63	1.01	1.51	1.18	1.85
$SE\ (\alpha_{ffwml})$	0.59	0.08	0.58	0.08	0.63	0.50	0.90	0.77	0.86
$\tilde{\alpha}_{ff}$	1.67	-0.08	1.49	-0.08	1.39	0.24	2.11	1.42	0.67
$SE\ (\tilde{\alpha}_{ff})$	0.56	0.06	0.54	0.07	0.60	0.32	0.89	0.79	0.82
$\tilde{\alpha}_{ffwml}$	1.58	0.05	1.26	0.08	1.44	0.34	1.98	0.82	1.27
$SE\ (\tilde{\alpha}_{ffwml})$	0.57	0.05	0.55	0.06	0.60	0.27	0.92	0.80	0.83
CS	1.16	-0.02	0.79	-0.08	1.12	0.32	1.81	1.63	0.99
$SE(CS)$	0.61	0.23	0.60	0.20	0.68	0.30	0.96	0.85	0.90

Table IV: Time series average of value weighted quintile scores for size, book-to-market, and momentum of trading strategies

	IID		TV-β		TV-RP		TV-α	TV-All		EW
	UN	RE	UN	RE	UN	RE	UN	UN	RE	
Panel A: 08/72-12/03										
Size	3.31	3.32	3.08	3.36	2.97	2.63	2.36	2.33	2.23	2.96
Book-to-market ratio	1.55	2.95	1.57	2.91	1.84	2.97	2.87	2.86	3.15	1.25
Momentum	4.04	3.04	3.83	3.05	3.92	2.93	4.23	4.17	2.92	12.88
Panel B: 08/72-12/87										
Size	3.31	3.23	3.03	3.30	2.85	2.74	2.40	2.35	2.41	1.13
Book-to-market ratio	1.44	3.01	1.43	2.97	1.87	2.85	2.74	2.69	3.08	1.19
Momentum	4.14	3.03	3.77	3.03	4.15	3.26	4.07	4.13	3.28	14.61
Panel C: 01/88-12/03										
Size	3.32	3.40	3.12	3.43	3.08	2.53	2.33	2.32	2.06	4.72
Book-to-market ratio	1.65	2.90	1.70	2.85	1.80	3.08	3.00	3.02	3.22	1.30
Momentum	3.96	3.04	3.89	3.06	3.71	2.61	4.38	4.20	2.58	11.21
Panel D: NBER designated expansions										
Size	3.29	3.32	3.07	3.38	2.96	2.60	2.40	2.37	2.19	3.09
Book-to-market ratio	1.57	2.95	1.59	2.90	1.82	3.02	2.91	2.90	3.20	1.25
Momentum	4.08	3.04	3.92	3.06	3.95	2.90	4.34	4.27	2.93	15.19
Panel E: NBER designated recessions										
Size	3.43	3.28	3.12	3.27	3.05	2.81	2.14	2.11	2.45	2.15
Book-to-market ratio	1.43	2.96	1.48	2.97	1.91	2.62	2.63	2.61	2.85	1.22
Momentum	3.83	3.04	3.29	2.95	3.75	3.13	3.55	3.55	2.90	-0.93
Panel F: Difference										
Size	-0.14	0.04	-0.05	**0.11**	-0.09	-0.21	**0.26**	**0.26**	**-0.26**	
Book-to-market ratio	0.14	-0.01	0.11	**-0.07**	-0.09	**0.40**	0.28	0.29	**0.35**	
Momentum	0.25	0.00	**0.63**	**0.11**	0.20	-0.23	**0.79**	**0.72**	0.03	

So far, we have shown

- Over the 1972-2003 investment period, portfolio strategies that condition on macro variables outperform the market by about 2% per month.
- Such strategies generate positive performance even when adjusted by the size, value, and momentum factors as well as by the size, book-to-market, and past return characteristics.
- In the period prior to the discovery of the macro variables, investment profitability is primarily attributable to the predictability in the equity premium.
- In the post-discovery period, the relation between the macro variables and the equity premium is attenuated considerably.
- Nevertheless, incorporating macro variables is beneficial because such variables drive stock-level alpha and beta variations.
- Predictability based strategies hold small, growth, and momentum stocks and load less (more) heavily on momentum (small) stocks during recessions.
- Such style rotation has turned out to be successful ex post.

