

Topics in Asset Pricing

Lecture Notes

Professor Doron Avramov

Chinese University of Hong Kong

Hebrew University of Jerusalem

Background, Objectives, and Pre-requisite

- The past few decades have been characterized by an extraordinary growth in the use of quantitative methods in the analysis of various asset classes; be it equities, fixed income securities, commodities, currencies, and derivatives.
- In response, financial economists have routinely been using advanced mathematical, statistical, and econometric techniques to understand the dynamics of asset pricing models, market anomalies, equity premium predictability, asset allocation, security selection, volatility, correlation, and the list goes on.
- This course attempts to provide a fairly deep understanding of such topical issues.
- It targets advanced master and PhD level students in finance and economics.
- Required: prior exposure to matrix algebra, distribution theory, Ordinary Least Squares, as well as kills in computer programing beyond Excel:
 - MATLAB and R are the most recommended for this course. OCTAVE could be used as well, as it is a free software, and is practically identical to MATLAB when considering the scope of the course.
 - STATA or SAS could be useful.

Topics to be covered

- From CAPM to market anomalies
- The credit risk effect in the cross section of asset returns
- Rational versus behavioural asset pricing
- Are market anomalies pervasive?
- Conditional CAPM
- Conditional versus unconditional portfolio efficiency
- Multi-factor models
- Interpreting factor models
- Machine learning approaches in asset pricing
- Consumption based asset pricing models
- The discount factor representation in asset pricing
- The equity premium puzzle
- The risk free rate puzzle
- The Epstein-Zin preferences
- Long run risk
- Habit formation
- Prospect theory

Topics to be covered

- Time series asset pricing tests
- Cross section asset pricing tests
- Stock return predictability
- Finite sample bias in predictive regressions
- The Campbell-Shiller present value model
- Vector auto regressions in asset pricing
- On the riskiness of stocks for the long run – Bayesian perspectives
- On the risk-return relation in the time series
- GMM: Theory and application
- The covariance matrix of regression slope estimates in the presence of heteroskedasticity and autocorrelation
- Bayesian Econometrics
- Bayesian portfolio optimization
- The Hansen Jagannathan Distance measure
- Spectral Analysis

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
- Class notes as well as published and working papers in finance and economics as listed in the reference list

From Rational Asset pricing to Market Anomalies

Expected Return

- Statistically, an expected asset return (in excess of the risk free rate) can be formulated as

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

where f_t denotes a set of K portfolio spreads realized at time t , β_i is a K vector of factor loadings, and α_i reflects the expected return component unexplained by factors, or model mispricing.

- The market model is a statistical setup with f represented by excess return on the market portfolio.
- An asset pricing model aims to identify economic or statistical factors that eliminate model mispricing.
- Absent alpha, expected return differential across assets is triggered by factor loadings only.
- The presence of model mispricing could give rise to additional cross sectional effects.
- If factors are not return spreads (e.g., consumption growth) α is no longer asset mispricing.
- The presence of factor structure with no alpha does not imply that asset pricing is essentially rational.
- Indeed, *comovement* of assets sharing similar styles (e.g., value, large cap) or belonging to the same industry could be attributed to biased investor's beliefs just as they could reflect risk premiums.
- Later, we discuss in more detail ways of interpreting factor models.

CAPM

- The CAPM of Sharpe (1964), Lintner (1965), and Mossin (1966) originates the literature on asset pricing models.
- The CAPM is an equilibrium model in a single-period economy.
- It imposes an economic restriction on the statistical structure of expected asset return.
- The unconditional version is one where moments are time-invariant.
- Then, the expected excess return on asset i is formulated as

$$\mathbb{E}(r_{i,t}^e) = \text{cov}(r_{i,t}, r_{m,t}^e) \frac{\mathbb{E}(r_{m,t}^e)}{\text{var}(r_{m,t}^e)} = \beta_{i,m} \mathbb{E}(r_{m,t}^e)$$

where $r_{m,t}^e$ is excess return on the market portfolio at time t .

- Asset risk is the covariance of its return with the market portfolio return.
- Risk premium, or the market price of risk, is the expected value of excess market return.
- In CAPM, risk means co-movement with the market.

CAPM

- The higher the co-movement the less desirable the asset is, hence, the asset price is lower and the expected return is higher.
- This describes the risk-return tradeoff: high risk comes along with high expected return.
- The market price of risk, common to all assets, is set in equilibrium by the risk aversion of investors.
- There are conditional versions of the CAPM with time-varying moments.
- For one, risk and risk premium could vary with macro economy variables such as the default spread and risk (beta) can vary with firm-level variables such as size and book-to-market.
- Time varying parameters could be formulated using a beta pricing setup (e.g., Ferson and Harvey (1999) and Avramov and Chordia (2006a)).
- Another popular approach is time varying pricing kernel parameters (e.g., Cochrane (2005)).
- Risk and risk premium could also obey latent autoregressive processes.
- Lewellen and Nagel (LN 2006) model beta variation in rolling samples using high frequency data.

Empirical Violations: Market Anomalies

- The CAPM is simple and intuitive and it is widely used among academic scholars and practitioners as well as in finance textbooks.
- However, there are considerable empirical and theoretical drawbacks at work.
- To start, the CAPM is at odds with anomalous patterns in the cross section of asset returns.
- Market anomalies describe predictable patterns (beyond beta) related to firm characteristics such as size, book-to-market, past return (short term reversals and intermediate term momentum), earnings momentum, dispersion, net equity issuance, accruals, credit risk, asset growth, capital investment, profitability, new 52-high, IVOL, and the list goes on.
- Harvey, Liu, and Zhu (2016) document 316 (some are correlated) factors discovered by academia.
- They further propose a t-ratio of at least 3 to make a characteristic significant in cross section regressions.
- Green, Hand, and Zhang (2013) find 330 return predictive signals.
- See also the survey papers of Subrahmanyam (2010) and Goyal (2012).

Multi-Dimension in the Cross Section?

- The large number of predictive characteristics leads Cochrane (2011) to conclude that there is a multi-dimensional challenge in the cross section.
- On the other hand, Avramov, Chordia, Jostova, and Philipov (2013, 2018) attribute the predictive ability of various factors to financial distress.
- They thus challenge the notion of multi-dimension in the cross section.
- Their story is straightforward: firm characteristics become extreme during financial distress, such as large negative past returns, large negative earnings surprises, large dispersion in earnings forecasts, large volatility, and large credit risk.
- Distressed stocks are thus placed in the short-leg of anomaly portfolios.
- As distressed stocks keep on losing value, anomaly profitability emerges from selling them short.
- This explains the IVOL effect, dispersion, price momentum, earnings momentum, among others, all of these effects are a manifestation of the credit risk effect.
- The *vast* literature on market anomalies is briefly summarized below.

The Beta Effect

- Friend and Blume (1970) and Black, Jensen, and Scholes (1972) show that high beta stocks deliver negative alpha, or they provide average return smaller than that predicted by the CAPM.
- Frazzini and Pedersen (2014) demonstrate that alphas and Sharpe ratios are almost monotonically declining in beta among equities, bonds, currencies, and commodities.
- They propose the BAB factor – a market neutral trading strategy that buys low-beta assets, leveraged to a beta of one, and sells short high-beta assets, de-leveraged to a beta of one.
- The BAB factor realizes high Sharpe ratios in US and other equity markets.
- What is the economic story?
- For one, high beta stocks could be in high demand by constrained investors.
- Moreover, Hong and Sraer (2016) claim that high beta assets are subject to speculative overpricing.
- Just like the beta-return relation is counter intuitive – an apparent violation of the risk return tradeoff – there are several other puzzling relations in the cross section of asset returns.
- The credit risk return relation (high credit risk low future return) is coming up next.

The credit risk return relation

- Dichev (1998), Campbell, Hilscher, and Szilagyi (2008), and Avramov, Chordia, Jostova, and Philipov (2009, 2013) demonstrate a negative cross-sectional correlation between credit risk and returns.
- Campbell, Hilscher, and Szilagyi (2008) suggest that such negative relation is a challenge to standard rational asset pricing models.
- Once again, the risk-return tradeoff is challenged.
- Campbell, Hilscher, and Szilagyi (2008) use failure probability estimated by a dynamic logit model with both accounting and equity market explanatory variables.
- Using the Ohlson (1980) O-score, the Z-score, or credit rating to proxy distress yields similar results.
- Dichev and Piotroski (2001) and Avramov, Chordia, Jostova, and Philipov (2009) document abnormal stock price declines following credit rating downgrades, and further the latter study suggests that market anomalies only characterize financially distressed firms.
- On the other hand, Vassalou and Xing (2004) use the Merton's (1974) option pricing model to compute default measures and argue that default risk is systematic risk, and Friewald, Wagner, and Zechner (2014) find that average returns are positively related to credit risk assessed through CDS spreads.
- I subscribe to the negative credit risk return relation, yet the contradicting findings beg for resolution.

Size Effect

Size effect: higher average returns on small stocks than large stocks. Beta cannot explain the difference. First papers go to Banz (1981), Basu (1983), and Fama and French (1992)

Table I

Average Returns, Post-Ranking β s and Average Size For Portfolios Formed on Size and then β : Stocks Sorted on ME (Down) then Pre-Ranking β (Across): July 1963 to December 1990

Portfolios are formed yearly. The breakpoints for the size (ME, price times shares outstanding) deciles are determined in June of year t ($t = 1963-1990$) using all NYSE stocks on CRSP. All NYSE, AMEX, and NASDAQ stocks that meet the CRSP-COMPUSTAT data requirements are allocated to the 10 size portfolios using the NYSE breakpoints. Each size decile is subdivided into 10 β portfolios using pre-ranking β s of individual stocks, estimated with 2 to 5 years of monthly returns (as available) ending in June of year t . We use only NYSE stocks that meet the CRSP-COMPUSTAT data requirements to establish the β breakpoints. The equal-weighted monthly returns on the resulting 100 portfolios are then calculated for July of year t to June of year $t + 1$.

The post-ranking β s use the full (July 1963 to December 1990) sample of post-ranking returns for each portfolio. The pre- and post-ranking β s (here and in all other tables) are the sum of the slopes from a regression of monthly returns on the current and prior month's returns on the value-weighted portfolio of NYSE, AMEX, and (after 1972) NASDAQ stocks. The average return is the time-series average of the monthly equal-weighted portfolio returns, in percent. The average size of a portfolio is the time-series average of monthly averages of $\ln(\text{ME})$ for stocks in the portfolio at the end of June of each year, with ME denominated in millions of dollars.

The average number of stocks per month for the size- β portfolios in the smallest size decile varies from 70 to 177. The average number of stocks for the size- β portfolios in size deciles 2 and 3 is between 15 and 41, and the average number for the largest 7 size deciles is between 11 and 22.

The All column shows statistics for equal-weighted size-decile (ME) portfolios. The All row shows statistics for equal-weighted portfolios of the stocks in each β group.

	All	Low- β	β -2	β -3	β -4	β -5	β -6	β -7	β -8	β -9	High- β
Panel A: Average Monthly Returns (in Percent)											
All	1.25	1.34	1.29	1.36	1.31	1.33	1.28	1.24	1.21	1.25	1.14
Small-ME	1.52	1.71	1.57	1.79	1.61	1.50	1.50	1.37	1.63	1.50	1.42
ME-2	1.29	1.25	1.42	1.36	1.39	1.65	1.61	1.37	1.31	1.34	1.11
ME-3	1.24	1.12	1.31	1.17	1.70	1.29	1.10	1.31	1.36	1.26	0.76
ME-4	1.25	1.27	1.13	1.54	1.06	1.34	1.06	1.41	1.17	1.35	0.98
ME-5	1.29	1.34	1.42	1.39	1.48	1.42	1.18	1.13	1.27	1.18	1.08
ME-6	1.17	1.08	1.53	1.27	1.15	1.20	1.21	1.18	1.04	1.07	1.02
ME-7	1.07	0.95	1.21	1.26	1.09	1.18	1.11	1.24	0.62	1.32	0.76
ME-8	1.10	1.09	1.05	1.37	1.20	1.27	0.98	1.18	1.02	1.01	0.94
ME-9	0.95	0.98	0.88	1.02	1.14	1.07	1.23	0.94	0.82	0.88	0.59
Large-ME	0.89	1.01	0.93	1.10	0.94	0.93	0.89	1.03	0.71	0.74	0.56

Value Effect

- Value effect: higher average returns on value stocks than growth stocks. Beta cannot explain the difference.
- Value firms: Firms with high E/P, B/P, D/P, or CF/P. The notion of value is that physical assets can be purchased at low prices.
- Growth firms: Firms with low ratios. The notion is that high price relative to fundamentals reflects capitalized growth opportunities.

Table I
Summary Statistics and Three-Factor Regressions for Simple Monthly Percent Excess Returns on 25 Portfolios Formed on Size and BE/ME: 7/63–12/93, 366 Months

R_f is the one-month Treasury bill rate observed at the beginning of the month (from CRSP). The explanatory returns R_M , SMB, and HML are formed as follows. At the end of June of each year t (1963–1993), NYSE, AMEX, and Nasdaq stocks are allocated to two groups (small or big, S or B) based on whether their June market equity (ME, stock price times shares outstanding) is below or above the median ME for NYSE stocks. NYSE, AMEX, and Nasdaq stocks are allocated in an independent sort to three book-to-market equity (BE/ME) groups (low, medium, or high; L, M, or H) based on the breakpoints for the bottom 30 percent, middle 40 percent, and top 30 percent of the values of BE/ME for NYSE stocks. Six size-BE/ME portfolios (S/L, S/M, S/H, B/L, B/M, B/H) are defined as the intersections of the two ME and the three BE/ME groups. Value-weight monthly returns on the portfolios are calculated from July to the following June. SMB is the difference, each month, between the average of the returns on the three small-stock portfolios (S/L, S/M, and S/H) and the average of the returns on the three big-stock portfolios (B/L, B/M, and B/H). HML is the difference between the average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low-BE/ME portfolios (S/L and B/L). The 25 size-BE/ME portfolios are formed like the six size-BE/ME portfolios used to construct SMB and HML, except that quintile breakpoints for ME and BE/ME for NYSE stocks are used to allocate NYSE, AMEX, and Nasdaq stocks to the portfolios.

BE is the COMPUSTAT book value of stockholders' equity, plus balance sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock. Depending on availability, we use redemption, liquidation, or par value (in that order) to estimate the book value of preferred stock. The BE/ME ratio used to form portfolios in June of year t is then book common equity for the fiscal year ending in calendar year $t - 1$, divided by market equity at the end of December of $t - 1$. We do not use negative BE firms, which are rare prior to 1980, when calculating the breakpoints for BE/ME or when forming the size-BE/ME portfolios. Also, only firms with ordinary common equity (as classified by CRSP) are included in the tests. This means that ADR's, REIT's, and units of beneficial interest are excluded. The market return R_M is the value-weight return on all stocks in the size-BE/ME portfolios, plus the negative BE stocks excluded from the portfolios.

Size	Book-to-Market Equity (BE/ME) Quintiles									
	Low	2	3	4	High	Low	2	3	4	High
Panel A: Summary Statistics										
	Means					Standard Deviations				
Small	0.31	0.70	0.82	0.95	1.08	7.67	6.74	6.14	5.85	6.14
2	0.48	0.71	0.91	0.93	1.09	7.13	6.25	5.71	5.23	5.94
3	0.44	0.68	0.75	0.86	1.05	6.52	5.53	5.11	4.79	5.48
4	0.51	0.39	0.64	0.80	1.04	5.86	5.28	4.97	4.81	5.67
Big	0.37	0.39	0.36	0.58	0.71	4.84	4.61	4.28	4.18	4.89

The International Value Effect

Table III

Annual Dollar Returns in Excess of U.S. T-Bill Rate for Market, Value, and Growth Portfolios: 1975–1995

Value and growth portfolios are formed on book-to-market equity (B/M), earnings/price (E/P), cashflow/price (C/P), and dividend/price (D/P), as described in Table II. We denote value (high) and growth (low) portfolios by a leading H or L; the difference between them is H – L. The first row for each country is the average annual return. The second is the standard deviation of the annual returns (in parentheses) or the *t*-statistic testing whether H – L is different from zero [in brackets].

	Market	HB/M	LB/M	H-LB/M	HE/P	LE/P	H-LE/P	HC/P	LC/P	H-LC/P	HD/P	LD/P	H-LD/P
U.S.	9.57 (14.64)	14.55 (16.92)	7.75 (15.79)	6.79 [2.17]	14.09 (18.10)	7.38 (15.23)	6.71 [2.28]	13.74 (16.73)	7.08 (15.99)	6.66 [2.08]	11.75 (13.89)	8.01 (17.04)	3.73 [1.22]
Japan	11.88 (28.67)	16.91 (27.74)	7.06 (30.49)	9.85 [3.49]	14.14 (26.10)	6.67 (27.62)	7.47 [4.00]	14.95 (31.59)	5.66 (29.22)	9.29 [3.03]	16.81 (35.01)	7.27 (27.51)	9.54 [2.53]
U.K.	15.33 (28.62)	17.87 (30.03)	13.25 (27.94)	4.62 [1.08]	17.46 (32.32)	14.81 (27.00)	2.65 [0.83]	18.41 (35.11)	14.51 (26.55)	3.89 [0.85]	15.89 (32.18)	12.99 (26.32)	2.90 [0.72]
France	11.26 (32.35)	17.10 (36.60)	9.46 (30.88)	7.64 [2.08]	15.68 (37.05)	8.70 (32.35)	6.98 [2.16]	16.17 (36.92)	9.30 (31.26)	6.86 [2.29]	15.12 (30.06)	6.25 (33.16)	8.88 [2.48]
Germany	9.88 (31.36)	12.77 (30.35)	10.01 (32.75)	2.75 [0.92]	11.13 (24.62)	10.58 (34.82)	0.55 [0.14]	13.28 (29.05)	5.14 (26.94)	8.13 [2.62]	9.99 (24.88)	10.42 (34.42)	-0.43 [-0.10]
Italy	8.11 (43.77)	5.45 (35.53)	11.44 (50.65)	-5.99 [-0.91]	7.62 (42.36)	12.99 (54.68)	-5.37 [-0.84]	11.05 (43.52)	0.37 (38.42)	10.69 [1.73]	10.07 (38.28)	12.68 (56.66)	-2.61 [-0.33]
Netherlands	13.30 (18.81)	15.77 (33.07)	13.47 (21.01)	2.30 [0.44]	14.37 (21.07)	9.26 (20.48)	5.11 [1.04]	11.66 (33.02)	11.84 (23.26)	-0.19 [-0.03]	13.47 (21.38)	13.05 (30.81)	0.41 [0.07]
Belgium	12.62 (25.88)	14.90 (28.62)	10.51 (27.63)	4.39 [1.99]	15.12 (30.47)	12.90 (27.88)	2.22 [0.78]	16.46 (28.84)	12.03 (25.57)	4.44 [1.27]	15.16 (26.47)	12.26 (29.26)	2.91 [1.29]
Switzerland	11.07 (27.21)	13.84 (30.00)	10.34 (28.57)	3.49 [0.80]	12.59 (31.44)	11.04 (28.81)	1.54 [0.36]	12.32 (36.58)	9.78 (27.82)	2.53 [0.41]	12.62 (31.00)	10.44 (27.83)	2.18 [0.63]
Sweden	12.44 (24.91)	20.61 (38.31)	12.59 (26.26)	8.02 [1.16]	20.61 (42.43)	12.42 (24.76)	8.19 [1.03]	17.08 (30.56)	12.50 (23.58)	4.58 [0.90]	16.15 (29.55)	11.32 (25.13)	4.83 [1.05]
Australia	8.92 (26.31)	17.62 (31.03)	5.30 (27.32)	12.32 [2.41]	15.64 (28.19)	5.97 (28.89)	9.67 [1.71]	18.32 (29.08)	4.03 (27.46)	14.29 [2.85]	14.62 (28.43)	6.83 (28.57)	7.79 [1.65]
Hong Kong	22.52 (41.96)	26.51 (48.68)	19.35 (40.21)	7.16 [1.35]	27.04 (44.83)	22.05 (40.81)	4.99 [0.82]	29.33 (46.24)	20.24 (42.72)	9.09 [1.37]	23.66 (38.76)	23.30 (42.05)	0.35 [0.09]
Singapore	13.31 (27.29)	21.63 (36.89)	11.96 (27.71)	9.67 [2.36]	15.21 (29.55)	13.12 (34.68)	2.09 [0.65]	13.42 (26.24)	8.03 (28.92)	5.39 [1.49]	10.64 (22.01)	13.10 (33.93)	-2.46 [-0.45]

1980

The Journal of Finance

Past Return Anomalies

- The literature has documented short term reversals, intermediate term momentum, and long term reversals.
- 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
- DeBondt and Thaler (1985, 1987) document long run reversals
- Momentum is the most heavily explored past return anomaly.
- Several studies document momentum robustness.
- Others document momentum interactions with firm, industry, and market level variables.
- There is solid evidence on momentum crashes following recovery from market downturns.
- More recent studies argue that momentum is attributable to the short-leg of the trade – and it difficult to implement in real time as losers stocks are difficult to short sale and arbitrage.

From Momentum Robustness to Momentum Crash

- 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 tends to strengthen, rather than discounts, momentum profitability.
- This is because momentum loads negatively on the market, size, and value factors.
- Momentum also seems to appear in bonds, currencies, commodities, as well as mutual funds and hedge funds.
- As Asness, Moskowitz, and Pedersen (2013) note: Momentum and value are everywhere.
- 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), Rouwenhorst (1998), and Titman and Wei (2010) document momentum in international markets (*not in Japan*).

From Momentum Robustness to Momentum Crash

- Korajczyk and Sadka (2004) find that momentum survives trading costs, whereas Avramov, Chordia, and Goyal (2006a) show that the profitability of short-term reversal disappears in the presence of trading costs.
- Fama and French (2008) show that momentum is among the few robust anomalies – it works also among large cap stocks.
- Geczy and Samonov (2013) examine momentum during the period 1801 through 1926 – probably the world's longest back-test.
- Momentum had been fairly robust in a cross-industry analysis, cross-country analysis, and cross-style analysis.
- The prominence of momentum has generated both behavioral and rational theories.
- Behavioral: Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), and Hon, Lim, and Stein (2000).
- Rational: Berk, Green, and Naik (1999), Johnson (2002), and Avramov and Hore (2017).

Momentum Crash

- In 2009, momentum delivers a negative 85% payoff.
- The negative payoff is attributable to the short side of the trade.
- Loser stocks had forcefully bounced back.
- Other episodes of momentum crashes were recorded.
- The down side risk of momentum can be immense.
- Daniel and Moskowitz (2017) is a good empirical reference while Avramov and Hore (2017) give theoretical support.
- In addition, both Stambaugh, Yu, and Yuan (2012) and Avramov, Chordia, Jostova, and Philipov (2007, 2013) show that momentum is profitable due to the short-leg of the trade.
- Based on these studies, loser stocks are difficult to short and arbitrage – hence, it is really difficult to implement momentum in real time.
- In addition, momentum does not work over most recent years.

Momentum Interactions

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

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, 2013) document that momentum concentrates in low rated stocks. Moreover, the credit risk effect seems to dominate the other interaction effects.

Potential industry-level interactions

- Moskowitz and Grinblatt (1999) show that industry momentum subsumes stock level momentum. That is, buy the past winning industries and sell the past losing industries.
- Grundy and Martin (2001) find no industry effects in momentum.

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 on the market performance over three years prior to the implementation of the momentum strategy.

Market sentiment

- Antoniou, Doukas, and Subrahmanyam (2010) and Stambaugh, Yu, and Yuan (2012) find that the momentum effect is stronger when market sentiment is high.
- The former paper suggests that this result is consistent with the slow spread of bad news during high-sentiment periods.
- Stambaugh, Yu, and Yuan (2015) use momentum along with ten other anomalies to form a stock level composite overpricing measure. For instance, loser stocks are likely to be overpriced due to impediments on short selling.

Other interactions at the aggregate level

- Chordia and Shivakumar (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.
- Avramov, Cheng, and Hameed (2016) show that momentum payoffs vary with market illiquidity - in contrast to “limits to arbitrage” momentum is profitable during highly liquid markets.

Momentum in Anomalies

- Avramov et al (Scaling Up Market Anomalies 2017) show that one could implement momentum among top and bottom anomaly portfolios.
- They consider 15 market anomalies, each of which is characterized by the anomaly conditioning variable., e.g., gross profitability, IVOL, and dispersion in analysts earnings forecast.
- There are 15 top (best performing long-leg) portfolios.
- There are 15 bottom (worst performing short-leg) portfolios.
- The trading strategy involves buying a subset (e.g., five) top portfolios and selling short a subset of bottom portfolios based on past one-month return or based on expected return estimated from time-series predictive regressions.
- Implementing momentum among anomalies delivers a robust performance even during the post-2000 period and during periods of low market sentiment.

Momentum Spillover from Stocks to Bonds

- Gebhardt, Hvidkjaer, and Swaminathan (2005) examine the interaction between momentum in the returns of equities and corporate bonds.
- They find significant evidence of a momentum spillover from equities to corporate bonds of the same firm.
- In particular, firms earning high (low) equity returns over the previous year earn high (low) bond returns in the following year.
- The spillover results are stronger among firms with lower-grade debt and higher equity trading volume.
- Beyond momentum spillover, Jostova et al (2013) find significant price momentum in US corporate bonds over the 1973 to 2008 period.
- They show that bond momentum profits are significant in the second half of the sample period, 1991 to 2008, and amount to 64 basis points per month.

Are there Predictable Patterns in Corporate Bonds?

- For the most part, anomalies that work on equities also work on corporate bonds.
- In addition, the same-direction mispricing applies to both stocks and bonds.
- See, for example, Avramov, Chordia, Jostova, and Philipov (2018).
- They document overpricing in stocks and the *corresponding* corporate bonds.
- Indeed, structural models of default, such as that originated by Merton (1974), impose a tight relation between equity and bond prices, as both are claims on the same firm assets.
- Then, if a characteristic x is able to predict stock returns, it must predict bond returns.
- On one hand, the empirical question is thus whether bond returns are over-predictable or under-predictable for a given characteristic.
- On the other hand, structural models of default have had difficult times to explain credit spreads and moreover bond and stock markets may not be integrated.
- Also, some economic theory claims that there might be wealth transfer from bond holders to equity holders – thus, one may suspect that equity overpricing must be followed by bond underpricing.

Time-Series Momentum

- Time-series momentum in an absolute strength strategy, while the price momentum is a relative strength one. Here, one takes long positions in those stocks having positive expected returns and short positions in stocks having negative expected returns, where expected return is assessed based on the following equation from Moskowitz, Ooi, and Pedersen (2012):

$$r_t^s / \sigma_{t-1}^s = \alpha + \beta_h r_{t-h}^s / \sigma_{t-h-1}^s + \varepsilon_t^s$$

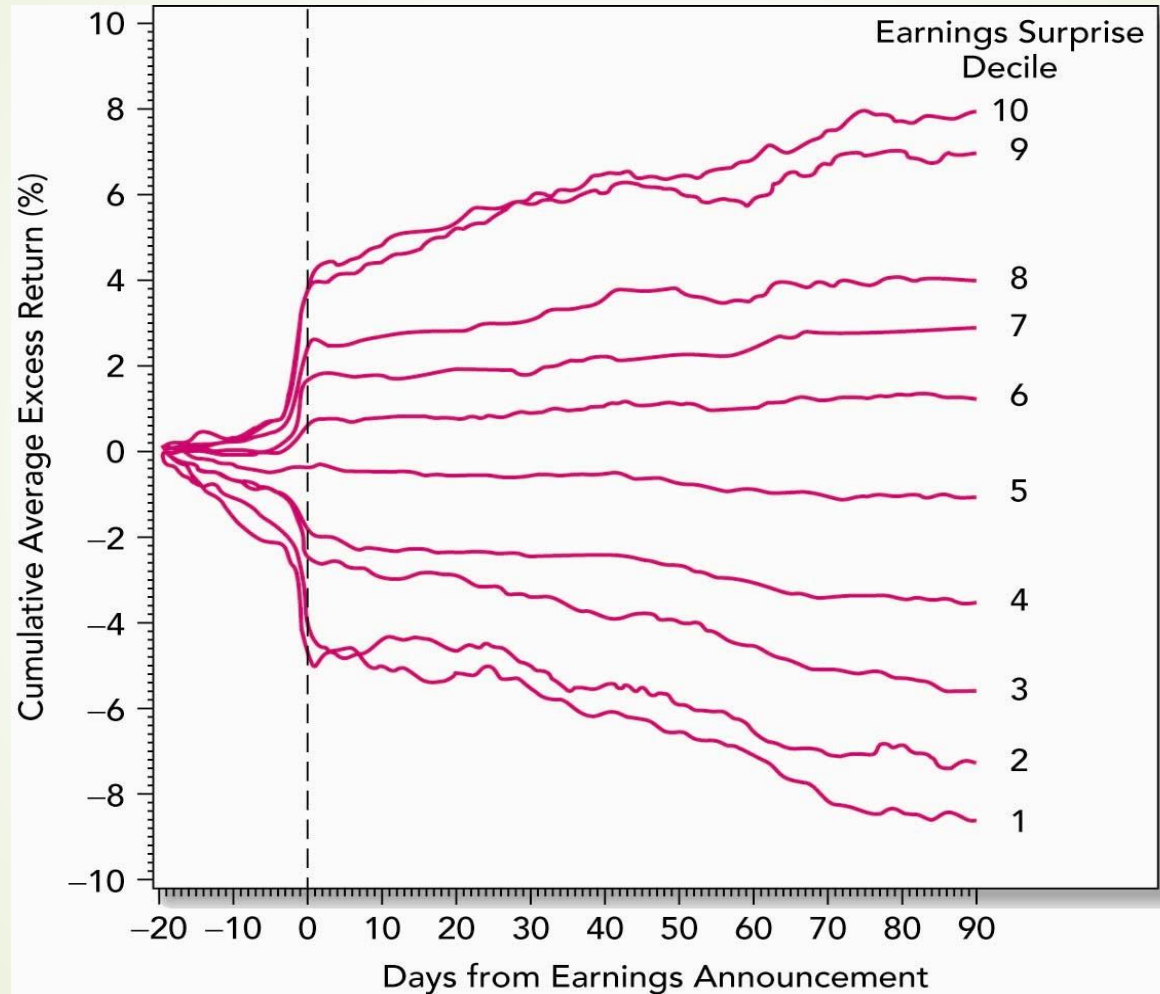
Earnings Momentum (see also next page)

- 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.

Revenue Momentum

- Chen, Chen, Hsin, and Lee (2010) study the inter-relation between price momentum, earnings momentum, and revenue momentum, concluding that it is ultimately suggested to combine all types rather than focusing on proper subsets.

Earnings Momentum: under-reaction?



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.

Capital Investment

- Titman, Wei, and Xie (2004) document a negative relation between capital investments and returns.
- 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 assets 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.

Idiosyncratic volatility (IVOL)

- Ang, Hodrik, Xing, and Zhang (2006, 2009) show negative cross section relation between IVOL and average return in both US and global markets.
- The AHXZ proxy for IVOL is the standard deviation of residuals from time-series regressions of excess stock returns on the Fama-French factors.

Counter intuitive relations

- The forecast dispersion, credit risk, betting against beta, and IV effects apparently violate the risk-return tradeoff.
- Investors seem to pay premiums for purchasing higher risk stocks.
- Intuition may suggest it should be the other way around.
- Avramov, Chordia, Jostova, and Philipov (2013, 2018) provide a plausible story: institutional and retail investors underestimate the severe implications of financial distress.
- Thus, financially distressed stocks (and bonds) are overpriced.
- As financially distressed firms exhibit high IVOL, high beta, high credit risk, and high dispersion – all the counter intuitive relations are explained by the overpricing of financially distressed stocks.

Return on Assets (ROA)

- Fama and French (2006) find that more profitable firms (ROA) have higher expected returns than less profitable firms.
- ROA is typically measured as income before extraordinary items divided by one quarter lagged total assets.

Quality Investing

- Novy-Marks describes seven of the most widely used notions of quality:
 - Sloan's (1996) accruals-based measure of earnings quality (coming next)
 - Measures of information uncertainty and financial distress (coming next)
 - Novy-Marx's (2013) gross profitability (coming next)
 - Piotroski's (2000) F-score measure of financial strength (coming next)
 - Graham's quality criteria from his "Intelligent Investor" (appendix)
 - Grantham's "high return, stable return, and low debt" (appendix)
 - Greenblatt's return on invested capital (appendix)

Quality investing: Accruals, Information uncertainty, and Distress

- **Accruals:** Sloan (1996) shows that firms with high accruals earn abnormal lower returns on average than firms with low accruals. Sloan suggests that investors overestimate the persistence of the accrual component of earnings when forming earnings expectations. Total accruals are calculated as changes in noncash working capital minus depreciation expense scaled by average total assets for the previous two fiscal years.
- **Information uncertainty:** Diether, Malloy, and Scherbina (2002) suggest that firms with high dispersion in analysts' earnings forecasts earn less than firms with low dispersion. Other measures of information uncertainty: firm age, cash flow volatility, etc.
- **Financial distress:** As noted earlier, Campbell, Hilscher, and Szilagyi (2008) find that firms with high failure probability have lower, not higher, subsequent returns. Campbell, Hilscher, and Szilagyi suggest that their finding is a challenge to standard models of rational asset pricing. The failure probability is estimated by a dynamic logit model with both accounting and equity market variables as explanatory variables. Using Ohlson (1980) O-score as the distress measure yields similar results. Avramov, Chordia, Jostova, and Philipov (2009) use credit ratings as a proxy for financial distress and also document the same phenomenon: higher credit rating firms earn higher returns than low credit rating firms.

Quality investing: 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.
- Novy-Marx argues that gross profits scaled by assets is the cleanest accounting measure of true economic profitability. The further down the income statement one goes, the more polluted profitability measures become, and the less related they are to true economic profitability.

Quality investing: F-Score

- The F-Score is due to Piotroski (2000).
- It 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 score firms exhibit poor fundamentals along these same dimensions.
- F-Score is computed as the sum of nine components which are either zero or one.
- It thus ranges between zero and nine, where a low (high) score represents a firm with very few (many) good signals about its financial conditions.

Illiquidity

- Illiquidity is not considered to be an anomaly.
- However, it is related to the cross section of average returns (as well as the time-series)
- Amihud (2002) proposes an illiquidity measure which is theoretically appealing and does a good job empirically.
- The Amihud measure is given by:

$$ILLIQ_{i,t} = \frac{1}{D_{i,t}} \sum_{t=1}^{D_{i,t}} \frac{|R_{itd}|}{DVOL_{itd}}$$

where: $D_{i,t}$ is the number of trading days in the month, $DVOL_{itd}$ is the dollar volume, R_{itd} is the daily return

- The illiquidity variable measures the price change per a unity volume.
- Higher change amounts to higher illiquidity

The turnover effect

- Higher turnover is followed by lower future return. See, for example, Avramov and Chordia (2006a).
- Swaminathan and Lee (2000) find that high turnover stocks exhibit features of high growth stocks.
- Turnover can be constructed using various methods. For instance, 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.

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 economic links generates about 1.5% per month.

Corporate Anomalies

- The corporate finance literature has documented a host of other interesting anomalies:
 - Stock Split
 - Dividend initiation and omission
 - Stock repurchase
 - Spinoff
 - Merger arbitrage
 - The long horizon performance of IPO and SEO firms.
- Finance research has documented negative relation between transactions of external financing and future stock returns: 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.
- See also discussion in the appendix.

Are anomalies pervasive?

- It is instructive to explore whether market anomalies are really pervasive.
- The evidence tilts towards the NO answer.
- 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.
- Schwert (2003) shows that anomalies (time-series and cross section) disappear or get attenuated following their discovery.
- Avramov, Chordia, and Goyal (2006) show that implementing short term reversal strategies yields profits that do not survive direct transactions costs and costs related to market impact.
- Wang and Yu (2010) find that the return on asset (ROA) anomaly exists primarily among firms with high arbitrage costs and high information uncertainty.
- Avramov, Chordia, Jostova, and Philipov (2007a,b, 2013, 2018) show that momentum, dispersion, credit risk, among many other effects, concentrate in a very small portion of high credit risk stocks and only during episodes of firm financial distress.
- In particular, investors tend to overprice distressed stocks. Moreover, distressed stocks display extreme values of firm characteristics – high IVOL, high dispersion, large negative past returns, and large negative earnings surprises. They are thus placed at the short-leg of anomaly portfolios. Anomaly profitability emerges only from the short-leg of a trade, as overpricing is corrected.

Are anomalies pervasive?

- Chordia, Subrahmanyam, and Tong (2014) and McLean and Pontiff (2014) find that several anomalies have attenuated significantly over time, particularly in liquid NYSE/AMEX stocks, and virtually none have significantly accentuated.
- Stambaugh, Yu, and Yuan (2012) associate anomalies with market sentiment.
- Following Miller (1977), there might be overpriced stocks due to costly short selling.
- As overpricing is prominent during high sentiment periods, anomalies are profitable only during such episodes and are attributable to the short-leg of a trade.
- Avramov, Chordia, Jostova, and Philipov (2013) and Stambaugh, Yu, and Yuan (2012) seem to agree that anomalies represent an un-exploitable stock overvaluation.
- But the sources are different: market level sentiment versus firm-level credit risk.
- Avramov, Chordia, Jostova, and Philipov (2018) reconcile the evidence: anomalies concentrate in the intersection of firm credit risk and market-wide sentiment.
- And the same mechanism applies for both stocks and corporate bonds.
- Beyond Miller (1977), there are other economic theories that permit overpricing.

Are anomalies pervasive?

- For instance, the Harrison and Kreps (1978) basic insight is that when agents agree to disagree and short selling is impossible, asset prices may exceed their fundamental value.
- The positive feedback economy of De Long, Shleifer, Summers, and Waldmann (1990) also recognizes the possibility of overpricing — arbitrageurs do not sell or short an overvalued asset, but rather buy it, knowing that the price rise will attract more feedback traders.
- Garlappi, Shu, and Yan (2008) and Garlappi and Yan (2011) argue that distressed stocks are overvalued due to shareholders' ability to extract value from bondholders during bankruptcy.
- Kumar (2009), Bailey, Kumar, and Ng (2011), and Conrad, Kapadia, and Xing (2014) provide support for lottery-type preferences among retail investors. Such preferences can also explain equity overpricing.
- Lottery-type stocks are stocks with low price, high idiosyncratic volatility, and positive return skewness.
- The idea of skewness preferring investors goes back to Barberis and Huang (2008) who build on the prospect theory of Kahneman and Tversky (1979) to argue that overpricing could prevail as investors overweight low-probability windfalls.
- Notice, however, that Avramov, Chordia, Jostova, and Philipov (2018) find that bonds of overpriced equity firms are also overpriced, thus calling into question the transfer of wealth hypothesis.
- Also, the upside potential of corporate bonds is limited relative to that of stocks — thus lottery-type preferences are less likely to explain bond overpricing.
- In sum, anomalies do not seem to be pervasive. They could emerge due to data mining, they typically characterize the short-leg of a trade, they concentrate in difficult to short and arbitrage stocks, and they might fail survive reasonable transaction costs.

Could anomalies also emerge from the long-leg?

- Notably, some work does propose the possibility of asset underpricing.
- Theoretically, in Diamond and Verrecchia (1987), investors are aware that, due to short sale constraints, negative information is withheld, so individual stock prices reflect an expected quantity of bad news. Prices are correct, on average, in the model, but individual stocks can be overvalued or undervalued.
- Empirically, Boehmer, Huszar, and Jordan (2010) show that low short interest stocks exhibit positive abnormal returns. Short sellers avoid those apparently underpriced stocks
- Also, the 52-week high anomaly tells you that stocks that are near their 52-week high are underpriced.
- Recently, Avramov, Kaplanski, and Subrahmanyam (2018) show that a ratio of short (fast) and long (slow) moving averages predict both the long and short legs of trades.
- The last two papers attribute predictability to investor's under-reaction due to the anchoring bias.
- Avramov, Kaplanski, and Subrahmanyam (2018) show theoretically why anchoring could result in positive autocorrelation in returns.
- The anchoring bias is the notion that agents rely too heavily on readily obtainable (but often irrelevant) information in forming assessments (Tversky and Kahneman, 1974).
- As an example of the anchoring bias, in Ariely, Loewenstein, and Prelec (2003), participants are asked to write the last two digits of their social security number and then asked to assess how much they would pay for items of unknown value. Participants having lower numbers bid up to more than double relative to those with higher numbers, indicating that they anchor on these two numbers.
- Such under-reaction could be long lasting as shown by Avramov, Kaplanski, and Subrahmanyam (2018).

Market Anomalies: Polar Views

- Scholars like Fama would claim that the presence of anomalies merely indicates the inadequacy of the CAPM.
- Per Fama, an alternative risk based model would capture all anomalous patterns in asset prices. Markets are in general efficient and the risk-return tradeoff applies. The price is right up to transaction cost bounds.
- Scholars like Shiller would claim that asset prices are subject to behavioral biases.
- Per Shiller, asset returns are too volatile to be explained by changing fundamental values and moreover higher risk need not imply higher return.
- Both Fama and Shiller won the Nobel Prize in Economics in 2013.
- Fama and Shiller represent polar views on asset pricing: rational versus behavioral.
- But whether or not markets are efficient seems more like a philosophical question.
- In his presidential address, Cochrane (2011) nicely summarizes this debate. See next page.

Rational versus Behavioral perspectives

- It is pointless to argue “rational” vs. “behavioral.”
- There is a discount rate and equivalent distorted probability that can rationalize any (arbitrage-free) data.
- “The market went up, risk aversion must have declined” is as vacuous as “the market went up, sentiment must have increased.” Any model only gets its bite by restricting discount rates or distorted expectations, ideally tying them to other data.
- The only thing worth arguing about is how persuasive those ties are in a given model and dataset.
- And the line between recent “exotic preferences” and “behavioral finance” is so blurred, it describes academic politics better than anything substantive.
- For example, which of Epstein and Zin (1989), Barberis, Huang, and Santos (2001), Hansen and Sargent (2005), Laibson (1997), Hansen, Heaton and Li (2008), and Campbell and Cochrane (1999) is really “rational” and which is really “behavioral?”
- Changing expectations of consumption 10 years from now (long run risks) or changing probabilities of a big crash are hard to tell from changing “sentiment.”

Rational versus Behavioral perspectives

- ▶ Yet another intriguing quote followed by a response.
- ▶ Cochrane (2011): Behavioral ideas - narrow framing, salience of recent experience, and so forth - are good at generating anomalous prices and mean returns in individual assets or small groups. They do not easily generate this kind of coordinated movement across all assets that looks just like a rise in risk premium. Nor do they naturally generate covariance. For example, “extrapolation” generates the slight autocorrelation in returns that lies behind momentum. But why should all the momentum stocks then rise and fall together the next month, just as if they are exposed to a pervasive, systematic risk?
- ▶ Kozak, Nagel, and Stantosh (KNS 2017a): The answer to this question could be that some components of sentiment-driven asset demands are aligned with covariances with important common factors, some are orthogonal to these factor covariances. Trading by arbitrageurs eliminates the effects of the orthogonal asset demand components, but those that are correlated with common factor exposures survive because arbitrageurs are not willing to accommodate these demands without compensation for the factor risk exposure.

Theoretical Drawbacks of the CAPM

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

- But eventual wealth could emerge from both investment, labor, and entrepreneurial incomes.
- Additional factors are 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, pro-cyclical stocks should offer higher average returns than countercyclical stocks, even if both stocks have the same market beta.
- Put another way, co-variation with recessions seems to matter in determining expected returns.
- You may correctly argue that the market tends to go down in recessions.
- Yet, recessions tend to be unusually severe or mild for a given level of market returns.

ICAPM

B. The CAPM assumes a static one-period model.

- Merton (1973) introduces a multi-period version of the CAPM - the inter-temporal CAPM (ICAPM).
- In ICAPM, 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.
- In ICAPM, an asset's risk should be measured via its covariance with the marginal utility of investors, and such covariance could be different from 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 inter-temporal models, equilibrium expected returns on risky assets may differ from the riskless rate even when they have no systematic (market) risk.
- But the ICAPM does not tell us which state variables are priced - this gives license to fish factors that work well in explaining the data albeit void any economic content.