Exploiting predictability in mutual fund returns

- Can we use our methodology to generate positive performance based on the universe of actively managed no-load equity mutual funds?
- What do we know about equity mutual funds?
  - In 2009 about $5 trillion is currently invested in U.S. equity mutual funds, making them a fundamental part of the portfolio of a domestic investor.
  - Previous work has shown that active fund management underperforms, on average, passive benchmarks.
  - Strategies that attempt to identify subsets of successful funds using information variables such as past returns or new money inflows ("hot hands" or "smart money" strategies) have underperformed when investment payoffs are adjusted the Fama-French and momentum benchmarks.
- Avramov and Wermers (2006) show that strategies that invest in no-load equity funds conditioning on macro variables (especially time varying alpha) generate substantial positive performance.

	The Dogmatist			The Skeptic					The Agnostic				
	ND	PD-1	PD-2	NS	PS-1	PS-2	PS-3	PS-4	NA	PA-1	PA-2	PA-3	PA-4
μ	6.39	6.58	6.74	6.92	5.56	6.06	15.14	11.5	7.07	5.26	5.89	16.52	12.12
std	0.16	0.16	0.16	0.21	0.21	0.18	0.28	0.21	0.21	0.21	0.18	0.28	0.22
SR	0.39	0.41	0.43	0.33	0.27	0.34	0.55	0.54	0.33	0.25	0.33	0.59	0.54
$skew$	-0.71	-0.68	0.03	-0.20	-0.38	-0.28	1.13	0.32	-0.18	-0.35	-0.31	1.05	0.21
α_{cpm}	-0.23	0.00	1.92	-0.38	-1.89	0.71	8.12	6.48	-0.23	-2.16	0.54	9.46	6.73
$P(\alpha_{cpm})$	0.66	1.00	0.38	0.88	0.36	0.79	0.07	0.08	0.93	0.31	0.84	0.04	0.08
α_{ff}	0.60	1.03	0.91	2.33	0.29	-0.51	11.14	6.86	2.49	0.06	-0.46	12.89	7.88
$P(\alpha_{ff})$	0.22	0.04	0.68	0.22	0.87	0.85	0.01	0.05	0.20	0.97	0.86	0.00	0.03
α_{wml}	0.37	0.62	3.92	-1.30	-2.49	0.83	5.98	4.95	-1.29	-2.83	0.51	8.46	6.20
$P(\alpha_{wml})$	0.46	0.22	0.07	0.45	0.13	0.76	0.14	0.17	0.47	0.10	0.85	0.04	0.10
$\tilde{\alpha}_{cpm}$	-0.09	0.14	1.85	0.14	-1.75	0.66	9.01	6.36	0.25	-1.99	0.49	10.52	6.76
$P(\tilde{\alpha}_{cpm})$	0.87	0.80	0.40	0.96	0.40	0.80	0.05	0.08	0.92	0.35	0.85	0.02	0.08
$\tilde{\alpha}_{ff}$	0.20	0.84	0.91	2.96	1.02	0.29	13.30	9.28	3.10	1.03	0.65	14.84	9.13
$P(\tilde{\alpha}_{ff})$	0.64	0.05	0.62	0.12	0.55	0.90	0.00	0.01	0.11	0.56	0.79	0.00	0.01
$\tilde{\alpha}_{wml}$	0.24	0.69	1.90	-0.84	-1.99	-0.37	9.44	7.88	-0.78	-2.01	-0.19	11.17	7.28
$P(\tilde{\alpha}_{wml})$	0.59	0.12	0.30	0.61	0.23	0.88	0.02	0.03	0.65	0.24	0.94	0.00	0.05

Andersen T.G., T Bollerslev, F.X. Diebold, and H. Ebens, 2001. The distribution of realized stock return volatility, *Journal of Financial Economics*, 61, 43-76.