Conditional CAPM

C. The CAPM is an unconditional model.

- Avramov and Chordia (2006a) show that various conditional versions of the CAPM do not explain anomalies.
- LN (2006) provide similar evidence yet in a quite different setup.
- LN nicely illustrate the distinct differences between conditional and unconditional efficiency.
- In particular, it is known from Hansen and Richards (1987) that a portfolio could be efficient period by period (conditional efficiency) but not unconditionally efficient.
- Here are the details (I try to follow LN notation):
- Let R_{it} be the excess return on asset i and let R_{Mt} be excess return on the market portfolio.
- Conditional moments for period t given $t-1$ are labeled with a t -subscript.
- The market conditional risk premium and volatility are γ_t and σ_t and the stock's conditional beta is β_t .
- The corresponding unconditional moments are denoted by γ , σ_M , and β^u
- Notice: $\beta \equiv E(\beta_t) \neq \beta^u$

Conditional CAPM

- The conditional CAPM states that at every time t the following relation holds:

$$E_{t-1}(R_t) = \beta_t \gamma_t$$

- Taking expectations from both sides

$$\begin{aligned} E(R_t) &= E(\beta_t)E(\gamma_t) + \text{cov}(\beta_t, \gamma_t) \\ &= \beta \gamma + \text{cov}(\beta_t, \gamma_t) \end{aligned}$$

- Notice that the unconditional alpha is defined as

$$\alpha^u = E(R_t) - \beta^u \gamma$$

where

$$\gamma = E(\gamma_t)$$

- Thus

$$\alpha^u = \beta \gamma + \text{cov}(\beta_t, \gamma_t) - \beta^u \gamma$$

$$\alpha^u = \gamma(\beta - \beta^u) + \text{cov}(\beta_t, \gamma_t)$$

Conditional CAPM

➤ Now let $\beta_t = \beta + \eta_t$.

➤ Conditional CAPM, $R_{it} = \beta_t R_{Mt} + \varepsilon_t$

$$= \beta R_{Mt} + \eta_t R_{Mt} + \varepsilon_t$$

➤ The unconditional covariance between R_{it} and R_{Mt} equals

$$\begin{aligned} \text{cov}(R_{it}, R_{Mt}) &= \text{cov}(\beta R_{Mt} + \eta_t R_{Mt} + \varepsilon_t, R_{Mt}) \\ &= \beta \sigma_M^2 + \text{cov}(\eta_t R_{Mt} + \varepsilon_t, R_{Mt}) \\ &= \beta \sigma_M^2 + E(\eta_t R_{Mt}^2) - E(\eta_t R_{Mt})E(R_{Mt}) \\ &= \beta \sigma_M^2 + \text{cov}(\eta_t, R_{Mt}^2) - \gamma \text{cov}(\eta_t, R_{Mt}) \\ &= \beta \sigma_M^2 + \text{cov}(\eta_t, \sigma_t^2) + \text{cov}(\eta_t, \gamma_t^2) - \gamma \text{cov}(\eta_t, \gamma_t) \\ &= \beta \sigma_M^2 + \text{cov}(\eta_t, \sigma_t^2) + \gamma \text{cov}(\eta_t, \gamma_t) + \text{cov}[\eta_t, (\gamma_t - \gamma)^2] \end{aligned}$$

Conditional CAPM

- In the end, we get

$$\beta^u = \beta + \frac{\gamma}{\sigma_M^2} \text{cov}(\beta_t, \gamma_t) + \frac{1}{\sigma_M^2} \text{cov}[\beta_t, (\gamma_t - \gamma)^2] + \frac{1}{\sigma_M^2} \text{cov}(\beta_t, \sigma_t^2)$$

So β^u differs from $E(\beta_t)$ if

- β_t covaries with γ_t
- β_t covaries with $(\gamma_t - \gamma)^2$
- β_t covaries with σ_t^2
- The stock unconditional alpha is

$$\alpha^u = \left[1 - \frac{\gamma^2}{\sigma_M^2}\right] \text{cov}(\beta_t, \gamma_t) - \frac{\gamma}{\sigma_M^2} \text{cov}[\beta_t, (\gamma_t - \gamma)^2] - \frac{\gamma}{\sigma_M^2} \text{cov}(\beta_t, \sigma_t^2)$$

- Notice that even if the conditional CAPM holds exactly we should expect to find deviations from the unconditional CAPM if **any** of the three covariance terms is nonzero.
- LN: if the conditional CAPM holds α^u should be rather small, apparently at odds with market anomalies.

D. Perhaps Multifactor Models?

- The poor empirical performance of the single factor CAPM motivated a search for multifactor models.
- Multiple factors have been inspired along the spirit of
 - The Arbitrage Pricing Theory — APT — (1976) of Ross
 - The inter-temporal CAPM (ICAPM) of Merton (1973).
- Distinguishing between the APT and ICAPM is often confusing.
- Cochrane (2001) argues that the biggest difference between APT and ICAPM 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

Multifactor Models

- 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 long in small firm stocks and short in large firm stocks (SMB)
 3. a spread portfolio long in high book-to-market stocks and short in low book-to-market stocks (HML)
- FF (1996) have shown that their model is able to explain many of the cross sectional effects known back then - excluding momentum.
- But meanwhile many new effects have been discovered that the FF-model fails to explain.
- FF (1993) argue that their factors are state variables in an ICAPM sense.
- Liew and Vassalou (2000) make a good case for that claim: they find that the FF factors forecast GDP growth
- But the FF model is empirically based while it voids any theoretical underpinning
- Moreover, the statistical tests promoting the FF model are based on 25 size book to market portfolios that already obey a factor structure, while results are less favorable focusing on industry portfolios or individual securities.
- Factor structure means that the first three eigen vectors of the covariance matrix of returns display similar properties to the market, size, and value factors. So perhaps nothing is really special about the FF model.

Multifactor Models

- The FF model is also unable to explain the IVOL effect, the credit risk effect, the dispersion effect, earnings momentum, net equity issues (net equity issued less the amount of seasoned equity retired), among many others.
- Out-of-sample, the FF model performs poorly.
- In fact, factor models typically do not perform well out-of-sample.
- Models based on cross section regressions with firm characteristics perform better (see, e.g., Haugen and Baker (2006) and the recently developing machine learning methods in finance) possibly due to estimation errors.
- In particular, in time-series asset pricing regressions, N times K factor loadings are estimated in addition to K risk premiums, while in cross section regressions only M slope coefficients, where N is the number of test assets, K is the number of factors, and M is the number of firm characteristics.
- Cross-section regressions thus require a smaller number of estimates.
- Shrinkage methods (e.g., Ridge and Lasso) attempt to improve the estimation of cross section regressions.
- Indeed, cross section regression coefficients are still estimated with errors and their computation implicitly requires the estimation of the inverse covariance matrix of all predictors, whose size grows quadratically with the number of firm characteristics.
- Moreover, firm characteristics are typically highly correlated – thus the regression suffers from the multi-collinearity problem.

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.
- Until 2003 we had five major factors to explain equity returns
 1. market
 2. SMB
 3. HML
 4. WML
 5. Liquidity

Multifactor Models

- Often bond portfolios such as the default risk premium and the term premium are also added (need to distinguish between risk premiums and yield spreads).
- Fama and French (2015) propose a five-factor model based on the original market, size, and book-to-market factors and adds investment and profitability factors.
- Hou, Xue, and Zhang (2015) propose four-factors: market, size, investment, and profitability.
- Both studies provide theoretical motivations for why these factors contain information about expected return.
- Hou, Xue, and Zhang (2015) rely on an investment-based pricing model, while Fama and French (2015) invoke comparative statics of a present-value relation.
- Stambaugh and Yuan (2016) propose two mispricing factors based on 11 anomalies studied in Stambaugh, Yu, and Yuan (2012).
- Avramov, Cheng, and Hameed (2018) employ the Stambaugh-Yuan factor in understanding performance of mutual funds.
- Controlling for this benchmark eliminates alphas of mutual funds that hold mispriced stocks.

What if factors are not pre-specified? The APT

- Chen, Roll, and Ross (1986) study pre-specified factors, presumably motivated by the APT.
- However, the APT is mostly silent on the return deriving factors.
- Considering latent (as opposed to pre-specified) factors is the basic tenet of APT.
- The APT is appealing as it requires minimal set of assumptions: that there are many assets, that trading is costless, and that a factor model drives returns.
- To analyse the model empirically, however, one must impose additional structure.
- First, as Shanken (1982) emphasizes, obtaining an exact rather than approximate factor pricing relation requires an assumption about market equilibrium.
- Second, some assumptions that ensure statistical identification are necessary.
- One possibility is to assume that returns are Gaussian, that their co-variances are constant, and that all co-movement in asset returns can be attributed to factor movements.
- Given these restrictions, it is possible to use maximum likelihood factor analysis to estimate factor loadings.

What if factors are not pre-specified? The APT

- Roll and Ross (1980) used these loadings to test exact APT pricing with constant factor risk premiums using simple cross-sectional regression tests.
- Lehmann and Modest (1988) use a more sophisticated factor decomposition algorithm to consider much larger cross-sections of returns under the same assumptions.
- Extending the results of Chamberlain and Rothschild (1982), Connor and Korajczyk (1986) introduced a novel method for factor extraction, which they called asymptotic principal components.
- Notice that asymptotic is with respect to the number of stocks, not time-series.
- The procedure allows for non-Gaussian returns.
- The central convergence result of CK states that given a large enough set of assets returns whose residuals are sufficiently uncorrelated, the realizations, over a fixed time period, of the unobserved factors (up to a non-singular translation) may be recovered to any desired precision.
- Jones (2001) extends CK to the case of heteroskedastic asset returns.
- Details on extracting latent factors are provided below.

Extracting latent factors

- CK assume countably infinite set of assets.
- We observe R^N , the $N \times T$ matrix of excess returns on the first N assets in the economy.
- We can write

$$R^N = B^N H + E^N$$

where B^N is the $N \times K$ matrix of factor loadings, H is a $K \times T$ matrix of factor risk premiums, and E^N contains the $N \times T$ regression residuals.

- Notice that

$$\begin{aligned} \frac{1}{N} R^{N'} R^N &= \frac{1}{N} H' B^{N'} B^N H + \frac{1}{N} H' B^{N'} E^N + \frac{1}{N} E^{N'} B^N H + \frac{1}{N} E^{N'} E^N \\ &= X^N + Y^N + Y^{N'} + Z^N \end{aligned}$$

- CK assume that $\frac{1}{N} B^{N'} B^N$ has a probability limit M , implying that $X^N \rightarrow H' M H$
- As the residual terms have zero means and are also serially uncorrelated, Y^N and $Y^{N'}$ have probability limits equal to zero.

Extracting latent factors

- The non-serial correlation and homoscedastic assumptions imply that there exists an average residual variance $\bar{\sigma}^2$ that is constant through time.

- Taking together all assumptions, we get

$$\frac{1}{N} R^{N'} R^N \rightarrow H' M H + \bar{\sigma}^2 I_T$$

or

$$\frac{1}{N} R^{N'} R^N \rightarrow F' F + \bar{\sigma}^2 I_T$$

- The K eigenvectors corresponding to the largest K eigenvalues of $\frac{1}{N} R^{N'} R^N$ are the latent factors.
- Each of the extracted factors is a T -vector.
- Notice that replacing H by F in time-series regressions of excess returns on factors does not have any effect on alpha estimates and their significance.
- One can use such PCA to test exact asset pricing as well as assess performance (alpha) of mutual funds and hedge funds.

Extracting latent factors

- ▶ Jones (2001) accounts for asset return heteroscedasticity but the non-serial correlation assumption is still preserved.
- ▶ Then $\frac{1}{N} R^{N'} R^N$ still converges to $H' M H + D$ while D is a $T \times T$ diagonal matrix with non equal diagonal entries.
- ▶ Put another way, the average idiosyncratic variance can freely change from one period to the next.
- ▶ Due to serially uncorrelated residuals Y^N and $Y^{N'}$ still have probability limits of zero.
- ▶ As $\frac{1}{N} R^{N'} R^N \rightarrow F' F + D$, it follows that $\frac{1}{N} D^{-\frac{1}{2}} R^{N'} R^N D^{-\frac{1}{2}} \rightarrow D^{-\frac{1}{2}} F' F D^{-\frac{1}{2}} + I_T$
- ▶ Or $\frac{1}{N} D^{-\frac{1}{2}} R^{N'} R^N D^{-\frac{1}{2}} \rightarrow Q' Q + I_T$
- ▶ By the singular value decomposition $Q = U' (\Lambda - I_K)^{\frac{1}{2}} V'$, where V contains the eigenvectors corresponding to the K largest eigenvalues of $\frac{1}{N} D^{-\frac{1}{2}} R^{N'} R^N D^{-\frac{1}{2}}$, and Λ contains the diagonal matrix of these eigenvalues in descending order.

Extracting latent factors

- Assuming $U = I_K$, we get $F' = QD^{\frac{1}{2}} = (\Lambda - I_K)^{\frac{1}{2}}V'D^{\frac{1}{2}}$.
- As D is unknown, Jones uses the iterative process:
 - Compute $C = \frac{1}{N}R^{N'}R^N$.
 - Guess an initial estimate of D , say \hat{D} .
 - Collect the K eigenvectors corresponding to the K largest eigenvalues of the matrix $\hat{D}^{-\frac{1}{2}}C\hat{D}^{-\frac{1}{2}}$.
 - Let Λ be the diagonal matrix with the K largest eigenvalues on the diagonal arranged in descending order, and let V denote the matrix of eigenvectors.
 - Compute an estimate of the factor matrix as
$$\tilde{F} = \hat{D}^{\frac{1}{2}}V(\Lambda - I_K)^{\frac{1}{2}}$$
 - Compute a new estimate \hat{D} as the diagonal of $C - \tilde{F}'\tilde{F}$.
 - Keep the iteration till convergence.

Understanding factor models

- Whether multi-factor models are based on pre-specified or latent factors, the stochastic discount factor (SDF) is represented as a function of a small number of portfolio returns.
- Such models are reduced-form because they are not derived from assumptions about investor beliefs, preferences, and technology that prescribe which factors should appear in the SDF.
- Reduced-form factor models in this sense also include theoretical models that analyze how cross-sectional differences in stocks' covariances with the SDF arise from firms' investment decisions.
- Berk, Green, and Naik (1999), Johnson (2002), Liu, Whited, and Zhang (2009), and Liu and Zhang (2014) belong into the reduced-form class because they make no assumptions about investor beliefs and preferences other than the existence of an SDF.
- These models show how firm investment decisions are aligned with expected returns in equilibrium, according to first-order conditions.
- But they do not give a clue about which types of beliefs, rational or otherwise, investors align their marginal utilities with asset returns through first-order conditions.
- Similarly, in the ICAPM (1973), the SDF is derived from the first-order condition of an investor who holds the market portfolio and faces exogenously given time-varying investment opportunities. This leaves open the question how to endogenously generate the time-variation in investment opportunities in a way that is consistent, in equilibrium, with the ICAPM investor's first-order condition and his choice to hold the market portfolio.

Understanding factor models

- In this context, researchers often take the view that a tight link between expected returns and factor loadings is consistent with rational rather than behavioral asset pricing.
- This view is also underlying arguments that a successful test or calibration of a reduced-form SDF provides a rational explanation of asset pricing anomalies.
- KNS (2017a), however, argue that the reduced-form factor model evidence does not help in discriminating between alternative hypotheses about investor beliefs.
- In particular, only minimal assumptions on preferences and beliefs of investors are required for a reduced-form factor model with a small number of factors to describe the cross-section of expected returns.
- These assumptions are consistent with plausible behavioral models of asset prices as much as they are consistent with rational ones.
- Thus, one cannot learn much about investor beliefs from the empirical evaluation of a reduced-form model.

Understanding factor models

- For test assets that are equity portfolios sorted on firm characteristics, the covariance matrix is typically dominated by a small number of factors.
- Then, the SDF can be represented as a function of these few dominant factors.
- Absence of “near arbitrage” opportunities implies that there are no investment opportunities with extremely high Sharp Ratios, which is to say that there are no substantial loadings on principal components with extremely low eigen values (see formal details on the next page).
- Hence, if assets have a small number of factors with large eigen values, then these factors must explain returns.
- Otherwise, near-arbitrage opportunities would arise, which would be implausible, even if one entertains the possibility that prices could be influenced substantially by irrational sentiment investors.
- This result is in the spirit of the Arbitrage Pricing Theory (APT) of Ross (1976).
- Ross (p. 354) suggests bounding the maximum squared Sharpe Ratio of any arbitrage portfolio at twice the squared SR of the market portfolio.
- Fama and French (1996) (p. 75) regard the APT as a rational pricing model.
- KNS disagree with this interpretation, as absence of near-arbitrage opportunities still leaves a lot of room for belief distortions to affect asset prices.
- In particular, belief distortions that are correlated with common factor covariances will affect prices, while belief distortions that are uncorrelated with common factor covariances will be neutralized by arbitrageurs who are looking for high-SR opportunities.

Understanding factor models

- The basic claim is that if a small number of factors dominate --- they have the largest eigenvalues --- then those factors must explain asset returns.

- To see why, consider the Hansen Jagannathan (1991) pricing kernel representation

$$M_t = 1 - b'(r_t - \mu)$$

where r_t is an N -vector of excess returns and b is the N -vector of pricing kernel coefficients.

- Imposing the asset pricing restriction $E(M_t r_t) = 0$, the pricing kernel can be represented as

$$M_t = 1 - \mu' V^{-1} (r_t - \mu),$$

where μ and V are the N -vector of mean excess returns and the $N \times N$ covariance matrix.

- That is, pricing kernel coefficients are weights of the mean-variance efficient portfolio.

- Notice that

$$\text{var}(M_t) = \mu' V^{-1} \mu$$

which is the highest admissible Sharpe ratio based on the N risky assets.

- Now let us express $V = Q \Lambda Q'$, where $Q = [q_1, \dots, q_N]$ is the collection of N principal components and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$ is the diagonal matrix of the corresponding eigenvalues.

- Then

$$\begin{aligned} V^{-1} &= Q \Lambda^{-1} Q' \\ &= \left(\frac{q_1 q_1'}{\lambda_1} + \frac{q_2 q_2'}{\lambda_2} + \dots + \frac{q_N q_N'}{\lambda_N} \right) \end{aligned}$$

Understanding factor models

- Assume further that the first PC is a level factor, or $q_1 = \frac{1}{\sqrt{N}} \mathbf{1}_N$.
- Check: $q_1' q_1 = 1$
- Moreover, to get $q_1' q_k = 0$ for $k = 2, \dots, N$ it must be the case that q_k is a combination of positive (long) and negative (short) entries.
- We get

$$\begin{aligned}\text{var}(M) &= \mu' V^{-1} \mu = \mu' Q \Lambda^{-1} Q' \mu \\ &= \frac{(\mu' q_1)^2}{\lambda_1} + \frac{(\mu' q_2)^2}{\lambda_2} + \dots + \frac{(\mu' q_N)^2}{\lambda_N} \\ &= \left(\frac{\mu_M}{\sigma_M} \right)^2 + N \text{var}(\mu_i) \sum_{k=2}^N \frac{\text{corr}(\mu_i, q_{ki})^2}{\lambda_k}\end{aligned}$$

where $\mu_M = \frac{1}{\sqrt{N}} q_1' \mu$ and $\sigma_M = \frac{\lambda_1}{N}$, since $\text{var}(q_{ik}) = 1 \ \forall k = 1, \dots, N$

- This expression for $\text{var}(M)$ is also the expression for the maximal Sharpe ratio.
- It shows that expected returns must line up with only the first few PCs --- otherwise dividing by a small eigenvalue leads to enormous squared Sharpe ratio.

Understanding factor models

- KNS show that if you extract $K = 1, \dots, 15$ factors, the maximum squared SR based on the extracted factors rises with K .
- However, out-of-sample --- it works very differently.
- That is, let R is the $T \times N$ matrix of asset returns based on the first part of the sample, and compute $f_1 = Rq_1, f_2 = Rq_2, \dots, f_K = Rq_K$.
- Then compute $\max SR^2 = \mu_F' V_F^{-1} \mu_F$ where μ_F and V_F are the mean vector and the covariance matrix of factors.
- Then apply q_1, \dots, q_K for out-of-sample R — the maximum SR^2 is much smaller
- Thus, mere absence of near arbitrage opportunities has limited economic content.
- For instance, the absence of near arbitrage opportunities could characterizes economics in which all cross-section variation in expected returns is attributable to sentiment.

Understanding Factor Models

- Let us revisit the HJ pricing kernel representation to reinforce the idea that only a small number of factors should explain asset returns.
- The pricing kernel is now represented using eigenvectors and eigenvalues:

$$\begin{aligned}M_t &= 1 - \mu' V^{-1} (r_t - \mu) \\ &= 1 - \mu' Q \Lambda^{-1} Q' (r_t - \mu)\end{aligned}$$

- Now, let $Q_t = Q' r_t$ and $\mu_Q = Q' \mu$, then

$$\begin{aligned}M_t &= 1 - \mu'_Q \Lambda^{-1} (Q_t - \mu_Q) \\ &= 1 - b'_Q (Q_t - \mu_Q)\end{aligned}$$

- The sample estimate of b_Q is given by $\hat{b}_Q = \hat{\Lambda}^{-1} \hat{\mu}_Q$.
- Assume that Λ is known, then $\text{var}(\hat{b}_Q) = \frac{1}{T} \Lambda^{-1}$ where T is the sample size.
- This expression tells you that the variance of the pricing kernel coefficients associated with the smallest eigenvalues is huge.
- The variance could even be more extreme when Λ is unknown.

Conditional Asset Pricing Revisited

- The conditional version of the pricing kernel is represented through time varying coefficients:

$$M_t = 1 - b'_{t-1}(r_t - \mu_{t-1})$$

where $\mu_{t-1} = E_{t-1}(r_t)$.

- The profession typically considers two formulations for time variation.
- First, b_{t-1} could vary with firm-level characteristics.
- Second, it could vary with macro-level variables.
- Notably, time-varying beta is different from time varying b , one does not imply the other.
- Considering firm level characteristics, it follows that $b_{t-1} = C_{t-1}b$, where C_{t-1} is an $N \times H$ matrix, H characteristics (e.g., size, profitability, past returns) for each of the N stocks, and b is an H –vector.
- Plugging b_{t-1} into the conditional version of the pricing kernel results in

$$M_t = 1 - b'[C'_{t-1}r_t - E_{t-1}(C'_{t-1}r_t)]$$

Conditional Asset Pricing Revisited

- The set of assets consists of H managed portfolios with realized returns $C'_{t-1}r_t$
- The vector b is again weights of the mean-variance efficient portfolio based on the managed portfolios.
- Next, we model time-variation with M macro variables, such as the dividend yield, the term spread, and the default spread, denoted by z_{t-1} :

$$b_{t-1} = bz_{t-1}$$

where b_{t-1} is an $N \times 1$ vector, b is an $N \times M$ matrix, and z_{t-1} is an $M \times 1$ vector.

- The pricing kernel representation is then

$$M_t = 1 - \text{vec}(b)'(r_t \otimes z_{t-1} - E_{t-1}(r_t \otimes z_{t-1}))$$

where $\text{vec}(b)$ is the vectorization of the matrix b .

- The pricing kernel parameters are again weights of the mean-variance efficient portfolio where the investment universe consists of $N \times M$ managed portfolios with realized returns $r_t \otimes z_{t-1}$.
- In what comes next, we revisit the presidential address of Cochrane and cover the literature that emerged in response to his so-called high-dimensionality challenge.

The high-dimensionality challenge per Cochrane (2011)

- Recall, Harvey, Liu, and Zhu (2016) document 316 factors discovered by academia.
- If you believe the results of Avramov, Chordia, Jostova, and Philipov noted earlier – then the dimension is too high: asset pricing anomalies concentrate in episodes of firm financial distress.
- Cochrane underlies the following challenges in understanding the cross section dispersion in average returns.
- First, which firm characteristics really provide independent information about average returns? Which are subsumed by others?
- Second, does each new anomaly variable also correspond to a new factor formed on those same anomalies? Momentum returns correspond to regression coefficients on a winner-loser momentum “factor.” Carry-trade profits correspond to a carry-trade factor. Do accruals return strategies correspond to an accruals factor?
- Third, how many of these new factors are really important? Can we again account for N independent dimensions of expected returns with $K < N$ factor exposures? Can we account for accruals return strategies by betas on some other factor, as with sales growth?
- Notice that factor structure is neither necessary nor sufficient for factor pricing. ICAPM and consumption-CAPM models do not predict or require that the multiple pricing factors will correspond to big co-movements in asset returns. And big co-movements, such as industry portfolios, need not correspond to any risk premium.
- There always is an equivalent single-factor pricing representation of any multifactor model, the mean-variance efficient portfolio. Still, the world would be much simpler if betas on only a few factors, important in the covariance matrix of returns, accounted for a larger number of mean characteristics.

The high-dimensionality challenge per Cochrane (2011)

- Fourth, eventually, we have to connect all this back to the central question of finance, why do prices move?
- Cochrane states: “to address these questions in the zoo of new variables, I suspect we will have to use different methods.”
- Indeed, financial economists have typically employed two methods to identify return predictors: (i) portfolio sorts using one or multiple characteristics and (ii) Fama and MacBeth (1973) cross section regressions.
- Portfolio sorts are subject to the curse of dimensionality when the number of characteristics is large, and linear regressions make strong functional-form assumptions and are sensitive to outliers.

Machine learning Methods in asset pricing

- In response to the Cochrane challenge and the call for different methods, there has been emerging literature that applies machine learning techniques in asset pricing.
- Machine learning typically prescribes a vast collection of high-dimensional models attempting to predict quantities of interest while imposing regularization methods.
- The paragraphs below provide general descriptions of machine learning methods and discuss how such methods have successfully been implemented in asset pricing.
- For perspective, it is well known that the ordinary least squares (OLS) estimator is the best linear unbiased estimator (BLUE) of the regression coefficients.
- “Best” means the lowest variance estimator among all other unbiased linear estimators.
- Notice that the regression errors do not have to be normal, nor do they have to be independent and identically distributed – but they have to be uncorrelated with mean zero and homoskedastic with finite variance.
- In the presence of either heteroskedasticity or autocorrelation, OLS is no longer BLUE.

Shortcomings of OLS

- With heteroskedasticity, we can still use the OLS estimators by finding heteroskedasticity-robust estimators of the variance, or we can devise an efficient estimator by re-weighting the data appropriately to incorporate heteroskedasticity.
- Similarly, with autocorrelation, we can find an autocorrelation-robust estimator of the variance. Alternatively, we can devise an efficient estimator by re-weighting the data appropriately to account for autocorrelation.
- Notice also that the requirement for an unbiased estimator is crucial since biased estimators exist with lower variance.
- This is where shrinkage methods come to play: the variance of the OLS estimator can be too high as the OLS coefficients are unregulated.
- If judged by Mean Squared Error (MSE), alternative biased estimators could be more attractive if they produce substantially smaller variance.
- In particular, let β denotes the true coefficient and let $\hat{\beta} = (X'X)^{-1}X'Y$, where X is a $T \times M$ matrix of explanatory variables and Y is a $T \times 1$ vector of the dependent variable.

Shortcomings of OLS

➤ Then

$$\begin{aligned}\text{MSE}(\hat{\beta}) &= E \left[(\hat{\beta} - \beta)' (\hat{\beta} - \beta) \right] \\ &= E \left\{ \text{tr} \left[(\hat{\beta} - \beta)' (\hat{\beta} - \beta) \right] \right\} \\ &= E \left\{ \text{tr} \left[(\hat{\beta} - \beta) (\hat{\beta} - \beta)' \right] \right\} \\ &= \text{tr} \left\{ E \left[(\hat{\beta} - \beta) (\hat{\beta} - \beta)' \right] \right\} \\ &= \sigma^2 \text{tr}[(X'X)^{-1}]\end{aligned}$$

- When predictors are highly correlated $\text{tr}[(X'X)^{-1}]$ can explode.
- Moreover, in the presence of many predictors, OLS delivers nonzero estimates to all coefficients – thus it is difficult to implement variable selection when the true model has a sparse representation.
- Also, the OLS solution is not unique if the design of X is not full rank.
- In sum, OLS often provide poor predictions, may be subject to over-fitting, does not penalize for model complexity, and could be difficult to interpret.
- From a Bayesian perspective, one can think of introducing priors on regression coefficients to make the coefficients smaller in magnitude or even zero them out altogether.
- Indeed, both Ridge and Lasso regressions could be casted in a Bayesian setup with informative priors.
- From a classical (non Bayesian) perspective, shrinkage methods are about penalizing complexity.

Economic restrictions on OLS

- The next pages discuss shrinkage methods.
- Meanwhile, here are two ways to potentially improve OLS estimates.
- Base case: the pooled OLS estimator corresponds to a panel (balanced) regression of future returns on firm attributes, where T and N represent again the time-series and cross-section dimensions.
- In particular, the objective is formulated as

$$\mathcal{L}(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left(r_{i,t+1} - f(c_{i,t}; \theta) \right)^2$$

where $f(c_{i,t}; \theta) = c'_{i,t} \theta$, $r_{i,t+1}$ is stock return at time $t+1$ per firm i , and $c_{i,t}$ is an M -vector of firm i attributes realized at time t .

- Predictive performance could be improved using an alternative optimization where stocks are weighted differently based on market size, volatility, credit quality, etc.:

$$\mathcal{L}(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T w_{i,t} \left(r_{i,t+1} - f(c_{i,t}; \theta) \right)^2$$

- An alternative economic based optimization takes account of the heavy tail displayed by stocks and the potential harmful effects of outliers. Then the objective is formulated such that squared (absolute) loss is applied to small (large) errors:

$$\mathcal{L}(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T H(r_{i,t+1} - f(c_{i,t}; \theta), \xi)$$

where ξ is a tuning parameter and

$$H(x, \xi) = \begin{cases} x^2, & \text{if } |x| \leq \xi \\ 2\xi|x| - \xi^2, & \text{if } |x| > \xi \end{cases}$$

Ridge Regression

- There are various shrinkage methods; We start with Ridge.
- Hoerl and Kennard (1970a, 1970b) introduce the Ridge regression

$$\min (Y - X\beta)'(Y - X\beta) \text{ s. t. } \sum_{j=1}^M \beta_j^2 \leq c$$

- The minimization can be rewritten as

$$\mathcal{L}(\beta) = (Y - X\beta)'(Y - X\beta) + \lambda(\beta'\beta)$$

- We get

$$\hat{\beta}^{\text{ridge}} = (X'X + \lambda I_M)^{-1} X'Y$$

- Notice that including λ makes the problem non-singular even when $X'X$ is non invertible.
- λ is the shrinkage parameter that controls the amount of regularization.
- It can be chosen using different methods including model selection criteria.
- As $\lambda \rightarrow 0$, we obtain the OLS estimator.
- As $\lambda \rightarrow \infty$, we have $\hat{\beta}^{\text{ridge}} = 0$, or intercept-only model.
- Of course, the ridge estimator is biased:

$$E(\hat{\beta}^{\text{ridge}}) = [I_M + \lambda(X'X)^{-1}]\beta \neq \beta$$

Interpretations of the Ridge Regression

Interpretation #1: Data Augmentation

- The ridge-minimization problem can be formulated as

$$\sum_{t=1}^T (y_t - x_t' \beta)^2 + \sum_{j=1}^M (0 - \sqrt{\lambda} \beta_j)^2$$

- Thus, the ridge-estimator is the usual OLS estimator where the data is transformed such that

$$X_\lambda = \begin{pmatrix} X \\ \sqrt{\lambda} I_M \end{pmatrix}, Y_\lambda = \begin{pmatrix} Y \\ 0_M \end{pmatrix}$$

- Then,

$$\hat{\beta}^{\text{ridge}} = (X_\lambda' X_\lambda)^{-1} X_\lambda' Y_\lambda$$

Interpretation #2: Eigen-values and Eigen-vectors

- By the singular value decomposition, we can express X as

$$X_{T \times M} = U_{T \times M} \Lambda_{M \times M}^{0.5} V_{M \times M}'$$

where $U = [U_1, \dots, U_M]$ is any $T \times M$ orthogonal matrix, $\Lambda^{0.5} = \begin{bmatrix} \lambda_1^{0.5} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_M^{0.5} \end{bmatrix}$ is an $M \times M$ matrix s. t.

Interpretations of the Ridge Regression

- $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$, and $V = [V_1, V_2, \dots, V_M]$ is an $M \times M$ orthogonal matrix.
- As $X'X = V\Lambda V'$, the matrix V denotes the eigenvectors of $X'X$ and $(\lambda_1, \dots, \lambda_M)$ are the corresponding eigenvalues
- How is the predicted value of Y related to eigen vectors?
- To answer, we would like first to find the eigenvectors and eigenvalues of the matrix Z

$$Z = X'X + \lambda I_M$$

- We know that for every $j=1,2,\dots,M$, the following holds by definition

$$(X'X)V_j = \lambda_j V_j$$

- Thus

$$\begin{aligned}(X'X + \lambda I_M)V_j &= (X'X)V_j + \lambda V_j \\ &= \lambda_j V_j + \lambda V_j = (\lambda_j + \lambda)V_j\end{aligned}$$

- Telling you that V still denotes the eigenvectors of Z and $\lambda_j + \lambda$ is the j -th eigenvalue.
- Notice now that if $A = V\Lambda V'$ then $A^N = V\Lambda^N V'$.
- Same eigenvectors while eigenvalues are raised to the power of N .

Interpretations of the Ridge Regression

- Then, the inverse of the matrix Z is given by

$$Z^{-1} = V \left[\text{diag} \left(\frac{1}{\lambda_1 + \lambda}, \frac{1}{\lambda_2 + \lambda}, \dots, \frac{1}{\lambda_M + \lambda} \right) \right] V'$$

- And

$$\begin{aligned} \hat{\beta}^{\text{ridge}} &= Z^{-1} X' Y = (X' X + \lambda I_M)^{-1} X' Y \\ &= V \left[\text{diag} \left(\frac{\lambda_1^{0.5}}{\lambda_1 + \lambda}, \frac{\lambda_2^{0.5}}{\lambda_2 + \lambda}, \dots, \frac{\lambda_M^{0.5}}{\lambda_M + \lambda} \right) \right] U' Y \end{aligned}$$

- And the fitted value is

$$\hat{Y}^{\text{ridge}} = X \hat{\beta}^{\text{ridge}} = \left[\sum_{j=1}^M \left(U_j \frac{\lambda_j}{\lambda_j + \lambda} U_j' \right) \right] Y$$

- Hence, ridge regression projects Y onto components with large λ_j .
- Or, ridge regression shrinks the coefficients of low variance components.

Informative Bayes Prior

Interpretation #3: Informative Bayes Prior

- Suppose that the prior on β is of the form:

$$\beta \sim N\left(0, \frac{1}{2M} I_M\right)$$

- Then, the posterior mean of β is:

$$(X'X + \lambda I_M)^{-1} X'Y$$

- We will discuss in detail Bayesian approaches later in the notes.
- Meanwhile, consider the family of priors for mean return μ :

$$\mu \sim N\left(0, \frac{\sigma^2}{s^2} V^\eta\right)$$

where $s^2 = \text{trace}(V)$ which is also the sum of the eigenvalues of V , or $s^2 = \sum_{j=1}^N \lambda_j$.

- To see why, notice that $\text{trace}(V) = \text{trace}(Q\Lambda Q') = \text{trace}(\Lambda Q'Q) = \text{trace}(\Lambda) = \sum_{j=1}^N \lambda_j$
- σ^2 is a constant controlling for the degree of confidence in the prior.
- The case $\eta = 1$ gives the asset pricing prior of Pastor (2000) and Pastor and Stambaugh (2000).
- The case $\eta = 2$ gives the prior motivated by Kozak, Nagel, and Santosh (2017b).

Informative Bayes Prior

- Let us again make the projection of the original space of returns into the space of principal components:

$$\begin{aligned}M_t &= 1 - \mu' V^{-1} (r_t - \mu) \\&= 1 - \mu' Q \Lambda^{-1} Q' (r_t - \mu) \\&= 1 - \mu'_Q \Lambda^{-1} (Q_t - \mu_Q) \\&= 1 - b'_Q (Q_t - \mu_Q)\end{aligned}$$

- As $\mu \sim N\left(0, \frac{\sigma^2}{s^2} V^\eta\right)$ it follows that $\mu_Q = Q' \mu$ has the prior distribution

$$\begin{aligned}\mu_Q = Q' \mu &\sim N\left(0, \frac{\sigma^2}{s^2} Q' V^\eta Q\right) \\&\sim N\left(0, \frac{\sigma^2}{s^2} Q' Q \Lambda^\eta Q' Q\right) \\&\sim N\left(0, \frac{\sigma^2}{s^2} \Lambda^\eta\right)\end{aligned}$$

Informative Bayes Prior

- As $b_Q = \Lambda^{-1} \mu_Q$ its prior distribution is formulated as (Λ is assumed known):

$$b_Q = \Lambda^{-1} \mu_Q \sim N\left(0, \frac{\sigma^2}{s^2} \Lambda^{\eta-2}\right)$$

- Notice that for $\eta < 2$ the variance of the b_Q coefficients associated with the smallest eigenvalues explodes.
- For $\eta = 2$, the pricing kernel coefficients $b = V^{-1} \mu$ have the prior distribution

$$b \sim N\left(0, \frac{\sigma^2}{s^2} I_N\right)$$

- Picking $\eta = 2$ makes the prior of b independent of V .
- Further, the likelihood of b is given by

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

where $\hat{\mu}$ is the sample mean return.

- Then, the posterior mean of b is given by $E(b) = (V + \lambda I_N)^{-1} \hat{\mu}$ where $\lambda = \frac{s^2}{T\sigma^2}$.
- The posterior variance is $var(b) = \frac{1}{T} (V + \lambda I_N)^{-1}$.
- Similar to the ridge regression with a tuning parameter λ .

Informative Bayes Prior

- ▶ The prior expected value of the squared SR is given by (V is assumed known):

$$\begin{aligned} E(SR^2) &= E(\mu' V^{-1} \mu) \\ &= E(\mu' Q \Lambda^{-1} Q' \mu) \\ &= E(\mu_Q' \Lambda^{-1} \mu_Q) \\ &= E\{\text{trace}(\mu_Q' \Lambda^{-1} \mu_Q)\} \\ &= \text{trace}[\Lambda^{-1} E(\mu_Q \mu_Q')] \\ &= \frac{\sigma^2}{s^2} \text{trace}[\Lambda^{\eta-1}] \end{aligned}$$

- ▶ The Pastor-Stambaugh prior ($\eta = 1$) tells you that

$$E(SR^2) = \frac{\sigma^2}{s^2} \text{trace}(I_N) = N \frac{\sigma^2}{s^2}$$

- ▶ That is, each principal component portfolio has the same expected contribution to the Sharpe ratio.
- ▶ If $\eta = 2$, then

$$E(SR^2) = \sum_{j=1}^N \frac{\sigma^2}{s^2} \lambda_j = \sigma^2$$

- ▶ Then, the expected contribution of each PC is proportional to its eigenvalue.
- ▶ This reinforces the notion that only the first few principal components could explain the cross section variation in expected returns.

Lasso

- We now consider various Lasso (least absolute shrinkage and selection operator) models.
- Tibshirani (1996) was the first to introduce Lasso.
- Lasso simultaneously performs variable selection and coefficient estimation via shrinkage.
- While the ridge regression implements an l_2 -penalty, Lasso is an l_1 -optimization:

$$\min (Y - X\beta)'(Y - X\beta) \text{ s. t. } \sum_{j=1}^M |\beta_j| \leq c$$

- The l_1 penalization approach is called basis pursuit in signal processing.
- We have again a non-negative tuning parameter λ that controls the amount of regularization:

$$\mathcal{L}(\beta) = (Y - X\beta)'(Y - X\beta) + \lambda \sum_{j=1}^M |\beta_j|$$

- Both Ridge and Lasso have solutions even when $X'X$ may not be of full rank (e.g., when there are more explanatory variables than time-series observations) or ill conditioned.

Lasso

- Unlike Ridge, the Lasso coefficients cannot be expressed in closed form.
- However, Lasso can provide with a set of sparse solutions.
- That improves the interpretability of regression models.
- Large enough λ , or small enough c , will set some coefficients exactly to zero.
- To understand why, notice that LASSO can be casted as having a Laplace prior β

$$P(\beta|\lambda) \propto \left(\frac{\lambda}{2\sigma}\right) \exp\left(-\frac{\lambda|\beta|}{\sigma}\right)$$

- In particular, Lasso obtains by combining Laplace prior and normal likelihood.
- Like the normal distribution, Laplace is symmetric.
- Unlike the normal distribution, Laplace has a spike at zero (first derivative is discontinuous) and it concentrates its probability mass closer to zero than does the normal distribution.
- This could explain why Lasso (Laplace prior) sets some coefficients to zero, while Ridge (normal prior) does not.

Adaptive Lasso and Bridge regression

- LASSO forces coefficients to be equally penalized.
- One modification is to assign different weights to different coefficients:

$$\mathcal{L}(\beta) = (Y - X\beta)'(Y - X\beta) + \lambda \sum_{j=1}^M w_j |\beta_j|$$

- It can be shown that if the weights are data driven and are chosen in the right way such weighted LASSO can have the so-called oracle properties even when the LASSO does not. This is the adaptive LASSO.
- For instance, w_j can be chosen such that it is equal to one divided by the absolute value of the corresponding OLS coefficient raised to the power of $\gamma > 0$. That is, $w_j = \frac{1}{|\beta_j|^\gamma}$ for $j = 1, \dots, M$.

- The adaptive LASSO estimates are given by

$$\mathcal{L}(\beta) = (Y - X\beta)'(Y - X\beta) + \lambda \sum_{j=1}^M w_j |\beta_j|$$

- The adaptive LASSO is a convex optimization problem and thus does not suffer from multiple local minima
- Frank and Friedman (1993) introduce bridge regression, which generalizes for ℓ_q penalty

$$\mathcal{L}(\beta) = (Y - X\beta)'(Y - X\beta) + \lambda \sum_{j=1}^M |\beta_j|^q$$

- Notice that $q = 0, 1, 2$, correspond to OLS, LASSO, and ridge, respectively.
- Moreover, the optimization is convex for $q \geq 1$ and the solution is sparse for $0 \leq q \leq 1$.

The Elastic Net

- The elastic net is yet another regularization and variable selection method.
- Zou and Hastie (2005) describe it as stretchable fishing net that retains “all big fish”.
- Using simulation, they show that it often outperforms Lasso in terms of prediction accuracy.
- The elastic net encourages a grouping effect, where strongly correlated predictors tend to be in or out of the model together.
- The elastic net is particularly useful when the number of predictors is much bigger than the number of observations.
- The naïve version of the elastic net is formulated through

$$\mathcal{L}(\beta) = (Y - X\beta)'(Y - X\beta) + \lambda_1 \sum_{j=1}^M |\beta_j| + \lambda_2 \sum_{j=1}^M \beta_j^2$$

- Thus, the elastic net combines l_1 and l_2 norm penalties.
- It still produces sparse representations.

Group Lasso

- Again, the data consist of Y , a T -vector of the dependent variable, and X , a $T \times M$ matrix of explanatory variables.
- Suppose that the M predictors are divided into L groups, with M_l denoting the number of predictors in group l .
- X_l represents the predictors corresponding to the l -th group, while β_l is the corresponding coefficient vector.
- Notice, $\beta = [\beta_1', \beta_2', \dots, \beta_L']'$.
- Assume that both Y and X have been centered.
- The group Lasso solves the convex optimization problem.

$$\mathcal{L}(\beta) = \left(Y - \sum_{l=1}^L X_l \beta_l \right)' \left(Y - \sum_{l=1}^L X_l \beta_l \right) + \lambda \left(\sum_{l=1}^L \beta_l' \beta_l \right)^{\frac{1}{2}}$$

- The group Lasso does not yield sparsity within a group.
- If a group of parameters is non-zero, they will all be non-zero.
- The sparse group Lasso criterion does yield sparsity

$$\mathcal{L}(\beta) = \left(Y - \sum_{l=1}^L X_l \beta_l \right)' \left(Y - \sum_{l=1}^L X_l \beta_l \right) + \lambda_1 \left(\sum_{l=1}^L \beta_l' \beta_l \right)^{\frac{1}{2}} + \lambda_2 \sum_{j=1}^M |\beta_j|$$

Non parametric shrinkage

- ▶ Lasso, Adaptive Lasso, Group Lasso, Ridge, Bridge, and Elastic net are all linear or parametric approaches for shrinkage.
- ▶ Group Lasso implements the same penalty to predictors belonging to some pre-specified group while different penalty applies to different groups.
- ▶ Some other parametric approaches (uncovered here) include the smoothed clip absolute deviation (SCAD) penalty of Fang and Li (2001) and Fang and Peng (2004) and the minimum concave penalty of Zhang (2010).
- ▶ In many application, however, there is little a priori justification for assuming that the effects of covariates take a linear form or belong to other known parametric family.
- ▶ Huang, Horowitz, and Wei (2010) thus propose to use a nonparametric approach: the adaptive group Lasso for variable selection.
- ▶ This approach is based on a spline approximation to the nonparametric components.
- ▶ To achieve model selection consistency, they apply the group Lasso in two steps.
- ▶ First, they use the group Lasso to obtain an initial estimator and reduce the dimension of the problem.
- ▶ Second, they use the adaptive group Lasso to select the final set of nonparametric components.