Amir, Yaron, and Ravi Bansal, 2004. Risks for the long run: a potential resolution of asset pricing puzzles. Journal of Finance 59, 1481-1509.

Andrews, Donald, 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817-858.

Amihud Y., and C. M. Hurvich, 2004, Predictive regressions: a reduced-bias estimation method. Journal of Financial and Quantitative Analysis 39, 813-841.

Amihud Y., Hurvich C. M., and Y. Wang, 2008, Multiple-Predictor regressions: hypothesis testing. Review of Financial Studies 22, 413-434.

Ang, Andrew, Robert Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006. The cross-section of volatility and expected returns. Journal of Finance 61, 259-299.

Ang, Andrew, Liu, Jun, Schwarz, Krista, 2008. Using Stocks or Portfolios in Tests of Factor Models. Working Paper.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2009. High Idiosyncratic Volatility and LowReturns: International and Further U.S. Evidence. Journal of Financial Economics 91, 1-23.

Avramov, Doron, 2002. Stock return predictability and model uncertainty. Journal of Financial Economics 64, 423-458.

Avramov, Doron, 2004. Stock return predictability and asset pricing models. The Review of Financial Studies 17, 699-738.

Avramov, Doron, and Tarun Chordia, 2006. Asset pricing models and financial market anomalies. Review of Financial Studies 19, 1001-1040.

Avramov, Doron, and Russ Wermers, 2006. Investing in mutual funds when returns are predictable. Journal of Financial Economics 81,339-377.

Avramov, Doron, Tarun Chordia, and Amit Goyal, 2006. The impact of trades on daily volatility. Review of Financial Studies 19, 1241-1277.

Avramov, Doron, and Satadru Hore. Momentum, information uncertainty, and leverage - an explanation based on recursive preferences. Working Paper.

Avramov, Doron, Gergana Jostova, and Alexander Philipov, 2007. Momentum and credit rating. Journal of Finance 62, 2503-2520.

Avramov, Doron, Gergana Jostova, and Alexander Philipov, 2009. Dispersion in analyst's earnings forecasts and credit rating. Journal of Financial Economics 91,83-101.

Avramov, Doron, Robert Kosowski, Narayan Y. Naik, and Melvyn Teo, 2009. Hedge funds, managerial skill, and macroeconomic variables. Journal of Financial Economics 99,672-692.

Avramov, Doron, and Zhou,Guofu, 2010. Bayesian Portfolio Analysis. Annual Review of Financial Economics 2,25-47.

Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov, 2011. Anomalies and financial distress. Working Paper.

Avramov, Doron, Scott Cederburg, and Satadru Hore, 2011. Cross-Sectional asset pricing puzzles: an equilibrium perspective. Working Paper.

Baks K, Metrick A, and Wachter J., 2001. Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation. Journal of Finance 56,45-85.

Ball, R., Brown, P., 1968. An empirical evaluation of accounting income numbers. Journal of Accounting Research 6, 159-178.

Bansal, Ravi, 2007. Long-run risks and financial markets. Review, Federal Reserve Bank of St. Louis July, 283-300.

Bansal, Ravi, Robert F. Dittmar, and Christian T. Lundblad, 2005. Consumption, dividends, and the cross-section of equity returns. Journal of Finance 60, 1639-1672.

Bansal, R., R. F. Dittmar, and D. Kiku, 2009. Cointegration and Consumption Risks in Equity Returns. Review of Financial Studies 22,1343-75.

Bansal, R. and Yaron, A.,2004. Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. The Journal of Finance 59, 1481-1509.

Banz, R. W., 1981. The relative efficiency of various portfolios: some further evidence: discussion. Journal of Finance 35, 281-283.

Barberis, Nicholas, 2000. Investing for the long run when returns are predictable. Journal of Finance 55, 225-264.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998. A model of investor sentiment. Journal of Financial Economics 49, 307-343.

Barro, Robert J., 2005. Equity premium and rare events. Working Paper.

Basu, Sankarshan, 1977. Investment performance of common stocks in relation to their price-earnings ratios: a test of the efficient market hypothesis. Journal of Finance 32, 663 - 682.

Bawa, V.S., and R.W. Klein, 1976. The effect of estimation risk on optimal portfolio choice. Journal of Financial Economics 3, 215–231.

Bawa, V.S., S.J. Brown, and R.W. Klein, 1979. Estimation risk and optimal portfolio choice. North-Holland, Amsterdam.

Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York (1985).

Berk, Jonathan, Richard Green, and Vasant Naik, 1999. Optimal investment, growth options, and security returns. Journal of Finance 54, 1553-1607.

Black, F., and R. Litterman,1992. Global Portfolio optimization. *Financial Analysts Journal* 48 , 28–43.

Bossaerts, Peter, and Pierre Hillion, 1999. Implementing statistical criteria to select return forecasting models: what do we learn? Review of Financial Studies 12, 405-428.

Brandt, Michael W., Goyal, Amit, Santa-Clara, Pedro, and Stroud, Jonathan R., 2005. A simulation approach to dynamic portfolio choice with an application to learning about return predictability. Review of Financial Studies 18, 831-874.

Brennan, M., Tarun Chordia, and Avanidhar Subrahmanyam, 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. Journal of Financial Economics 49,345-373.

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

Breeden, Douglas, Gibbons, Michael, and Robert Litzenberger, 1989. Empirical tests of the consumption-oriented CAPM. Journal of Finance 44, 231-262.

Brown, S. J,1976. Optimal portfolio choice under uncertainty. Ph.D. Dissertation. University of Chicago.

Brown, S. J., 1978. The portfolio choice problem: comparison of certainty equivalence and optimal Bayes portfolios. *Communications in Statistics — Simulation and Computation* 7, 321–334.

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

Campbell, John, Jens Hilscher, and Jan Szilagyi, 2008. In search of distress risk. Journal of Finance 63, 2899-2939.

Campbell, John, and John Cochrane, 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205-251.

Campbell, John, and John Cochrane, 2000. Explaining the poor performance of consumption-based asset pricing models. Journal of Finance 55, 2863-2878.

Campbell, John, Andrew Lo, and A. Craig MacKinlay, 1997. The Econometrics of Financial Markets. Princeton University Press.

Campbell, J. Y., and S. B. Thompson. 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Review of Financial Studies 21,1509-31.

Carhart, Mark, 1997. On persistence in mutual fund performance. Journal of Finance 52, 57-82.

Chen, Long, Novy-Marx, Robert, and Zhang, Lu, 2010. An Alternative Three-Factor Model. Working Paper.

Chen, Nai-fu, Richard Roll, and Stephen Ross, 1986. Economic forces and the stock market. Journal of Business 59, 383-403.

Chen, Nai-fu 1991. Financial investment opportunities and the macroeconomy. Journal of Finance 46, 529-554.

Chordia, Tarun, and Lakshmanan Shivakumar, 2002. Momentum, business cycle and time-varying expected returns. Journal of Finance 57, 985-1019.

Connor, Gregory, and Robert Korajczyk, 1988. Risk and return in an equilibrium APT: application of a new test methodology. Journal of Financial Economics 21, 255-289.

Cochrane, John, 1996.A cross-sectional test of a production-based asset pricing model. Journal of Political Economy 104, 572-621.

Cochrane, John, 2001. Long-term debt and optimal policy in the fiscal theory of the price level. Econometrica 69, 69-116.

Cochrane, John, 2007. Financial Markets and the Real Economy. Handbook of the Equity Premium, 237-325. Amsterdam, The Netherlands: Elsevier.

Cochrane, J. H. 2008. The Dog That Did Not Bark: A Defense of Return Predictability. Review of Financial Studies 21,1533-1575.