Nonparametric Models

- In finance, Cochrane (2011) notes that portfolio sorts are the same thing as nonparametric cross section regressions.
- Drawing on Huang, Horowitz, and Wei (2010), Freyberger, Neuhier, and Weber (2017) study this equivalence formally.
- The cross section of stock returns is modelled as a non-linear function of firm characteristics:

$$r_{it} = m_t (C_{1,it-1}, \dots, C_{S,it-1}) + \epsilon_{it}$$

- Notation:

- r_{it} is the return on firm i at time t .
 - m_t is a function of S firm characteristics C_1, C_2, \dots, C_S .
 - Notice, m_t itself is not stock specific but firm characteristics are.
 - They assume an additive model of the following form
- $$m_t(C_1, \dots, C_S) = \sum_{s=1}^S m_{t,s}(C_s)$$
- As the additive model implies that $\frac{\partial^2 m_t(c_1, \dots, c_S)}{\partial c_s \partial c_{s'}} = 0$ for $s \neq s'$, apparently there should not be no cross dependencies between characteristics.
 - Such dependencies can still be accomplished via interacting characteristics.

Nonparametric Models

- For each characteristic s , let $F_{s,t}(\cdot)$ be a strictly monotone function and let $F_{s,t}^{-1}(\cdot)$ denote its inverse.
- Define $\tilde{C}_{s,it-1} = F_{s,t}(C_{s,it-1})$ such that $\tilde{C}_{s,it-1} \in [0,1]$.
- That is, characteristics are monotonically mapped into the $[0,1]$ interval.
- An example for $F_{s,t}(\cdot)$ is the rank function: $F_{s,t}(C_{s,it-1}) = \frac{\text{rank}(C_{s,it-1})}{N_t+1}$, where N_t is the total number of firms at time t .
- The aim then is to find \tilde{m}_t such that

$$m_t(C_1, \dots, C_S) = \tilde{m}_t(\tilde{C}_{1,it-1}, \dots, \tilde{C}_{S,it-1})$$

- In particular, to estimate the \tilde{m}_t function, the normalized characteristic interval $[0,1]$ is divided into L subintervals ($L+1$ knots): $0 = x_0 < x_1 < \dots < x_{L-1} < x_L = 1$.
- To illustrate, consider the equal spacing case.
- Then, $x_l = \frac{l}{L}$ for $l=0, \dots, L-1$ and the intervals are:

$$\tilde{I}_1 = [x_0, x_1), \tilde{I}_l = [x_{l-1}, x_l) \text{ for } l=2, \dots, L-1, \text{ and } \tilde{I}_L = [x_{L-1}, x_L]$$

Nonparametric Models

- Each firm characteristic is transformed into its corresponding interval.
- Estimating the unknown function $\tilde{m}_{t,s}$ nonparametrically is done by using quadratic splines.
 - The function $\tilde{m}_{t,s}$ is approximated by a quadratic function on each interval \tilde{I}_l .
 - Quadratic functions in each interval are chosen such that $\tilde{m}_{t,s}$ is continuous and differentiable in the whole interval $[0,1]$.
 - $\tilde{m}_{t,s}(\tilde{c}) = \sum_{k=1}^{L+2} \beta_{tsk} \times p_k(\tilde{c})$, where $p_k(\tilde{c})$ are basis functions and β_{tsk} are estimated slopes.
 - In particular, $p_1(y) = 1, p_2(y) = y, p_3(y) = y^2$, and $p_k(y) = \max\{y - x_{k-3}, 0\}^2$ for $k = 4, \dots, L + 2$
 - In that way, you can get a continuous and differentiable function.
- To illustrate, consider the case of two characteristics, e.g., size and book to market (BM), and 3 intervals.
- Then, the \tilde{m}_t function is:

$$\begin{aligned} \tilde{m}_t(\tilde{c}_{i,size}, \tilde{c}_{i,BM}) = & \\ = & \beta_{t,size,1} \times 1 + \beta_{t,size,2} \times \tilde{c}_{i,size} + \beta_{t,size,3} \times \tilde{c}_{i,size}^2 + \beta_{t,size,4} \times \max\{\tilde{c}_{i,size} - 1/3\}^2 + \beta_{t,size,5} \times \max\{\tilde{c}_{i,size} - 2/3\}^2 \\ & + \beta_{t,BM,1} \times 1 + \beta_{t,BM,2} \times \tilde{c}_{i,BM} + \beta_{t,BM,3} \times \tilde{c}_{i,BM}^2 + \beta_{t,BM,4} \times \max\{\tilde{c}_{i,BM} - 1/3\}^2 + \beta_{t,BM,5} \times \max\{\tilde{c}_{i,BM} - 2/3\}^2 \end{aligned}$$

Adaptive group Lasso

- The estimation of \tilde{m}_t is done in two steps:
- First step, estimate the slope coefficients b_{sk} using the group Lasso routine:

$$\hat{\beta}_t = \underset{b_{sk:s=1,\dots,S;k=1,\dots,L+2}}{\operatorname{argmin}} \sum_{i=1}^{N_t} \left(r_{it} - \sum_{s=1}^S \sum_{k=1}^{L+2} b_{sk} \times p_k(\tilde{C}_{s,it-1}) \right)^2 + \lambda_1 \sum_{s=1}^S \left(\sum_{k=1}^{L+2} b_{sk}^2 \right)^{1/2}$$

- All together, the number of b_{sk} coefficients is $S \times (L + 2)$.
- The second term is penalty applied to the spline expansion.
- λ_1 is chosen such that it minimizes the Bayesian Information Criterion (BIC).
- The essence of group Lasso is either to include or exclude all $L+2$ spline terms associated with a given characteristic.
- While this optimization yields a sparse solution there are still many characteristics retained.
- To include only characteristics with a strong predictive the adaptive Lasso is then employed.

Adaptive group Lasso

- To implement adaptive group Lasso, define the following weights using estimates for b_{sk} from the first step:

$$w_{ts} = \begin{cases} \left(\sum_{k=1}^{L+2} \tilde{b}_{sk}^2 \right)^{-\frac{1}{2}} & \text{if } \sum_{k=1}^{L+2} \tilde{b}_{sk}^2 \neq 0 \\ \infty & \text{if } \sum_{k=1}^{L+2} \tilde{b}_{sk}^2 = 0 \end{cases}$$

- Then estimate again the coefficients b_{sk} using the above-estimated weights w_{ts}

$$\hat{\beta}_t = \underset{b_{sk:s=1,\dots,S;k=1,\dots,L+2}}{\operatorname{argmin}} \sum_{i=1}^{N_t} \left(r_{it} - \sum_{s=1}^S \sum_{k=1}^{L+2} b_{sk} \times p_k(\tilde{C}_{s,it-1}) \right)^2 + \lambda_2 \sum_{s=1}^S \left(w_{ts} \sum_{k=1}^{L+2} b_{sk}^2 \right)^{1/2}$$

- λ_2 is chosen such that it minimizes BIC.
- The weights w_{ts} guarantee that we do not select any characteristic in the second step that was not selected in the first step.

Machine learning methods – summary

- Machine learning methods have been implemented in response to the call to go beyond cross section regressions and portfolio sorts to deal with the ever growing dimensionality of the cross section.
- Indeed, machine methods have good potential for improving risk premium measurements, while regular cross sectional regressions and portfolio sorts are ill suited for that aim.
- However, as implied by Gu, Kelly, and Xiu (2018), improved measurements are reduced form statistical correlations, while machine learning methods do not really tell us anything about economic mechanism or equilibria.
- Or, there is no sense in which we can rely on machine learning methods to identify deep fundamental associations among asset prices, firm characteristics, conditioning macro or firm level variables.

Sparse versus non sparse solutions

- This research controversy crosses disciplines that deal with model uncertainty and variable selection.
- Zou and Hastie (2005) advocate in favor of sparse representation.
- Fu (1998) advocates using general cross-validation to select the shrinkage parameter (q) and the tuning parameter (λ) in a bridge regression setup and shows that bridge regression performs well. The bridge regression could produce non sparse solutions.
- In asset pricing, Freyberger, Neuhierl, and Weber (2017) use the adaptive group LASSO to select characteristics and to estimate how they affect expected returns non-parametrically. Feng, Giglio, and Xiu (2017) implements a methodology that marries the double selection LASSO method of Belloni, Chernozhukov, and Hansen (2014b).
- Both these studies adopt Lasso-style estimation with l_1 -penalty and they suggest a relatively high degree of redundancy among equity predictors.
- In contrast, Kozak, Nagle, and Santosh (2017b) advocate against sparse solutions.
- They propose a Bayesian approach that shrinks the SDF coefficients towards zero.
- Their formulation is similar to a ridge regression with an important difference.
- As informative priors are about the pricing kernel parameters the degree of shrinkage is not equal for all assets. Rather there is more shrinkage to SDF coefficients associated with low-eigenvalue PCs.
- Notice that while the q -model or the Fama-French five factor model can imply sparse solution a caveat is in order. Take the present-value relation - it can indeed motivate why book-to-market and expected profitability could jointly explain expected returns. However, the notion that expected profitability is unobserved gives license to fish a large number of observable firm characteristics that predict future returns through their ability to predict future profitability.

Consumption Based Asset Pricing Models

Perhaps the consumption CAPM does a better job?

- Most prominent in the class of inter-temporal models is the consumption CAPM (CCAPM).
- The CCAPM is a single state variable model; real consumption growth is the single factor.
- Consumption-based asset pricing models have been among the leading multi-period general equilibrium models in financial economics for the past four decades.
- The Consumption Capital Asset Pricing Model (CCAPM) was first derived in the late 1970s in successively more general models by Rubinstein (1976), Breeden and Litzenberger (1978), and Breeden (1979).
- Lucas (1978) did not derive the CCAPM formula, yet his work on Euler equations was also helpful to many empiricists in subsequent consumption-based asset pricing tests.
- The CCAPM states that the expected excess return on any risky asset should be proportional to its “consumption beta”, or covariation with consumption growth.
- Financial assets with higher sensitivities of returns to movements in real consumption spending have more systematic risk and should have proportionately higher excess returns

The consumption CAPM

- Such assets pay off more when consumption is high and marginal utility is low, and pay less when consumption is low and marginal utility is high, so they are worth less in price and have higher equilibrium returns.
- This CCAPM differs from the CAPM as real consumption growth is not perfectly correlated with market returns.
- In a multi-period model, market wealth can be high and still have high marginal utility if the investment opportunity set is good, as shown by Merton (1973) and Breeden (1984).
- The first two decades of CCAPM tests produced mixed results tilting towards the model rejection
- Tests of the special case of the CCAPM under constant relative risk aversion by Hansen and Singleton (1983), Mehra and Prescott (1985), and others rejected the model.
- Chen, Roll, and Ross (1986) found no significant consumption factor priced in the presence of other factors, including industrial production, junk bond returns, and inflation hedges
- Grossman, Melino, and Shiller (1987), Breeden, Gibbons, and Litzenberger (BGL, 1989) and Wheatley (1988) examined measurement issues in consumption (such as time aggregation) and their biases on measures of volatility and consumption betas.

The consumption CAPM

- BGL found a significant positive coefficient on consumption betas; and separately a significant positive coefficient on market betas; however, both the CCAPM and the CAPM were rejected.
- BGL derived a useful result: estimation of consumption betas relative to returns on a consumption mimicking portfolio, which allows greater number and frequency of observations and more precise estimates of consumption betas.
- The very strong theory in support of the CCAPM, contrasted with weak early empirical support, motivated researchers to improve both their theoretical and empirical modeling.
- On the theoretical side, Pye (1972, 1973) and Greenig (1986) developed time-multiplicative utility functions.
- Then Sundaresan (1989), Constantinides (1990), and Abel (1990) modeled habit formation.
- Epstein-Zin (1989) and Weil (1989) (often jointly referred to as EZ-W) developed preference structures that displayed time-complementarity in utility for consumption streams, allowing researchers to separate effects of different levels of intra-temporal relative risk aversion (RRA) from levels of the elasticity of intertemporal substitution (EIS).

The consumption CAPM

- Campbell and Cochrane (1999) later produced a model with the habit formation approach.
- With a subsistence level of consumption for a “representative individual,” their model allows for dramatic rises in relative risk aversion as surplus consumption (above habit) goes towards zero in severe recessions.
- With the flexibility afforded by this model, they were able to fit many aspects of empirical data on stock and bond returns as related to real consumption growth, especially the risk premium on the stock market.
- Mehra and Prescott (1985) find that the equity premium is too high (the “equity premium puzzle”), given the low volatility of real consumption growth.
- Mankiw and Zeldes (1991) considered that many households did not own stock at all or in significant amounts, a situation called “limited participation.”
- They pointed out that there is no reason that the Euler equation should hold for households who are not investing.
- They found that for households who actually owned stocks, the implied estimates of relative risk aversion were much more reasonable than for households who did not own stocks.
- Heaton and Lucas (1992, 1996) examined “incomplete markets” that did not permit full hedging of labor income, thus causing consumers to have more volatile consumption streams.

The consumption CAPM

- Vissing-Jorgensen (2002) focused on estimating the “elasticity of intertemporal substitution,” which determines how much consumers change their expected consumption growth rate when interest rates or expected returns on assets change.
- She finds the EIS to be quite different for stockholders than for non-stockholders, generally getting plausible estimates for those who chose to invest in stocks and bonds and based on trading off current consumption versus future consumption.
- Also on the empirical side, advances were made in examining changes in **conditional** means, variances, and covariances and testing conditional versions of the CAPM and CCAPM, as in Harvey (1991) and Ferson and Harvey (1991, 1999).
- Lettau and Ludvigson (2001a,b) use deviations of consumption from total wealth (which includes a human capital estimate in addition to stock market wealth) as a conditioning or “scaling” variable for changing mean returns.
- They find results quite compatible with Merton’s (1973) and Breeden’s (1984) intertemporal theories, in that high consumption versus wealth is a predictor of future investment returns, as consumers optimally smooth forward those changes in expected returns

The consumption CAPM

- Lettau and Ludvigson also find significant differences in the movements of consumption betas of value vs. growth stocks during recessions
- They find that value stocks tend to have much larger increase in consumption betas during recessions, when risks and risk premiums are high, which helps to explain the findings of higher returns on value stocks than predicted by the unconditional CCAPM betas.
- More recently, Bansal and Yaron (2004) consider the “long run risk” caused by small, persistent shocks in the drift and volatility of real consumption growth.
- They show that variance of real consumption growth grows more than proportionally with time, which is consistent with the persistence of growth shocks.
- They also provide evidence that the conditional volatility of consumption is time-varying, which leads naturally to time-varying risk premia.
- Much subsequent research has been done on this long run risk model, most notably in the paper by Bansal, Dittmar, and Kiku (2009).
- They show that aggregate consumption and aggregate dividends have a co-integrating relation

The consumption CAPM

- They observe that “the deviation of the level of dividends from consumption (the so-called error correction variable) is important for predicting dividend growth rates and returns at all horizons” (1, 5 and 10 years).
- Imposing co-integration allows them to predict 11.5% of the variation in 1-year returns, whereas only 7.5% of the variation is predicted without co-integration
- Their conditional consumption betas account for about 75% of the cross-sectional variation in risk premia for the one-year horizon, and 85% for long horizons
- After Grossman, Melino, and Shiller (1987) and Breeden, Gibbons, and Litzenberger (1989) raised concerns about measuring consumption, Parker and Julliard (2005) showed that it is important to measure “ultimate consumption betas,” since consumption changes are slow-moving, and could take 2-3 years for the full effects to be observed.
- Using measures of these ultimate consumption betas, they were able to explain much of the Fama-French (1992) effects for size and book-to-market portfolios
- Let us derive a simple discrete time version of the CCAPM
- The derivation is based on the textbook treatment of Cochrane (2007).

A simple derivation of the consumption CAPM

- ▶ We use the power utility to describe investor preferences:

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

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

- ▶ 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 (w.l.o.g.) a two-period investor whose income at time s is e_s and let y be the amount of security she chooses to buy.

A simple derivation of the consumption CAPM

- The investor's problem is

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

$$c_t = e_t - p_t y,$$

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

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,

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

- The left hand side 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 simple derivation of the consumption CAPM

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

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

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 is the fundamental discount factor view of asset pricing theories.
- Observe from the representation that the gross risk-free rate, the rate known at time t and uncorrelated with the discount factor, is given by $R_{t,t+1}^f = 1/\mathbb{E}_t(\xi_{t+1})$.
- When investor preferences are described by the power utility function, as in equation ($u(c_t) = \frac{1}{1-\gamma} c_t^{1-\gamma}$ for $\gamma \neq 1$), 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 risk-free rate is

$$\begin{aligned} r_{t,t+1}^f &= -\ln(\rho) - \ln \mathbb{E}_t [\exp(-\gamma \Delta \ln c_{t+1})], \\ &= -\ln(\rho) + \gamma \mathbb{E}_t(\Delta \ln c_{t+1}) - \frac{\gamma^2}{2} \sigma_t^2(\Delta \ln c_{t+1}). \end{aligned}$$

A simple derivation of the consumption CAPM

- To derive an explicit form for the risk-free rate we have used the useful relation that if x is normally distributed then

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

- We can see from the risk-free rate equation that the EIS (elasticity of inter-temporal substitution) is $1/\gamma$ which creates some problems, as discussed below.
- From the fundamental representation, we also obtain a beta pricing model of the form

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

- 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

A simple derivation of the consumption CAPM

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

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

where

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

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

- This is the discrete time version of the consumption CAPM.
- The relation is exact in continuous time.
- The asset's risk is defined as the covariance between the asset return and consumption growth.
- The risk premium is the product of the risk aversion and the volatility of consumption growth
- Notice from the asset pricing equation that the asset expected return is larger as the covariance between the asset return and consumption growth gets larger.
- 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 commanding higher expected return.

Theoretically, 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 been unsuccessful

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), the risk-free rate puzzle, and the excess volatility puzzle.

The Equity Premium Puzzle

- To illustrate, let us manipulate the first order condition: $1 = \mathbb{E}_t(\xi_{t+1}R_{t+1})$ (for notational clarity I will suppress the time dependence)

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

- Dividing both sides of $(E(\xi)E(R) + \rho_{\xi,R}\sigma(\xi)\sigma(R))$ by $E(\xi)\sigma(R)$ leads to

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

which implies that

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

- The left hand side is known as the Sharpe ratio.

The Equity Premium Puzzle

- The highest Sharpe ratio is associated with portfolios lying on the efficient frontier.
- Notice that the slope of the frontier is governed by the volatility of the discount factor.
- Under the CCAPM it follows that

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

- When log pricing kernel is normally distributed, the right hand side can be shown to be equal to (proof in the appendix)

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

which can be approximated by

$$\gamma \sigma (\Delta \ln c).$$

- 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%.

The Equity Premium Puzzle

- 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.

The Risk-Free Rate Puzzle

- Using the standard CCAPM framework also gives rise to the risk-free rate puzzle.
- Recall, we have shown that

$$r_{t,t+1}^f = -\ln(\rho) + \gamma \mathbb{E}_t(\Delta \ln c_{t+1}) - \frac{\gamma^2}{2} \sigma_t^2(\Delta \ln c_{t+1})$$

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

How could the Equity Premium and Risk-Free Puzzles be resolved ?

- ▶ 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 risk-free rate. So higher risk aversion reinforces the risk-free puzzle.
- ▶ Perhaps the stock returns over the last 50 years are good luck rather than an equilibrium compensation for risk.
- ▶ 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.

How could the Equity Premium and Risk-Free Puzzles be resolved ?

- 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 risk-free rate, whereas risk aversion is about preferences over random variables.
- In Epstein and Zin (1989) and Weil (1990) recursive inter-temporal utility functions, the risk aversion is separated from the elasticity of inter-temporal substitution, thereby separating the equity premium from the risk-free 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 risk-free 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. i.e., the potential for rare economic disasters explains a lot of asset-pricing puzzles including high equity premium, low risk free rate, and excess volatility.

How could the Equity Premium and Risk-Free Puzzles be resolved ?

- 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

$$g \sim N(\mu_g, \sigma_g^2).$$

- 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, σ_g is a random variable obeying the inverted gamma distribution.

- Then g has the Student-t distribution. We will show this result later upon digging into Bayesian Econometrics.

- The student t distribution captures both the high equity premium and low risk-free rate.

- In what follows, I will elaborate on the three most successful consumption models: long run risk, habit formation, and prospect theory.

- Beforehand, it is useful to get familiarity with the E-Z preferences.