Cohen, Lauren, and Andrea Frazzini, 2008. Economic links and predictable returns. Journal of Finance 63, 1977-2011.

Cooper, Michael, Roberto Gutierrez, and Bill Marcum, 2001. On the predictability of stock returns in real time. Unpublished working paper, Purdue University.

Cooper, Michael, Roberto Gutierrez, and Allaudeen Hameed, 2004. Market States and momentum. Journal of Finance 59, 1345-1365.

Cooper, Michael, Huseyin Gulen, and Michael J. Schill, 2008. Asset Growth and the Cross-Section of Stock Returns. Journal of Finance 63, 1609-1651.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997. Measuring Mutual Fund Performance with Characteristic-Based Benchmarks.Journal of Finance 52,1035-1058.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998. Investor psychology and security market under- and overreactions. Journal of Finance 53, 1839-1885.

Daniel, Kent, and Sheridan Titman, 2011. Testing factor-model explanations of market anomalies. Working Paper.

Da, Zhi, and Yun, Hayong, Electricity Consumption and Asset Prices. Working Paper.

DeBondt, Werner, and Richard Thaler, 1985. Does the stock market overreact? Journal of Finance 40, 793-805.

DeBondt, Werner, and Richard Thaler, 1987. Further evidence on investor overreaction and stock market seasonality. Journal of Finance 42, 557-581.

Dichev, Ilia, 1998. Is the risk of bankruptcy a systematic risk? Journal of Finance 53,557-581 .

Dichev, Ilia, Piotroski, J., 2001. The long-run stock returns following bond ratings changes. Journal of Finance 56, 173-203.

Diether, Karl, Christopher Malloy, and Anna Scherbina, 2002. Differences of opion and the cross section of stock returns. Journal of Finance 57, 2113-2141.

Duffie, Darrell, and Larry Epstein, 1992. Stochastic differential utility. Econometrica 60, 353-394.

Epstein, Larry, and Stanley Zin, 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. Econometrica 57, 937-969.

Fama, Eugene, 1998, Market efficiency, long-term returns, and behavioral finance, Journal of Financial Economics, 49, 283-306

Fama, Eugene, and Kenneth French, 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23-49.

Fama, Eugene, and Kenneth French, 1992. The cross-section of expected stock returns. Journal of Finance 47, 427-465.

Fama, Eugene, and Kenneth French, 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

Fama, Eugene, and Kenneth French, 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55-84.

Fama, Eugene F., and Kenneth R. French, 2006. Profitability, investment, and average returns. Journal of Financial Economics 82, 491-518.

Fama, Eugene, and Kenneth French, 2008. Dissecting anomalies. Journal of Finance 63, 1653-1678.

Fama, Eugene, MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 81, 607-636.

Fama, Eugene, and G. William Schwert, 1977. Asset returns and inflation. Journal of Financial Economics 5, 115-146.

Ferson, Wayne, and Campbell Harvey, 1999. Conditioning variables and the cross section of stock returns. Journal of Finance 54, 1325-1360.

Ferson, Wayne E., Nallareddy, Suresh and Xie, Biqin,2011. The 'out-of-sample' performance of long-run risk models. Working Paper.

Ferson, Wayne E., and Robert A. Korajczyk. Do arbitrage pricing models explain the predictability of stock returns? Journal of business 68, 309-349.

Foster, F. Douglas, Tom Smith, and Robert Whaley, 1997. Assessing the goodness-of-fit of asset pricing models: the distribution of maximal R^2. Journal of Finance 52, 591-607.

Frost PA, Savarino JE. 1986. An empirical Bayes approach to efficient portfolio selection. J. Financ. Quant. Anal 21,293C305.

Fung, William, and David Hsieh, 2001. The risk in hedge fund strategies: theory and evidence from trend followers. Review of Financial Studies 14, 313-341.

Fung, William, and David Hsieh, 2004. Hedge fund benchmarks: a risk based approach. Financial Analysts Journal 60, 65-80.

Geczy CC, Stambaugh RF, Levin D. 2005. Investing in socially responsible mutual funds. Working Paper.

Geweke, John, and Guofu Zhou, 1996. Measuring the pricing error of the arbitrage pricing theory. Review of Financial Studies 9, 557-587.

Gibbons, Michael, Stephen Ross, and Jay Shanken, 1989. A test of the efficiency of a given portfolio. The Econometric Society 57, 1121-1152.

Gomes, Joao, Leonid Kogan, and Lu Zhang, 2003. Equilibrium cross section of returns. Journal of Political Economy 111, 693-732.

Goyal, Amit, and Ivo Welch, 2006. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455-1508.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980. On the Impossibility of Informationally Efficient Markets. The American Economic Review 70, 393-408.

Grossman, Sanford, and Robert Shiller, 1981. The determinants of the variability of stock market prices. American Economic Review 71, 222 - 27.

Grundy, Bruce, and J. Spencer Martin, 2001. Understanding the nature of the risks and the source of the rewards to momentum investing. Review of Financial Studies 14, 29-78.

Hansen, Lars, 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029-1054.

Hansen, L. P., J. P. Heaton, and N. Li, 2008. Consumption strikes back?: Measuring long run risk. Journal of Political Economy 116, 260-302.

Hansen, Lars, and Ravi Jagannathan, 1997. Assessing specification errors in stochastic discount factor models. Journal of Finance 52, 557-590.

Hansen, Lars, and Kenneth Singleton, 1982. Generalized instrumental variables estimation of nonliner rational expectations models. Econometrica 50, 1269-1286.

Hansen, Lars, and Kenneth Singleton, 1983. Stochastic consumption, risk aversion and the temporal behavior of asset returns. Journal of Political Economy 91, 249-265.

Harvey, Campbell, 1989. Time-varying conditional covariances in tests of asset pricing-models. Journal of Financial Economics 24, 289-317.

Harvey, Campbell, and Akhtar Siddique, 2000. Conditional skewness in asset pricing tests. Journal of Finance 55, 1263-1295.

Harrison, J. Michael, and David Kreps, 1979. Martingales and arbitrage in multiperiod securities markets. Journal of Economic Theory 20, 381-408.

Haugen, Robert, and Nardin Baker, 1996. Commonality in the determinants of expected stock returns. Journal of Financial Economics 41, 401-439.

He G, Litterman R. 1999. The intuition behind Black-Litterman model portfolios. Working Paper. Goldman Sachs Quant. Resour. Group.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993. Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974-1988. Journal of Finance 48, 93-130.

Hon, Harrison, Terrence Lim, and Jeremy Stein, 2000. Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies. Journal of Finance 55, 265-295.

Jagannathan, R. and Wang, Y. (2007), Lazy Investors, Discretionary Consumption, and the Cross-Section of Stock Returns. The Journal of Finance 62, 1623-1661.

Jegadeesh, Narasimhan, 1990. Evidence of predictable behavior in security returns. Journal of Finance 45, 881-898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65-91.

Jegadeesh, Narasimhan, and Sheridan Titman, 2001. Profitability of momentum strategies: an evaluation of alternate explanations. Journal of Finance 56, 699-720.

Jegadeesh, Narasimhan, and Sheridan Titman, 2002. Cross-sectional and time-series determinants of momentum returns. Review of Financial Studies 15, 143-157.

Johnson, Timothy C., 2002. Rational momentum effects. Journal of Finance 57(2),585-608.

Jones, C., and Shanken J. 2005. Mutual fund performance and learning across funds. Journal of Financial Economics 78,507-52.

Jorion P. 1986. Bayes-Stein estimation for portfolio analysis. J. Financ. Quant. Anal. 21, 279-92.

Kandel, Shmuel, and Robert Stambaugh, 1995. Portfolio inefficiency and the cross-section of expected returns. Journal of Finance 50, 157-184.

Kandel, Shmuel, and Robert Stambaugh, 1996. On the predictability of stock returns: an asset allocation perspective. Journal of Finance 51, 385-424.