The Epstein-Zin preferences

- Epstein and Zin follow the work by Kreps and Porteus to introduce a class of preferences that breaks the link between risk aversion and EIS. The basic formulation is

$$U_t = \left\{ (1 - \beta) C_t^{1 - \frac{1}{\psi}} + \beta \left(E_t(U_{t+1}^{1-\gamma}) \right)^{\frac{1 - \frac{1}{\psi}}{1 - \gamma}} \right\}^{\frac{1}{1 - \frac{1}{\psi}}}$$

- This utility function can also be rewritten as

$$U_t = \left\{ (1 - \beta) C_t^{\frac{1-\gamma}{\theta}} + \beta \left(E_t(U_{t+1}^{1-\gamma}) \right)^{\frac{1}{\theta}} \right\}^{\frac{\theta}{1-\gamma}} \quad \text{where } \theta = \frac{1-\gamma}{1 - \frac{1}{\psi}}$$

- To give some intuition, consider the case where $\psi = 1$:

$$U_t = (1 - \beta) \log C_t + \beta \theta \log E_t \left[\exp \left(\frac{U_{t+1}}{\theta} \right) \right]$$

- With normally distributed U_t we get that the conditional variance of utility matters:

$$U_t = (1 - \beta) \log C_t + \beta E_t(U_{t+1}) + \frac{1}{2} \times \frac{\beta}{\theta} \text{Var}_t(U_{t+1})$$

The first two terms correspond to the time-additive case while the third is the E-Z addition.

The Epstein-Zin preferences

- The Euler equation of the Epstein-Zin preferences is given by,

$$1 = E_t \left[\left\{ \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\psi}} \right\}^{\theta} \left\{ \frac{1}{R_{m,t+1}} \right\}^{1-\theta} R_{i,t+1} \right]$$

The pricing kernel

- For the market portfolio itself, the Euler equation takes the form:

$$1 = E_t \left[\left\{ \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\psi}} R_{m,t+1} \right\}^{\theta} \right]$$

- As $\theta = 1$ this collapses to the familiar expression for power utility
- Now assume that $(C_t, R_{m,t})$ are jointly homoscedastic and log-normally distributed, then:

$$1 = E_t \exp \left[\theta \ln \beta - \frac{\theta}{\psi} \ln \left(\frac{C_{t+1}}{C_t} \right) + \theta \ln R_{m,t+1} \right]$$

The Epstein-Zin preferences

- Notice that the form in the exponential has a normal distribution with time-varying mean (E_t) and constant variance, due to the homoscedasticity assumption, (V), both are given by:

$$E_t = \theta \ln \beta - \frac{\theta}{\psi} E_t(c_{t+1}) + \theta E_t r_{m,t+1}$$
$$V = \left(\frac{\theta}{\psi}\right)^2 \sigma_c^2 + \theta^2 \sigma_m^2 - \frac{2\theta^2}{\psi} \sigma_{cm}$$

where $c_{t+1} = \ln \left(\frac{c_{t+1}}{c_t} \right)$

- Thus,

$$1 = E_t \exp \left[\theta \ln \beta - \frac{\theta}{\psi} \ln \left(\frac{c_{t+1}}{c_t} \right) + \theta \ln R_{m,t+1} \right] = \exp \left[E_t + \frac{1}{2} V \right]$$

- Taking logs from both sides it follows that

$$0 = \theta \ln \beta - \frac{\theta}{\psi} E_t(c_{t+1}) + \theta E_t r_{m,t+1} + \frac{1}{2} \left[\left(\frac{\theta}{\psi} \right)^2 \sigma_c^2 + \theta^2 \sigma_m^2 - \frac{2\theta^2}{\psi} \sigma_{cm} \right]$$

The Epstein-Zin preferences

- Then we get

$$E_t(c_{t+1}) = \mu_m + \psi E_t r_{m,t+1}$$

where

$$\mu_m = \psi \ln \beta + \frac{1}{2} \left(\frac{\theta}{\psi} \right) \text{Var}_t [c_{t+1} - \psi r_{m,t+1}]$$

- We can also understand the cross section of returns: the pricing kernel representation is

$$E_t \exp \left[\theta \ln \beta - \frac{\theta}{\psi} \ln \left(\frac{c_{t+1}}{c_t} \right) + (\theta - 1) \ln R_{m,t+1} + \ln R_{i,t+1} \right] = 1$$

- Then expected return is given by

$$E_t(r_{i,t+1}) - r_{f,t+1} = -\frac{\sigma_i^2}{2} + \theta \frac{\sigma_{ic}}{\psi} + (1 - \theta) \sigma_{im}$$

The Epstein-Zin preferences

- With $\theta = 1$ the model collapses to the loglinear CCAPM.
- With $\gamma = 1$ ($\theta = 0$) the logarithmic version of the static CAPM holds
- Otherwise, the E-Z is a linear combination of the CAPM and CCAPM
- Dividing one by the expected value of the pricing kernel yields

$$r_{f,t+1} = -\log\beta + \frac{1}{\psi} E_t(c_{t+1}) + \frac{\theta-1}{2} \sigma_m^2 - \frac{\theta}{2\psi^2} \sigma_c^2$$

- Notice that the log risk free rate under E-Z preferences is no longer the reciprocal of EIS (ψ).
- This is important to disentangle the equity premium and risk free rate puzzles.
- Indeed, recursive preferences perform better in matching the data.
- The long run risk model (coming next) has employed the Epstein Zin preferences.
- In addition, multiple papers on long run asset allocation (see e.g., the textbook treatment of Campbell and Viceira) have formulated investor's preferences using EZ.

The Long Run Risk model

- The LRR of Bansal and Yaron (2004) has been one of the most successful asset pricing theory over the last decade.
- LRR models feature a small but highly persistent component in consumption growth 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)).
- The evidence is based on calibration experiments.

Long Run Risk Model

- The aggregate log consumption and dividend growth rates, $g_{c,t+1}$ and $g_{d,t+1}$, contain a small persistent component, g_t , and a fluctuating component reflecting economic uncertainty, σ_t

$$g_{t+1} = \rho g_t + \varphi_e \sigma_t e_{t+1}$$

$$g_{c,t+1} = \mu_c + g_t + \sigma_t \eta_{t+1}$$

$$g_{d,t+1} = \mu_d + \phi g_t + \varphi_d \sigma_t u_{t+1} + \pi \sigma_t \eta_{t+1}$$

$$\sigma_{t+1}^2 = \sigma^2 + v_1(\sigma_t^2 - \sigma^2) + \sigma_w w_{t+1}$$

- The shocks e_{t+1} , η_{t+1} , u_{t+1} , and w_{t+1} are iid normal
- To produce an equity risk premium that is consistent with the data, the Epstein-Zin investor must have preferences such that $\gamma > 1/\psi$
- Under these conditions, the price-dividend ratio tends to be high when expected consumption and dividend growth rates are high due to the persistent component g_t and when economic uncertainty σ_t^2 is relatively low.

Long Run Risk Model

- Bansal and Yaron consider the E-Z pricing kernel representation derived earlier

$$E_t \exp \left[\theta \ln \beta - \frac{\theta}{\psi} \ln \left(\frac{C_{t+1}}{C_t} \right) + (\theta - 1) \ln R_{m,t+1} + \ln R_{i,t+1} \right] = 1$$

- Their notation : $E_t \exp \left[\theta \ln \delta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{a,t+1} + r_{i,t+1} \right] = 1$
- Notice that $r_{a,t+1}$ is the log return on an asset that delivers aggregate consumption as its dividends each period.
- Need to solve for $r_{a,t+1}$ using the Campbell-Shiller present value formula to be developed here

$$r_{a,t+1} = k_0 + k_1 z_{t+1} - z_t + g_{t+1}$$

where z is the log price-to-dividend ratio.

- Then they guess $z_t = A_0 + A_1 x_t + A_2 \sigma_t^2$ and substitute this into the Euler equation.

Long Run Risk Model

- Ferson, Nallareddy, and Xie (2011) examine the out of sample performance of the LRR paradigm.
- They examine both the 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.
- Beeler and Campbell (2012) display several weaknesses of LRR models:
 - Too strong predictability of consumption growth by the price-dividend ratio
 - Too small variance risk premium
 - Too strong predictive power of return volatility by the price-dividend ratio
 - Too small discount rate risk versus cash flow risk
- In response, Zhou and Zhu (2013) propose an extra volatility factor that helps resolve these weaknesses

The habit formation model

- In the literature on habit formation there are two types of models that govern the specification of the habit level.
- In the internal habit models, habit is a function of previous consumption path.
- In the external habit models, habit is a function of the consumption of a peer group or the aggregate consumption
- Another modelling twist is about how consumption is related to the habit level.
- Abel (1990) and Chan and Kogan (2002) are examples of ratio models in which utility is defined over the ratio of consumption to habit
- Constantinides (1990) and Campbell and Cochrane (1999) are examples of difference models wherein utility is defined over the difference between consumption and habit.
- Importantly, the relative risk aversion is constant in ratio models but time-varying in difference models.

The habit formation model

- Time-varying relative risk aversion implies that the price of risk is time-varying
- This feature lines up with the ever growing literature on time-series predictability.
- In difference habit models, expected returns are time varying due to changes in the effective risk aversion of the agents as we show below.
- Following extended periods of low consumption growth, investors require higher risk premiums as consumption approaches the habit.
- This variation in discount rates in the model produces stock market returns that are predictable using the price-dividend ratio
- In the Campbell and Cochrane (1999) model, agents in the model maximize expected utility

$$E \left[\sum_{t+1}^{\infty} \left(\delta^t \frac{(C_t - H_t)^{1-\gamma}}{1-\gamma} \right) \right]$$

where C_t is consumption and H_t is the habit level

The habit formation model

- Aggregate consumption growth and dividend growth follow lognormal processes,

$$\begin{aligned}\log(\bar{C}_{t+1}/\bar{C}_t) &= g_c + \sigma_c \eta_{t+1} \\ \log(\bar{D}_{t+1}/\bar{D}_t) &= g_d + \sigma_d \epsilon_{t+1}\end{aligned}$$

where \bar{D}_{t+1} is the aggregate dividend on stocks. The innovations η_{t+1} and ϵ_{t+1} are iid and come from a standard bivariate normal distribution with correlation ω

- The consumption of the identical agents must equal aggregate consumption in equilibrium,

$$C_t = \bar{C}_t$$

- The surplus ratio is the sole state variable that describes time variation in valuation levels and expected returns in the model.
- The price-dividend ratio is an increasing function of the surplus ratio
- Expected returns are decreasing in the surplus of consumption over the habit level.
- Expected returns are thus a decreasing function of the price-dividend ratio

The habit formation model

- The evolution of the habit can be described by the log surplus ratio

$$\bar{s}_t = \log \left(\frac{\bar{c}_t - H_t}{\bar{c}_t} \right)$$

which measures the surplus of aggregate consumption over the habit level

- The log surplus ratio follows the process

$$\bar{s}_{t+1} = (1 - \phi)\bar{s} + \phi\bar{s}_t + \lambda(\bar{s}_t)(\bar{c}_{t+1} - \bar{c}_t - g_c)$$

where ϕ governs the persistence and \bar{s} governs the long-run mean of the log surplus ratio, g_c is expected consumption growth, and $\lambda(\bar{s}_t)$ is a sensitivity function that modulates the effect of unexpected consumption growth on the habit.

The habit formation model

- In what follows, it is useful to consult the class notes by Simone Gilchrise: 'Asset pricing models with Habit Formation'
- Campbell and Cochrane use the following sensitivity function:

$$\lambda(\bar{s}_t) = \frac{1}{\bar{S}} \sqrt{1 - 2(\bar{s}_t - \bar{S})} - 1$$

where $0 \leq \bar{S} \leq s_{\max}$ and \bar{S} and s_{\max} are respectively the steady-state and upper bound of the surplus-consumption ratio given by

$$\begin{aligned}\bar{S} &= \sigma \sqrt{\frac{\gamma}{1-\phi}} \\ s_{\max} &= \bar{S} + \frac{1 - \bar{S}^2}{2}\end{aligned}$$

- These two values for \bar{S} and s_{\max} are one possible choice that Campbell and Cochrane justify to make the habit locally predetermined.

The habit formation model

- This sensitivity function allows them to have a constant risk-free interest rate.
- To see this, note that the risk-free rate is

$$r_{t+1}^f = -\log \beta + \gamma g - \gamma(1 - \phi)(\bar{s}_t - \bar{s}) - \frac{\gamma^2 \sigma^2}{2} (1 + \lambda(\bar{s}_t))^2$$

- Two effects where s_t appears: intertemporal substitution and precautionary savings.
- When s_t is low, households have a low IES which drives the risk free rate up.
- High risk aversion induces more precautionary savings which drives the risk free rate down
- Campbell and Cochrane offset these two effects by picking λ such that

$$\gamma(1 - \phi)(\bar{s}_t - \bar{s}) + \frac{\gamma^2 \sigma^2}{2} (1 + \lambda(\bar{s}_t))^2 = \text{constant}$$

- If agent's utility is $v(C, H) = u(C - H)$ instead of $u(C)$ and H grows over time so that its distance to C is always rather small, then a given percentage change in C generates a larger percentage change in $C - H$:

$$\frac{\Delta(C - H)}{C - H} = \frac{\Delta C}{C} \frac{C}{C - H} > \frac{\Delta C}{C}$$

The habit formation model

- This is just a “leverage” effect coming from the “subsistence level” H .
- Hence for a given volatility of C , we get more volatility of marginal utility of consumption $\frac{du}{dC}$.
- This allows to come closer to the Hansen-Jagannathan bounds: marginal utility of consumption is volatile, which is essential to resolve the equity premium puzzle.
- When agents’ consumption becomes closer to the habit level h , they fear further negative shocks since their utility is concave
- The relative risk aversion is time varying:

$$RRA(C) = \frac{-Cv'''(C)}{v'(C)} = \frac{-Cu'''(C-H)}{u'(C-H)}$$

- With preferences given by $u(C) = \frac{C^{1-\gamma}}{1-\gamma}$, direct calculation yields

$$RRA(C) = \gamma \frac{C}{C-H}$$

- As $C \rightarrow H$, $RRA(C) \rightarrow \infty$.
- Hence “time-varying risk aversion”, and hence time-varying risk premia

Key mechanism

- Time-varying local risk-aversion coefficient:

$$\gamma_t = -\frac{CU_{CC}}{U_C} = \frac{\gamma}{\bar{S}_t}$$

- Counter-cyclical market price of risk.
- To show it, let us start from

$$SR_t = \left| \frac{E_t(R_{t+1}^e)}{\sigma_t(R_{t+1}^e)} \right| \leq \frac{\sigma_t(M_{t+1})}{E_t(M_{t+1})} = MPR_t$$

- MPR_t is the market price of risk.
- As noted earlier, equality holds for assets that are perfectly correlated with the SDF.
- In this model the market price of risk is:

$$MPR_t = \gamma\sigma(1 + \lambda(\bar{s}_t)) = \frac{\gamma\sigma}{\bar{S}} \sqrt{1 - 2(\bar{s}_t - \bar{s})}$$

- At the steady-state, $\bar{SR} = \frac{\gamma\sigma}{\bar{S}}$, but the market price of risk is countercyclical, and hence so is the Sharpe ratio.

Key mechanism

- Several more insights:
- Volatility of returns is higher in bad times
- The model also matches the time-series predictability evidence: dividend growth is not predictable, but returns are predictable, and the volatility of the price-to-dividend ratio is accounted for by this latter term (“discount rate news”).
- Long-run equity premium: because of mean-reversion in stock prices, excess returns on stocks at long horizons are even more puzzling than the standard one-period ahead puzzle.
- Campbell and Cochrane note that if the state variable is stationary, the long-run standard deviation of the SDF will not depend on the current state. Key point: in their model, $S^{-\gamma}$ is not stationary – variance is growing with horizon!

The prospect theory model

- In the prospect theory model of Barberis et al. (2001), identical agents extract utility not only from their consumption but also from fluctuations in their financial wealth.
- Prospect theory investors are loss averse, as they are more concerned about losses than gains.
- Investors track their gains and losses relative to a slow-moving benchmark, and their effective risk aversion is higher (lower) when they have accumulated losses (gains).
- This specification triggers intertemporal variation in risk aversion and more volatile asset prices relative to the benchmark case with symmetric preferences
- Formally, prospect theory agents maximize utility of the form

$$E \left[\sum_{t+1}^{\infty} \left(\delta^t \frac{C_t^{1-\gamma}}{1-\gamma} + b_0 \bar{C}_t^{-\gamma} \delta^{t+1} v(X_{t+1}, S_t, s_t^*) \right) \right]$$

- The first part of the utility function corresponds to standard power utility over the agent's consumption
- The second part reflects loss aversion preferences
- The $b_0 \bar{C}_t^{-\gamma}$ term is a scaling factor

The prospect theory model

- The function v depends on
 - The value of the agent's stock holdings S_t
 - The change in financial wealth $X_{t+1} \equiv S_t R_{t+1} - S_t R_{f,t}$ (where R_{t+1} and $R_{f,t}$ are the returns on stocks and the risk-free asset)
 - s_t^* which is the historical benchmark level of stocks S_t^* given as a fraction of the stock value S_t (i.e., $s_t^* \equiv S_t^*/S_t$)
- The state variable s_t^* is assumed to sluggishly evolve and is modeled by

$$s_{t+1}^* = \eta \left(s_t^* \frac{\bar{R}}{R_{t+1}} \right) + (1 - \eta)$$

which \bar{R} is a parameter chosen such that s_t^* is equal to one on average

- $\eta \in [0,1]$ corresponds to the memory of the agents
- If $\eta = 0$ the benchmark level quickly adapts and is equal to the stock price at every time t , $S_t^* = S_t$
- When η is greater than zero, however, the benchmark level reflects a longer memory of the agent with respect to past gains and losses

The prospect theory model

- Overall, the function v captures loss aversion such that agents are more sensitive to losses below their historical benchmark.
- Aggregate consumption and dividend growth rates follow the same iid processes as in the habit formation model

$$\begin{aligned} \log(\bar{C}_{t+1}/\bar{C}_t) &= g_c + \sigma_c \eta_{t+1} \\ \log(\bar{D}_{t+1}/\bar{D}_t) &= g_d + \sigma_d \epsilon_{t+1} \end{aligned}$$

- Given that dividend growth is IID, all variation in the price-dividend ratio is driven through a valuation channel with time-varying expected returns.
- For example, the gains from an unexpected positive dividend shock reduce effective risk aversion.
- A corresponding decrease in expected returns is accompanied by an increase in the price-dividend ratio that amplifies the effects of dividend shocks in the model.
- In equilibrium, the price-dividend ratio and expected returns are functions of the sole state variable s_t^* .

When prospect theory meets data

- Barberis, Mukherjee, and Wang (2015) develop testable predictions of prospect theory.
- They outline the challenges in applying prospect theory outside the laboratory.
- In particular, prospect theory entails two steps.
- First, since a prospect theory agent is assumed to derive utility from gains and losses, the agent forms a mental representation of the gains and losses characterizing the random outcome.
- In experimental settings, the answer is clear: laboratory subjects are typically *given* a representation for any risk they are asked to consider – a 50:50 bet to win \$110 or lose \$100, say.
- Outside the laboratory, however, the answer is less clear: how does an investor who is thinking about a stock represent that stock in his mind?
- Second, the agent evaluates this representation – this distribution of gains and losses – to see if it is appealing.
- The valuation step is straightforward: Tversky and Kahneman (1992) provide detailed formulas that specify the value that a prospect theory agent would assign to any given distribution of gains and losses.

When prospect theory meets data

- Barberis et al suggest that, for many investors, their mental representation of a stock is given by the *distribution of the stock's past returns*.
- Indeed, the past return distribution is a good and easily accessible proxy for the object agents are interested in - the distribution of the stock's *future* returns.
- This belief may be mistaken: a stock with a high mean return over the past few years typically could have low subsequent return (De Bondt and Thaler, 1985); and a stock whose past returns are highly skewed need not exhibit high skewness in its future returns.
- Nonetheless, many investors may *think* that a stock's past return distribution is a good approximation of its future return distribution, and therefore adopt the past return distribution as their mental representation of the stock.

Estimating and Evaluating Asset Pricing Models

Why Caring About Asset Pricing Models

- An essential question that arises is why both academics and practitioners invest huge resources attempting to develop and test of asset pricing models.
- It turns out that pricing models have crucial roles in various applications in financial economics – both asset pricing as well as corporate finance.
- In the following, I list five major applications.

1 – Common Risk Factors

- Pricing models characterize the risk profile of a firm.
- In particular, systematic risk is no longer stock return volatility – rather it is the loadings on risk factors.
- For instance, in the single factor CAPM the market beta – or the co-variation with the market – characterizes the systematic risk of the firm
- Likewise, in the single factor CCAPM the consumption growth beta – or the co-variation with consumption growth – characterizes the systematic risk of the firm.
- In the multi-factor Fama-French (FF) model there are three sources of risk – the market beta, the SMB beta, and the HML beta.
- Under FF, other things being equal (*ceteris paribus*), a firm is riskier if its loading on SMB beta is higher.
- Under FF, other things being equal (*ceteris paribus*), a firm is riskier if its loading on HML beta is higher

2 – Moments for Asset Allocation

- ▶ Pricing models deliver moments for asset allocation.
- ▶ For instance, the tangency portfolio takes on the form

$$w_{TP} = \frac{V^{-1}\mu^e}{\iota'V^{-1}\mu^e}$$

- ▶ Under the CAPM, the vector of expected returns and the covariance matrix are given by:

$$\mu^e = \beta\mu_m^e$$

$$V = \beta\beta'\sigma_m^2 + \Sigma$$

where Σ is the covariance matrix of the residuals in the time-series asset pricing regression.

- ▶ The corresponding quantities under the FF model are

$$\mu^e = \beta_{MKT}\mu_m^e + \beta_{SML}\mu_{SML} + \beta_{HML}\mu_{HML}$$

$$V = \beta\Sigma_F\beta' + \Sigma$$

where Σ_F is the covariance matrix of the market, size, and book-t-market factors.

3 – Discount Factor

- Expected return is the discount factor, commonly denoted by k , in present value formulas in general and firm evaluation in particular:

$$PV = \sum_{t=1}^T \frac{CF_t}{(1+k)^t}$$

- In practical applications, expected returns are typically assumed to be constant over time, an unrealistic assumption.
- Indeed, thus far we have examined models with constant beta and constant risk premiums

$$\mu^e = \beta' \lambda$$

where λ is a K-vector of risk premiums.

- When factors are return spreads the risk premium is the mean of the factor.
- Later we will consider models with time varying factor loadings.

4 - Benchmarks

- Factors in asset pricing models could serve as benchmarks for evaluating performance of active investments.
- In particular, performance is the intercept (alpha) in the time series regression of excess fund returns on a set of benchmarks (typically four benchmarks in mutual funds and more so in hedge funds):

$$r_t^e = \alpha + \beta_{MKT} \times r_{MKT,t}^e + \beta_{SMB} \times SMB_t + \beta_{HML} \times HML_t + \beta_{WML} \times WML_t + \varepsilon_t$$

5 - Corporate Finance

- There is a plethora of studies in corporate finance that use asset pricing models to risk adjust asset returns.
- Here are several examples:
 - Examining the long run performance of IPO firm.