Kendall, M. G.,1954. Note on the bias in the estimation of autocorrelation. Biometrika 41,403–404.

Keim, Donald, and Robert Stambaugh, 1986. Predicting returns in the stock and the bond markets. Journal of Financial Economics 17, 357-390.

Kirby, Chris, 1997. Measuring the predictable variation in stock and bond returns. Review of Financial Studies 10, 579-630.

Kirby, Chris, 1998. The restrictions on predictability implied by rational asset pricing models. Review of Financial Studies 11, 343-382.

Klein, R. W., and V. S. Bawa, 1976, "The Effect of Estimation Risk on Optimal Portfolio Choice." *Journal of Financial Economics* 3, 215–231.

Korajczyk, Robert, and Ronnie Sadka, 2004. Are momentum profits robust to trading costs? Journal of Finance 59, 1039-1082.

Lee, Charles, and Bhaskaran Swaminathan, 2000. Price momentum and trading volume. Journal of Finance 55, 2017-2069.

Lehmann, B., 1990. Fads, martingales, and market efficiency. Quarterly Journal of Economics 105, 1-28.

Lettau, Martin, and Sydney Ludvigson, 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56, 815-849.

Lettau, Martin, and Sydney Ludvigson, 2001. Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying. Journal of Political Economy 10, 1238-1286.

Lewellen,Jonathan, Stefan Nagel, Jay Shanken, 2010. A skeptical appraisal of asset pricing tests. Journal of Financial Economics 96, 175-194.

Liew, J., Vassalou, M., 2000. Can book-to-market, size and momentum be risk factors that predict growth? Journal of Financial Economics 57, 221-245.

Lintner, John, 1965. The valuation of risky assets and the selection of risky instruments in stock portfolios and capital budgets. The Review of Economics and Statistics 47 , 13-37 .

Lo, Andrew, and A. Craig MacKinlay, 1990. Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies 3, 431-467.

Lucas, Robert, 1978. Asset prices in an exchnage economy. Econometrica 46, 1429-1445.

MacKinlay AC, Pastor L., 2000. Asset pricing models: implications for expected returns and portfolio selection. Review of Financial Studies 13,883-916.

Mankiw, N. Gregory, and Matthew Shapiro, 1986. Do we reject to often? Small sample properties of tests of rational expectations models. Economic Letters 20, 139-145.

Markowitz, Harry M, 1952.Mean-variance analysis in portfolio choice and capital markets. Journal of Finance 7,77-91.

McCulloch, Robert, and Peter E. Rossi, 1990, Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory, *Journal of Financial Economics* 28, 7–38.

Mehra, Rajnish, and Edward Prescott, 1985. The equity premium: a puzzle. Journal of Monetary Economics 15, 145-161.

Menzly, Lior, Tano Santos, and Pietro Veronesi, 2004. Understanding predictability. Journal of Political Economy 112, 1-47.

Merton, Robert, 1973. An intertemporal capital asset pricing model. Econometrica 41, 867-887.

Moskowitz, Tobias, and Mark Grinblatt, 1999. Do industries explain momentum? Journal of Finance 54, 1249-1290.

Newey, Whitney, and Kenneth West, 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-708.

Novy-Marx, Robert, 2010. The Other Side of Value: Good Growth and the Gross Profitability Premium. Working Paper

Parker, J. A., and C. Julliard, 2005. Consumption risk and the cross-section of expected returns. Journal of Political Economy 113, 185-222.

Pastor, Lubos, 2000. Portfolio selection and asset pricing models. Journal of Finance 55, 179-223.

Pastor, L., Stambaugh, R., 2000. Comparing asset pricing models: an investment perspective. Journal of Financial Economics 56, 335-381.

Pastor, L., Stambaugh, R., 2002a. Mutual fund performance and seemingly unrelated assets. Journal of Financial Economics 63, 315-349.

Pastor, L., Stambaugh RF. 2002b. Investing in equity mutual funds. Journal of Financial Economics 63,351-80.

Pastor, Lubos, and Robert Stambaugh, 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642-685.

Piotroski,Joseph D., 2000. Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers. Journal of Accounting Research 38,1-41.

Rapach, David, Strauss,Jack, and Zhou,Guofu, 2010.Out-of-Sample equity premium prediction: combination forecasts and links to the real economy. Review of Financial Studies 23, 821-862.

Richardson, S. A., and R. G. Sloan, 2003. External Financing and Future Stock Returns. Working Paper.

Rietz, T., 1988. The equity risk premium a solution. Journal of Monetary Economics 22, 117-131.

Roll, Richard, 1977. A critique of the asset pricing theory's tests: part i. Journal of Financial Economics 4, 129-176.

Roll, Richard, and Stephen Ross, 1980. An empirical investigation of the arbitrage pricing theory. Journal of Finance 35, 1073-1103.

Roll, Richard, and Stephen Ross, 1994. On the cross-sectional relation between expected returns and betas. Journal of Finance 49, 101-121.

Ross, Stephen, 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341-360.

Rouwenhorst, K. Geert, 1998. International Momentum Strategies. Journal of Finance 53, 267-284.

Rubinstein, Mark, 1976. Valuation of uncertain income streams and pricing of options. Bell Journal of Economics 7, 407-425.

Savov, A., 2011. Asset Pricing with Garbage. The Journal of Finance 66, 177-201.

Schwert, G. William, 2003. Anomalies and market efficiency, Handbook in economic and finance, Edited by G.M. Constantinides, M. Harris, and R. Stulz.

Sloan, Richard, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings? The Accounting Review 71, 289-315.

Shanken, Jay, 1990. Intertemporal asset pricing: an empirical investigation. Journal of Econometrics 45, 99-120.

Shanken, Jay, 1992. On the estimation of beta-pricing models. Review of Financial Studies 5, 1-55.

Sharpe, William, 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425-442.

Stambaugh, Robert, 1999. Predictive regressions. Journal of Financial Economics 54, 375-421.

Tierney, Luke, 1994. Markov-chains for exploring posterior distributions. Annals of Statistics 22, 1701-1728.

Titman, Sheridan, K. C. John Wei, and Feixue Xie, 2004.Capital Investments And Stock Returns. Journal of Financial and Quantitative Analysis 39, 677-700.

Tu J, Zhou G. 2010. Incorporating economic objectives into Bayesian priors: portfolio choice under parameter uncertainty. Journal of Financial Quantitative Analysis 45,959-986.

Wang, Huijun, and Jianfeng Yu, 2010. Dissecting The Profitability Premium. Working Paper.

Weil, Philippe, 1990. Non-expected utility in macroeconomics. Quarterly Journal of Economics 105, 29-43.

Weitzman, Martin, 2007. Subjective expectations and asset-return puzzles. American Economic Review 97, 1102-1130.

Wheatley,Simon,1988. Some tests of the consumption-based asset pricing model. Journal of Monetary Economics 22,193-215.

White, Hal 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817-838.

Winkler, Robert L., 1973, Bayesian Models for Forecasting Future Security Prices, Journal of Financial and Quantitative Analysis 8, 387–405.

Winkler, Robert L., and Christopher B. Barry, 1975, A Bayesian Model for Portfolio Selection and Revision, Journal of Finance 30, 179–192.

Zellner, Arnold, 1971. An introduction to bayesian inference in econometrics. Wiley, New York.

Zellner, Arnold, and V. Karuppan Chetty, 1965. Prediction and decision problem in regression models from the Bayesian point of view. Journal of the American Statistical Association 60, 608-615.

Zheng, Lu, 1999. Is money smart? a study of mutual fund investors's fund selection ability. Journal of Finance 54, 901-933.

Zhang, X. Frank, 2006, Information uncertainty and stock returns. Journal of Finance 61, 105-136.

Zhou, G., 2009. Beyond Black-Litterman: letting the data speak. J. Portfolio Management 36, 36-45.