5 - Corporate Finance

- Examining the long run performance of SEO firms
- Analyzing abnormal performance of stocks going through splits and reverse splits.
- Analyzing mergers and acquisitions
- Analyzing the impact of change in board of directors.
- Studying the impact of corporate governance on the cross section of average returns.
- Studying the long run impact of stock/bond repurchase.

Methods of evaluation model pricing abilities

- The finance literature has used three main approaches to evaluate asset pricing models: calibration, cross sectional and time series asset pricing tests, and out-of-sample fit.
- With calibration (e.g., Long run risk, habit formation, prospect theory), values for the parameters of the underlying model are chosen, and the model is solved at these parameter values for the prices of financial assets.
- The model-generated series of prices and returns are examined to see if their moments match key moments of actual asset prices.
- In asset pricing tests, model parameters are optimally chosen to fit a panel of economic series and asset returns.
- Standard errors for the parameter estimates quantify their precision.
- Statistical hypothesis tests about the parameters are conducted and the residuals of the model are examined to assess the fit to the sample.
- Estimation typically challenges a model in more dimensions at once than calibration. For example, a calibrated parameter value may not be the value that maximizes the likelihood, indicating that more issues are going on in the data than captured by the calibration.

- The practical utility of an asset pricing model ultimately depends on its ability to fit out-of-sample returns, as most practical applications are, in some sense, out of sample
- For example, firms want to estimate costs of capital for future projects, portfolio and risk managers want to know the expected compensation for future risks, and academic researchers will want to make risk adjustments to expected returns in future data.
- Many of these applications rely on out-of-sample estimates for the required or ex ante expected return, where the model parameters are chosen based on available data.
- This perspective leads naturally to the mean squared pricing error (MSE) criterion: a better model produces lower MSE.
- Ex post out-of-sample performance of asset allocation decisions would establish a solid economic metric for model pricing abilities.
- More on asset allocation is in the Bayesian section of the class notes.
- This section provides the econometric paradigms of asset pricing tests.

Time Series Tests

- ▶ Time series tests are designated to examine the validity of models in which factors are portfolio based, or factors that are return spreads.
- ▶ Example: the market factor is the return difference between the market portfolio and the risk-free asset.
- ▶ Consumption growth is not a return spread.
- ▶ Thus, the consumption CAPM cannot be tested using time series regressions, unless you form a factor mimicking portfolio (FMP) for consumption growth
- ▶ FMP is a convex combination of asset returns having the maximal correlation with consumption growth.
- ▶ The statistical time series tests have an appealing economic interpretation
- ▶ Testing the CAPM amounts to testing whether the market portfolio is the tangency portfolio.
- ▶ Testing multi-factor models amounts to testing whether some optimal combination of the factors is the tangency portfolio.

Testing the CAPM

- ▶ Run the time series regression:

$$r_{1t}^e = \alpha_1 + \beta_1 r_{mt}^e + \varepsilon_{1t}$$

\vdots

$$r_{Nt}^e = \alpha_N + \beta_N r_{mt}^e + \varepsilon_{Nt}$$

- ▶ The null hypothesis is:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_N = 0$$

- ▶ In the following, four times series test statistics will be described:

- ▶ WALD;
- ▶ Likelihood Ratio;
- ▶ GRS (Gibbons, Ross, and Shanken (1989));
- ▶ GMM.

The Distribution of a

- Recall, a is asset mispricing.
- The time series regressions can be rewritten using a vector form as:

$$\underset{N \times 1}{r_t^e} = \underset{N \times 1}{\alpha} + \underset{N \times 1}{\beta} \cdot \underset{1 \times 1}{r_{mt}^e} + \underset{N \times 1}{\varepsilon_t}$$

- Let us assume that

$$\underset{N \times 1}{\varepsilon_t} \overset{iid}{\sim} N\left(0, \underset{N \times N}{\Sigma}\right)$$

for $t = 1, 2, 3, \dots, T$

- Let $\Theta = (\alpha', \beta', \text{vech}(\varepsilon)')'$ be the set of all parameters.
- Under normality, the likelihood function for ε_t is

$$L(\varepsilon_t | \theta) = c |\Sigma|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (r_t^e - \alpha - \beta r_{mt}^e)' \Sigma (r_t^e - \alpha - \beta r_{mt}^e) \right]$$

where c is the constant of integration (recall the integral of a probability distribution function is unity).

- Moreover, the IID assumption suggests that

$$L(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N | \theta) = c^T |\Sigma|^{-\frac{T}{2}} \\ \times \exp \left[-\frac{1}{2} \sum_{t=1}^T (r_t^e - \alpha - \beta r_{mt}^e)' \Sigma^{-1} (r_t^e - \alpha - \beta r_{mt}^e) \right]$$

- Taking the natural log from both sides yields

$$\ln(L) \propto -\frac{T}{2} \ln(|\Sigma|) - \frac{1}{2} \sum_{t=1}^T (r_t^e - \alpha - \beta r_{mt}^e)' \Sigma^{-1} (r_t^e - \alpha - \beta r_{mt}^e)$$

- Asymptotically, we have $\theta - \hat{\theta} \sim N(0, \Sigma(\theta))$

where

$$\Sigma(\theta) = \left[-E \left[\frac{\partial^2 \ln(L)}{\partial \theta \partial \theta'} \right] \right]^{-1}$$

- Let us estimate the parameters

$$\frac{\partial \ln(L)}{\partial \alpha} = \Sigma^{-1} \left[\sum_{t=1}^T (r_t^e - \alpha - \beta r_{mt}^e) \right]$$

$$\frac{\partial \ln(L)}{\partial \beta} = \Sigma^{-1} \left[\sum_{t=1}^T (r_t^e - \alpha - \beta r_{mt}^e) \times r_{mt}^e \right]$$

$$\frac{\partial \ln(L)}{\partial \Sigma} = -\frac{T}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \left[\sum_{t=1}^T \varepsilon_t \varepsilon_t' \right] \Sigma^{-1}$$

- Solving for the first order conditions yields

$$\hat{\alpha} = \hat{\mu}^e - \hat{\beta} \cdot \hat{\mu}_m^e$$

$$\hat{\beta} = \frac{\sum_{t=1}^T (r_t^e - \hat{\mu}^e) (r_{mt}^e - \hat{\mu}_m^e)}{\sum_{t=1}^T (r_{mt}^e - \hat{\mu}_m^e)^2}$$

➤ Moreover,

$$\begin{aligned}\hat{\Sigma} &= \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t' \\ \hat{\mu}^e &= \frac{1}{T} \sum_{t=1}^T r_t^e \\ \mu_m^e &= \frac{1}{T} \sum_{t=1}^T r_{mt}^e\end{aligned}$$

- Recall our objective is to find the variance-covariance matrix of $\hat{\alpha}$.
- Standard errors could be found using the information matrix:

$$I(\theta) = - \left\{ E \left[\begin{array}{ccc} \frac{\partial^2 \ln(L)}{\partial \alpha \partial \alpha'} & \frac{\partial^2 \ln(L)}{\partial \alpha \partial \beta'} & \frac{\partial^2 \ln(L)}{\partial \alpha \partial \Sigma'} \\ \frac{\partial^2 \ln(L)}{\partial \beta \partial \alpha'} & \frac{\partial^2 \ln(L)}{\partial \beta \partial \beta'} & \frac{\partial^2 \ln(L)}{\partial \beta \partial \Sigma'} \\ \frac{\partial^2 \ln(L)}{\partial \Sigma \partial \alpha'} & \frac{\partial^2 \ln(L)}{\partial \Sigma \partial \beta'} & \frac{\partial^2 \ln(L)}{\partial \Sigma \partial \Sigma'} \end{array} \right] \right\}$$

The Distribution of the Parameters

- Try to establish yourself the information matrix.
- Notice that $\hat{\alpha}$ and $\hat{\beta}$ are independent of $\hat{\Sigma}$ - thus, you can ignore the second derivatives with respect to Σ in the information matrix if your objective is to find the distribution of $\hat{\alpha}$ and $\hat{\beta}$.
- If you aim to derive the distribution of $\hat{\Sigma}$ then focus on the bottom right block of the information matrix.

The Distribution of $\hat{\alpha}$

- We get:

$$\hat{\alpha} \sim N \left(\alpha, \frac{1}{T} \left[1 + \left(\frac{\hat{\mu}_m^e}{\hat{\sigma}_m} \right)^2 \right] \Sigma \right)$$

- Moreover,

$$\begin{aligned} \hat{\beta} &\sim N \left(\beta, \frac{1}{T} \cdot \frac{1}{\hat{\sigma}_m^2} \Sigma \right) \\ T\hat{\Sigma} &\sim W(T-2, \Sigma) \end{aligned}$$

- Notice that $W(x, y)$ stands for the Wishart distribution with $x = T - 2$ degrees of freedom and a parameter matrix $y = \Sigma$.

The Wald Test

➤ Recall, if

$$X \sim N(\mu, \Sigma) \text{ then } (X - \mu)' \hat{\Sigma}^{-1} (X - \mu) \sim \chi^2(N)$$

➤ Here we test

$$H_0: \hat{\alpha} = 0$$

$$H_1: \hat{\alpha} \neq 0$$

where

$$\hat{\alpha} \stackrel{H_0}{\sim} N(0, \Sigma_{\alpha})$$

➤ The Wald statistic is $\hat{\alpha}' \hat{\Sigma}_{\alpha}^{-1} \hat{\alpha} \sim \chi^2(N)$, which becomes:

$$J_1 = T \left[1 + \left(\frac{\hat{\mu}_m^e}{\hat{\sigma}_m} \right)^2 \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} = T \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + S\hat{R}_m^2}$$

where $S\hat{R}_m$ is the Sharpe ratio of the market factor.

Algorithm for Implementation

- The algorithm for implementing the statistic is as follows:
- Run separate regressions for the test assets on the common factor:

$$\begin{array}{l} r_{Tx1}^e = X_{Tx2} \theta_1 + \varepsilon_{Tx1} \\ \vdots \\ r_N^e = X_{Tx2} \theta_N + \varepsilon_N \end{array} \quad \text{where} \quad X_{Tx2} = \begin{bmatrix} 1, r_{m1}^e \\ \vdots \\ 1, r_{mT}^e \end{bmatrix} \quad \theta_i = [\alpha_i, \beta_i]'$$

- Retain the estimated regression intercepts

$$\hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_N]' \quad \text{and} \quad \hat{\varepsilon}_{TxN} = [\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N]$$

- Compute the residual covariance matrix

$$\hat{\Sigma} = \frac{1}{T} \hat{\varepsilon}' \hat{\varepsilon}$$

- Compute the sample mean and the sample variance of the factor.
- Compute J_1 .

The Likelihood Ratio Test

- We run the unrestricted and restricted specifications:

$$\text{un: } r_t^e = \alpha + \beta r_{mt}^e + \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma)$$

$$\text{res: } r_t^e = \beta^* r_{mt}^e + \varepsilon_t^* \quad \varepsilon_t^* \sim N(0, \Sigma^*)$$

- Using MLE, we get:

$$\hat{\beta}^* = \frac{\sum_{t=1}^T r_t^e r_{mt}^e}{\sum_{t=1}^T (r_{mt}^e)^2}$$

$$\hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^* \hat{\varepsilon}_t^{*'}$$

$$\hat{\beta}^* \sim N\left(\beta, \frac{1}{T} \left[\frac{1}{\hat{\mu}_m^2 + \hat{\sigma}_m^2} \right] \Sigma\right)$$

$$T\hat{\Sigma}^* \sim W(T-1, \Sigma)$$

The LR Test

$$LR = \ln(L^*) - \ln(L) = -\frac{T}{2} [\ln |\hat{\Sigma}^*| - \ln |\hat{\Sigma}|]$$

$$J_2 = -2LR = T [\ln |\hat{\Sigma}^*| - \ln |\hat{\Sigma}|] \sim \chi^2(N)$$

- Using some algebra, one can show that

$$J_1 = T \left(\exp \left(\frac{J_2}{T} \right) - 1 \right)$$

- Thus,

$$J_2 = T \cdot \ln \left(\frac{J_1}{T} + 1 \right)$$

GRS (1989) – Finite Sample Test

Theorem: let

$$\underset{N \times 1}{X} \sim N(0, \Sigma)$$

let

$$\underset{N \times N}{A} \sim W(\tau, \Sigma)$$

where $\tau \geq N$ and A and X are independent. Then:

$$\frac{\tau - N + 1}{N} X' A^{-1} X \sim F_{N, \tau - N + 1}$$

► In our context:

$$X = \sqrt{T} \left[1 + \left(\frac{\hat{\mu}_m}{\hat{\sigma}_m} \right)^2 \right]^{-\frac{1}{2}} \hat{\alpha} \stackrel{H_0}{\sim} N(0, \Sigma)$$

$$A = T\hat{\Sigma} \sim W(\tau, \Sigma)$$

where

$$\tau = T - 2$$

► Then:

$$J_3 = \left(\frac{T - N - 1}{N} \right) \left[1 + \left(\frac{\hat{\mu}_m}{\hat{\sigma}_m} \right)^2 \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F(N, T - N - 1)$$

GMM

The GMM test statistic (derivation comes up later in the notes) is given by

$$J_4 = T \hat{\alpha}' (R(D_T' S_T^{-1} D_T)^{-1} R')^{-1} \cdot \hat{\alpha} \stackrel{H_0}{\sim} \chi^2(N)$$

where

$$R_{N \times 2N} = \begin{bmatrix} I_N & 0 \\ & N \times N \end{bmatrix}$$
$$D_T = - \begin{bmatrix} 1, & \hat{\mu}_m^e \\ \hat{\mu}_m^e, & (\hat{\mu}_m^e)^2 + \hat{\sigma}_m^2 \end{bmatrix} \otimes I_N$$

- Assume no serial correlation but heteroscedasticity:

$$S_T = \frac{1}{T} \sum_{t=1}^T (x_t x_t' \otimes \hat{\varepsilon}_t \hat{\varepsilon}_t')$$

where

$$x_t = [1, r_{mt}^e]'$$

- Under homoscedasticity and serially uncorrelated moment conditions: $J_4 = J_1$.
- That is, the GMM statistic boils down to the WALD.

The Multi-Factor Version of Asset Pricing Tests

$$\underset{Nx1}{r_t^e} = \underset{Nx1}{\alpha} + \underset{NxK}{\beta} \cdot \underset{Kx1}{F_t} + \underset{Nx1}{\varepsilon_t}$$

$$J_1 = T(1 + \hat{\mu}_F' \hat{\Sigma}_F^{-1} \hat{\mu}_F)^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim \chi(N)$$

J_2 follows as described earlier.

$$J_3 = \frac{T - N - K}{N} (1 + \hat{\mu}_F' \hat{\Sigma}_F^{-1} \hat{\mu}_F)^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F_{(N, T-N-K)}$$

where $\hat{\mu}_F$ is the mean vector of the factor based return spreads

- $\hat{\Sigma}_F$ is the variance covariance matrix of the factors.
- For instance, considering the Fama-French model:

$$\hat{\mu}_F = \begin{bmatrix} \hat{\mu}_m^e \\ \hat{\mu}_{SMB} \\ \hat{\mu}_{HML} \end{bmatrix} \quad \hat{\Sigma}_F = \begin{bmatrix} \hat{\sigma}_m^2, \hat{\sigma}_{m,SMB}, \hat{\sigma}_{m,HML} \\ \hat{\sigma}_{SMB,m}, \hat{\sigma}_{SMB}^2, \hat{\sigma}_{SMB,HML} \\ \hat{\sigma}_{HML,m}, \hat{\sigma}_{HML,SMB}, \hat{\sigma}_{HML}^2 \end{bmatrix}$$

The Economics of Time Series Test Statistics

- Let us summarize the first three test statistics:

$$J_1 = T \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + S \hat{R}_m^2}$$

$$J_2 = T \cdot \ln \left(\frac{J_1}{T} + 1 \right)$$

$$J_3 = \frac{T - N - 1}{N} \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + S \hat{R}_m^2}$$

The Economics of Time Series Test Statistics

- The J_4 statistic, the GMM based asset pricing test, is actually a Wald test, just like J_1 , except that the covariance matrix of asset mispricing takes account of heteroscedasticity and often even potential serial correlation.
- Notice that all test statistics depend on the quantity

$$\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

- GRS show that this quantity has a very insightful representation.
- Let us provide the steps.

Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

- Consider an investment universe that consists of $N + 1$ assets - the N test assets as well as the market portfolio.
- The expected return vector of the $N + 1$ assets is given by

$$\hat{\lambda}_{(N+1) \times 1} = \begin{bmatrix} \hat{\mu}_m^e & \hat{\mu}^{e'} \end{bmatrix}'$$

where $\hat{\mu}_m^e$ is the estimated expected excess return on the market portfolio and $\hat{\mu}^e$ is the estimated expected excess return on the N test assets.

Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

- The variance covariance matrix of the $N+1$ assets is given by

$${}_{(N+1) \times (N+1)} \hat{\Phi} = \begin{bmatrix} \hat{\sigma}_m^2 & \hat{\beta}' \hat{\sigma}_m^2 \\ \hat{\beta} \hat{\sigma}_m^2 & \hat{V} \end{bmatrix}$$

where $\hat{\sigma}_m^2$ is the estimated variance of the market factor, $\hat{\beta}$ is the N -vector of market loadings, and \hat{V} is the covariance matrix of the N test assets.

- Notice that the covariance matrix of the N test assets is

$$\hat{V} = \hat{\beta} \hat{\beta}' \hat{\sigma}_m^2 + \hat{\Sigma}$$

- The squared tangency portfolio of the $N + 1$ assets is

$$S\hat{R}_{TP}^2 = \hat{\lambda}' \hat{\Phi}^{-1} \hat{\lambda}$$

- Notice also that the inverse of the covariance matrix is

$$\hat{\Phi}^{-1} = \begin{bmatrix} (\hat{\sigma}_m^2)^{-1} + \hat{\beta}' \hat{\Sigma}^{-1} \hat{\beta} & -\hat{\beta}' \hat{\Sigma}^{-1} \\ -\hat{\Sigma}^{-1} \hat{\beta} & \hat{\Sigma}^{-1} \end{bmatrix}$$

Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

- Thus, the squared Sharpe ratio of the tangency portfolio could be represented as

$$S\hat{R}_{TP}^2 = \left(\frac{\hat{\mu}_m^e}{\hat{\sigma}_m} \right)^2 + \left[(\hat{\mu}^e - \hat{\beta}\hat{\mu}_m^e)' \hat{\Sigma}^{-1} (\hat{\mu}^e - \hat{\beta}\hat{\mu}_m^e) \right]$$

$$S\hat{R}_{TP}^2 = S\hat{R}_m^2 + \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

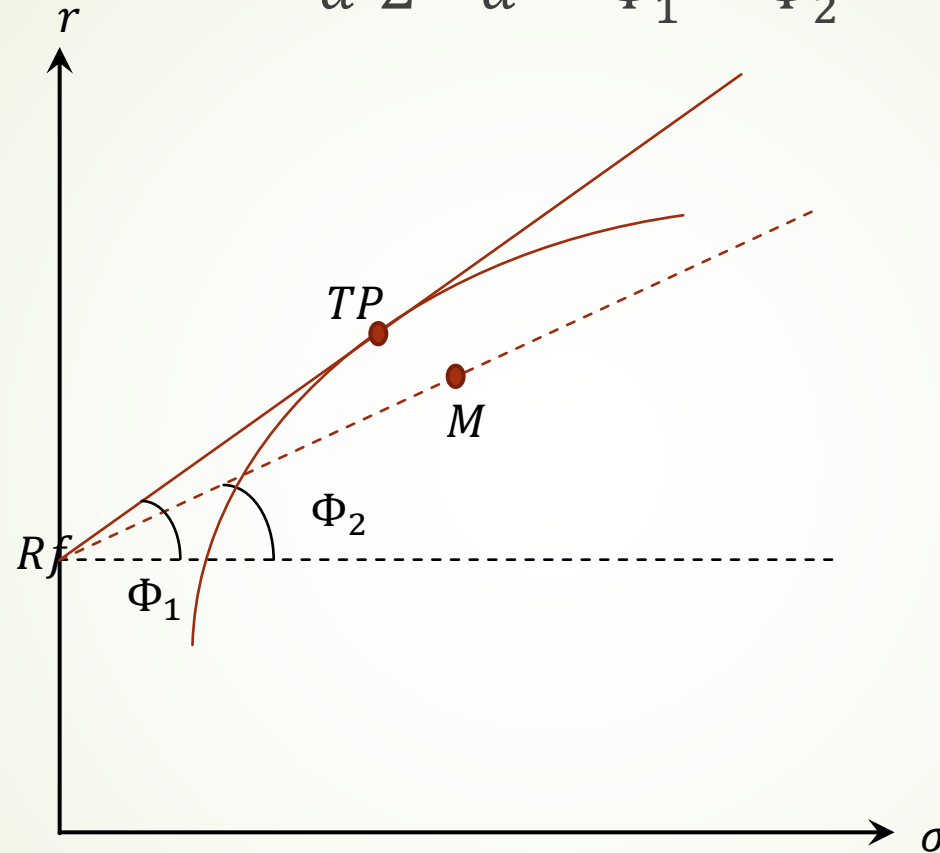
or

$$\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} = S\hat{R}_{TP}^2 - S\hat{R}_m^2$$

- In words, the $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ quantity is the difference between the squared Sharpe ratio based on the $N + 1$ assets and the squared Sharpe ratio of the market portfolio.
- If the CAPM is correct then these two Sharpe ratios are identical in population, but not identical in sample due to estimation errors.
- The test statistic examines how close the two sample Sharpe ratios are.
- Under the CAPM, the extra N test assets do not add anything to improving the risk return tradeoff.
- The geometric description of $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ is given in the next slide.

Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

$$\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} = \Phi_1^2 - \Phi_2^2$$



Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

- So we can rewrite the previously derived test statistics as

$$J_1 = T \frac{S\hat{R}_{TP}^2 - S\hat{R}_m^2}{1 + S\hat{R}_m^2} \sim \chi^2(N)$$

$$J_3 = \frac{T - N - 1}{N} \times \frac{S\hat{R}_{TP}^2 - S\hat{R}_m^2}{1 + S\hat{R}_m^2} \sim F(N, T - N - 1)$$

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.
- 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.

Cross Sectional Regressions

- 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

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

- Then run the cross-section regression of average returns on the betas

$$\bar{r}_i = \beta_i' \lambda + v_i.$$

- Notation: λ – the regression coefficient – is the risk premium, and v_i — the regression disturbance – is the pricing error.
- Assume for analytic tractability that there is a single factor, let $\bar{r} = [\bar{r}_1, \bar{r}_2, \dots, \bar{r}_N]'$, and let

$$\beta = [\beta_1, \beta_2, \dots, \beta_N]'$$

Cross Sectional Regressions

- The OLS cross-sectional estimates are

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

$$\hat{v} = \bar{r} - \hat{\lambda} \beta.$$

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

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

$$\text{cov}(\hat{v}) = \frac{1}{T} (I - \beta (\beta' \beta)^{-1} \beta') \Sigma (I - \beta (\beta' \beta)^{-1} \beta').$$

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

$$\hat{v}' \text{cov}(\hat{v})^{-1} \hat{v} \sim \chi_{N-1}^2.$$

- We could also test whether a factor is priced

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

- Notice that we assume β is known.

- However, β is estimated in the time-series regression and therefore is unknown.

- So we have the EIV problem.

Cross Sectional Regressions

- 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

$$\sigma_{eiv}^2(\hat{\lambda}) = \sigma^2(\hat{\lambda})Y + \frac{1}{T}\Omega_f$$

$$\text{cov}_{eiv}(\hat{v}) = \text{cov}(\hat{v})Y,$$

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

- 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.

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

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

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

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

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

Fama and MacBeth (FM) Procedure

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

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

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

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

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

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

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

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

- Then we can test whether all pricing errors are zero using the WALD test statistic.

Avramov and Chordia (AC) Procedure

- Let us now present the methodology in Avramov and Chordia (2006) for asset pricing tests based on individual stocks.
- In particular, assume that returns are generated by a conditional version of a K -factor model

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

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

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

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

Avramov and Chordia (AC) Procedure

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

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

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.

- The 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, run the cross-sectional regression

$$R_{jt}^* = c_{0t} + \sum_{m=1}^M c_{mt} Z_{mjt-1} + e_{jt},$$

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 ($R_{jt}^* = c_{0t} + \sum_{m=1}^M c_{mt} Z_{mjt-1} + e_{jt}$).

Avramov and Chordia (AC) Procedure

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

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

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 using the generic form

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

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.

Avramov and Chordia (AC) Procedure

- 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

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

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

Avramov and Chordia (AC) Procedure

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

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

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

- Then, R_{jt}^* in $(R_{jt}^* = c_{0t} + \sum_{m=1}^M c_{mt} Z_{mjt-1} + e_{jt})$, the dependent variable in the cross-section regression, is given by $\alpha_j + u_{jt}$.

- The time series regression ($r_{jt} = \alpha_j + \beta_{j1} r_{mt} + \beta_{j2} z_{t-1} r_{mt} + \beta_{j3} Size_{jt-1} r_{mt} + \beta_{j4} z_{t-1} Size_{jt-1} r_{mt} + \beta_{j5} BM_{jt-1} r_{mt} + \beta_{j6} z_{t-1} BM_{jt-1} r_{mt} + u_{jt}$),

is run over the entire sample.

- While this entails the use of future data in calculating the factor loadings, Fama and French (1992) indicate that this forward looking does not impact any of the results.
- For perspective, it is useful to compare our approach to earlier studies

Other Procedures

- 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 the beta-scaling procedure in Avramov and Chordia 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.
- The beta pricing specification of Avramov and Chordia does not have that unconditional multifactor interpretation since the firm-level $Size_j$ and BM_j are asset specific – that is, they are uncommon across all test assets.

Understanding Stock Return Predictability by Macro-wide Variables

Stock Return Predictability

- Empirical evidence shows that returns were predictable by financial ratios, such as the price-dividend or price-earnings ratio.
- Later other variables, such as the spread between long-term and short-term bond yields, the consumption-wealth ratio, macroeconomic variables, and corporate decision variables were also shown to have predictive ability.
- The literature has expanded its interest to returns on other asset classes, such as government bonds, currencies, real estate, and commodities, and to many countries.
- Initially, the finding of predictability was interpreted as evidence against the efficient market hypothesis.
- Fama (1991) proposed the alternative explanation of time-varying expected returns.

Stock Return Predictability Based on Macro Variables

- Indeed, in the past twenty years, research in asset pricing has proposed several equilibrium models with efficient markets that generate time variation in expected returns: models with time-varying risk aversion (Campbell and Cochrane, 1999), time-varying aggregate consumption risk (Bansal and Yaron, 2004; Bansal, Kiku, and Yaron, 2009), time-varying consumption disasters (Gabaix, 2009), time-variation in risk-sharing opportunities among heterogeneous agents (Lustig and Van Nieuwerburgh, 2005), or time-variation in beliefs (Timmermann, 1993; Detemple and Murthy, 1994).
- The evidence on predictability is typically based upon the joint system
$$r_t = a + \beta' z_{t-1} + u_t$$
$$z_t = c + \rho z_{t-1} + v_t$$
- 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, based on expected stock return.

Stock Return Predictability - Is The Evidence on Predictability Robust?

- Predictability based on macro variables is still a research controversy:
 - Asset pricing theories do not identify variables that predict asset returns. More specifically, Menzly, Santos, and Veronesi (2006) like the other studies cited on the previous page provide theoretical validity for predictability – but that is an ex post
 - 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 time-series predictability based on the dividend yield tends to attenuate and even disappears after its discovery.
- Indeed, the power of macro variables to predict the equity premium substantially deteriorates during the post-discovery period.

Stock Return Predictability - Potential Data Mining

- Repeated visits of the same database lead to a problem that statisticians refer to as data mining (also model over-fitting 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 over-fitting data in tests of financial models.
- In the context of predictability, data mining has been investigated by Foster, Smith, and Whaley (FSW - 1997).

Stock Return Predictability - Foster, Smith, and Whaley (1997)

- FSW adjust the test size for potential over-fitting, using theoretical approximations as well as simulation studies.
- They assume that
 1. M potential predictors are available.
 2. All possible regression combinations are tried.
 3. Only $m < M$ predictors 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.

Stock Return Predictability - The Poor Out-of-Sample Performance of Predictive Regressions

- Bossaerts and Hillion (1999) and Goyal and Welch (2006) are a good reference. While a highly visible paper, Goyal and Welch do not tell us something new beyond what we know from Bossaerts and Hillion and Avramov (2002).
- The adjustment of the test size for specification search could 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.
- BH and GW propose examining predictability using model selection criteria.
 - Suppose there are M potential predictors: then here are 2^M competing specifications.
 - Select one specification based on model selection criteria e.g., adjusted R^2 , AIC, or SIC.
 - The winning model maximizes the tradeoff between goodness of fit and complexity.
- The selected model (regardless of the criterion used) always retains predictors – not a big surprise – indicating in-sample predictability.

Stock Return Predictability

- Implicitly you assign a $\frac{1}{2^M}$ probability that the IID model is correct - so you are biased in favor of detecting predictability.
- The out-of-sample performance of the selected model is always a disaster.
- The Bayesian approach of model averaging could improve out-of-sample performance (see Avramov (2002)).
- BMA can be extended to account for time varying parameters.
- There are three major 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 dividend growth rates. Cochrane invokes the present value relation. Coming up soon.
- Campbell 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 and considering a small subset of models.

Predictive regressions- Finite sample bias in the slope coefficients

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

$$r_t = a + \beta z_{t-1} + u_t$$

$$z_t = c + \rho z_{t-1} + v_t$$

- 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 Kandall (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

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

- The bias can easily be derived.

Finite Sample Bias

- Note that the OLS estimates of β and ρ are

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

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

where

$$R = [r_1, r_2, \dots, r_T]', \quad Z = [z_1, z_2, \dots, z_T]'$$

$$U = [u_1, u_2, \dots, u_T]', \quad V = [v_1, v_2, \dots, v_T]'$$

$$X = [\iota_T, Z_{-1}], \quad \iota_T \text{ is a } T\text{-dimension vector of ones}$$

$$Z_{-1} = [z_0, z_1, \dots, z_{T-1}]'$$

Finite Sample Bias

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

$$u_t = \frac{\sigma_{uv}}{\sigma_v^2} v_t + e_t$$

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

- Hence, the predictive regression slope can be rewritten as

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

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

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

The Campbell-Shiller (CS) present value formula

- ▶ Let R_{t+1} be the simple net return on a risky asset

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} - 1$$

- ▶ Let $r_t = \log(1 + R_t)$, $p_t = \log(P_t)$, $d_t = \log(D_t)$.
- ▶ Then,

$$\begin{aligned} r_{t+1} &= \log(P_{t+1} + D_{t+1}) - \log P_t \\ &= p_{t+1} - p_t + \log(1 + \exp(d_{t+1} - p_{t+1})) \end{aligned}$$

The CS approximation

- ▶ We are ready now to implement the CS approximation.
- ▶ In particular, we approximate the function $f(x_t) = \log[1 + \exp(x_t)]$ around \bar{x}
- ▶ The first order Taylor approximation is given by

$$f(x_t) = \log[1 + \exp(\bar{x})] + \frac{\exp(\bar{x})}{1 + \exp(\bar{x})} (x_t - \bar{x})$$

The CS approximation

➔ Hence

$$\begin{aligned} \log[1 + \exp(d_{t+1} - p_{t+1})] &\approx \log[1 + \exp(\bar{d} - \bar{p})] \\ &\quad - \frac{\exp(\bar{d} - \bar{p})}{1 + \exp(\bar{d} - \bar{p})} (\bar{d} - \bar{p}) \\ &\quad + \frac{\exp(\bar{d} - \bar{p})}{1 + \exp(\bar{d} - \bar{p})} (d_{t+1} - p_{t+1}) \end{aligned}$$

➔ Letting $\rho \equiv \frac{1}{1 + \exp(\bar{d} - \bar{p})}$, it follows that

$$\begin{aligned} 1 - \rho &= \frac{\exp(\bar{d} - \bar{p})}{1 + \exp(\bar{d} - \bar{p})} \\ \bar{d} - \bar{p} &= \log\left(\frac{1}{\rho} - 1\right) \end{aligned}$$

➔ We get

$$\log[1 + \exp(d_{t+1} - p_{t+1})] \approx$$

$$-\log \rho - (1 - \rho) \log\left(\frac{1}{\rho} - 1\right) + (1 - \rho)(d_{t+1} - p_{t+1})$$

The CS approximation

- Now let

$$k \equiv -\log \rho - (1 - \rho) \log \left(\frac{1}{\rho} - 1 \right)$$

- And the realized return could be approximated as

$$r_{t+1} \approx k + \rho p_{t+1} + (1 - \rho) d_{t+1} - p_t$$

- Rearranging,

$$p_t \approx k + \rho p_{t+1} - r_{t+1} + (1 - \rho) d_{t+1}$$

$$p_t - d_t \approx k + \rho(p_{t+1} - d_{t+1}) - r_{t+1} + d_{t+1} - d_t$$

- If the *dividend-to-price* ratio is constant then the approximation holds exactly with

$$\rho = \frac{1}{1 + \frac{D}{P}}$$

- Iterating forward and assuming that $\lim_{j \rightarrow \infty} \rho^j p_{t+j} = 0$ (no bubbles), we get

$$p_t - d_t = \frac{k}{1 - \rho} + \sum_{j=0}^{\infty} \rho^j (\Delta d_{t+1+j} - r_{t+1+j})$$

The CS approximation

- ▶ This equation holds ex post and also ex ante, conditioning both left and right hand side of the equation by time t information.
- ▶ Thus, variation in the *dividend-to-price* ratio predicts dividend growth or expected returns.
- ▶ This point is central in the literature about equity premium predictability.
- ▶ To illustrate, using the present value model, for finite horizon, we have

$$dp_t = d_t - p_t = \sum_{j=1}^J \rho^{j-1} r_{t+j} - \sum_{j=1}^J \rho^{j-1} \Delta d_{t+j} + \rho^J (p_{t+J} - d_{t+J})$$

- ▶ Thus,

$$\text{var}(dp_t) = \text{cov}\left(dp_t, \sum_{j=1}^J \rho^{j-1} r_{t+j}\right) - \text{cov}\left(dp_t, \sum_{j=1}^J \rho^{j-1} \Delta d_{t+j}\right) + \rho^J \text{cov}(dp_t, dp_{t+J})$$

- ▶ Dividing by $\text{var}(dp_t)$ yields

$$1 = b_r^{(J)} - b_{rd}^{(J)} + \rho^J b_{dp}^{(J)}$$

where the b s on the right hand side are the slope coefficients in the regressions:

$$\begin{aligned}\sum_{j=1}^J \rho^{j-1} r_{t+j} &= a_r + b_r^{(J)} dp_t + \varepsilon_{t+J}^r \\ \sum_{j=1}^J \rho^{j-1} \Delta d_{t+j} &= a_d + b_d^{(J)} dp_t + \varepsilon_{t+J}^d \\ dp_{t+J} &= a_{dp} + b_{dp}^{(J)} dp_t + \varepsilon_{t+J}^{dp}\end{aligned}$$

The CS approximation

- Cochrane (2011) presents the long-run regression coefficients

	$b_r^{(J)}$	$b_d^{(J)}$	$b_{dp}^{(J)}$
Direct regression $J = 15$	1.01	-0.11	-0.11
Implemented by VAR $J = 15$	1.05	0.27	0.22
Implemented by VAR $J = \infty$	1.35	0.35	0.00

- That says that all the variation in the dividend to price ratio corresponds to variation in expected returns.
- None corresponds to variation in expected dividend growth or bubbles.
- This reinforces the notion of equity premium predictability by the dividend yield.

The CS approximation

- We can further show that the unexpected return can be formulated as

$$\begin{aligned}\eta_{t+1} &= r_{t+1} - E_t(r_{t+1}) = \\ &= E_{t+1} \left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} \right] - E_t \left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} \right] - \left(E_{t+1} \left[\sum_{j=1}^{\infty} \rho^j r_{t+1+j} \right] - E_t \left[\sum_{j=1}^{\infty} \rho^j r_{t+1+j} \right] \right)\end{aligned}$$

- Hence, the unexpected stock return is the sum of two components

$$\eta_{t+1} = \eta_{d,t+1} - \eta_{r,t+1}$$

- Or unexpected stock returns must be associated with changes in expectations of future dividends or real returns
- To illustrate let us assume that

$$E_t[r_{t+1}] = r + x_t$$

$$x_{t+1} = \phi x_t + \varepsilon_{t+1}$$

The CS approximation

➤ Then

$$\begin{aligned} p_{rt} &\equiv E_t \left[\sum_{j=0}^{\infty} \rho^j r_{t+1+j} \right] \\ &= \frac{r}{1-\rho} + \frac{x_t}{1-\rho\phi} \\ &\approx \frac{r}{1-\rho} + \frac{x_t}{1-\phi} \end{aligned}$$

- So if expected return is persistent, a 1% increase in expected return has a greater effect on the stock price.
- Notice:

$$\begin{aligned} E_{t+1} \left[\sum_{j=1}^{\infty} \rho^j r_{t+1+j} \right] &= \frac{\rho r}{1-\rho} + \frac{\rho x_t}{1-\rho\phi} \\ E_t \left[\sum_{j=1}^{\infty} \rho^j r_{t+1+j} \right] &= \frac{\rho r}{1-\rho} + \frac{\rho\phi x_t}{1-\rho\phi} \end{aligned}$$

➤ And we get,

$$\eta_{r,t+1} = \frac{\rho \varepsilon_{t+1}}{1-\rho\phi}$$

Campbell and Voulteenaho (CV)

- Recall, the unexpected return is $\eta_{t+1} = \eta_{d,t+1} - \eta_{r,t+1}$, which based on CV can be rewritten as $\eta_{t+1} = N_{CF,t+1} - N_{DR,t+1}$
- That is to say that unexpected return is attributable to news about future cash flows $N_{CF,t+1}$, represented through stream of dividends or consumption, as well as news about future discount rates $N_{DR,t+1}$.
- An increase in expected future CFs is associated with a capital gain today while an increase in discount rates is associated with a capital loss today.
- Such return components can also be interpreted approximately as permanent and transitory shocks to wealth.
- In particular, returns generated by CF news are never reversed, whereas returns generated by DR news are offset by lower future returns.
- From this perspective, conservative long-term investors are more averse to CF risk than to DR risk.

Vector Auto regression Specification

- Consider now the VAR

$$Z_{t+1} = AZ_t + u_{t+1}$$

where

$$Z_t = \begin{pmatrix} r_t \\ d_t - p_t \\ r_{bt} \end{pmatrix}$$

and u_{t+1} is a vector of IID shocks.

- By the CS decomposition we have

$$\begin{aligned} \eta_{t+1} &= r_{t+1} - E_t(r_{t+1}) = \\ &= (E_{t+1} - E_t) \left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} \right] - (E_{t+1} - E_t) \left[\sum_{j=1}^{\infty} \rho^j r_{t+1+j} \right] \end{aligned}$$

- Notice:

$$E_t(Z_{t+1+j}) = A^{j+1}Z_t$$

Vector Auto regression Specification

➤ Then,

$$E_t(r_{t+1+j}) = e_1' A^{j+1} Z_t$$

$$E_{t+1}(r_{t+1+j}) = e_1' A^j Z_{t+1}$$

$$\text{where } e_1 = [1, 0, 0]'$$

➤ So:

$$\begin{aligned}(E_{t+1} - E_t) r_{t+1+j} &= e_1' (A^j Z_{t+1} - A^{j+1} Z_t) \\ &= e_1' (A^j (A Z_t + u_{t+1}) - A^{j+1} Z_t) \\ &= e_1' A^j u_{t+1}\end{aligned}$$

➤ The discount rate news is given by

$$\begin{aligned}N_{DR,t+1} &= (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \\ &= e_1' \sum_{j=1}^{\infty} \rho^j A^j u_{t+1} \\ &= e_1' \rho A (I - \rho A)^{-1} u_{t+1} \\ &= e_1' \lambda u_{t+1}\end{aligned}$$

Vector Auto regression Specification

- The vector product $e_1' \lambda$ captures the long-run significance of each individual VAR shock to DR expectations.
- The greater the absolute value of the variable's coefficient in the return prediction equation (the top row of A) the greater the weight the variable receives in the discount rate news formula.
- More persistent variables should also receive more weight as captured by $(I - \rho A)^{-1}$
- Similarly, the CF news is given by

$$\begin{aligned} N_{CF,t+1} &= (E_{t+1} - E_t) \left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} \right] \\ &= r_{t+1} - E_t(r_{t+1}) + e_1' \rho A (I - \rho A)^{-1} u_{t+1} \\ &= e_1' Z_{t+1} - e_1' A Z_t + e_1' \rho A (I - \rho A)^{-1} u_{t+1} \\ &= e_1' u_{t+1} + e_1' \rho A (I - \rho A)^{-1} u_{t+1} \\ &= e_1' [I + \rho A (I - \rho A)^{-1}] u_{t+1} \\ &= e_1' [I + \lambda] u_{t+1} \end{aligned}$$

Vector Auto regression Specification

- Notice that the stock return variance can be expressed as

$$V(1,1) = e_1' V e_1 \text{ where } V = \text{cov}(Z_t, Z_t')$$

- The variance component due to DR news

$$\text{Var}(N_{DR,t+1}) = e_1' \rho A (I - \rho A)^{-1} V [(I - \rho A)^{-1}]' (\rho A)' e_1$$

- The component of variance due to CF news

$$\text{Var}(N_{CF,t+1}) = e_1' [I + \rho A (I - \rho A)^{-1}] V [I + \rho A (I - \rho A)^{-1}]' e_1$$

- Moreover, since the market return contains two components, both of which are not highly correlated, then different types of stocks may have different betas with these components.
- In particular the cash flow beta

$$\beta_{i,CF} \equiv \frac{\text{Cov}(r_{it}, N_{CF,t})}{\text{Var}[r_{M,t}^e - E_{t-1}(r_{M,t}^e)]}$$

- Likewise, the discount rate beta is

$$\beta_{i,DR} \equiv \frac{\text{Cov}(r_{it}, N_{DR,t})}{\text{Var}[r_{M,t}^e - E_{t-1}(r_{M,t}^e)]}$$

Vector Auto regression Specification

- And we have

$$\beta_{i,M} = \beta_{i,CF} + \beta_{i,DR}$$

- Both betas could be represented as

$$\beta_{i,CF} \equiv (e'_1 + e'_1 \lambda) \frac{\text{Cov}(r_{it}, u_t)}{\text{Var}[r_{M,t}^e - E_{t-1}(r_{M,t}^e)]}$$

$$\beta_{i,DR} \equiv -e'_1 \lambda \frac{\text{Cov}(r_{it}, u_t)}{\text{Var}[r_{M,t}^e - E_{t-1}(r_{M,t}^e)]}$$

where $\text{Cov}(r_{it}, u_t)$ is a vector of covariance between firm i 's stock return and the innovations in the state variables

- Campbell (1993) derives an approximate discrete-time version of the ICAPM.
- Based on this ICAPM, CV show that

$$E_t[r_{i,t+1}] - r_{f,t+1} = \gamma \sigma_M^2 \beta_{i,CF} + \sigma_M^2 \beta_{i,DR}$$

- Notice that based on empirical studies γ the relative risk aversion parameter is well beyond one
- Thus, higher risk premium is required on the CF beta than on the DR beta

Vector Auto regression Specification

- For this reason, CV call the CF beta the “bad beta” and the DR beta the “good beta”
- The intuition is that a higher unexpected market equity return, due to DR news, implies lower future growth opportunities. Thus a positive DR beta, stock that pays more when growth opportunity shrinks, is welcome by investors.
- Typically in empirical studies, DR news is directly modeled while the CF news is calculated as the residual.
- Chen and Zhao (2006) show that this approach has a serious limitation because the DR news cannot be measured accurately enough.
- To illustrate, this point Chen and Zhao (2006) apply the approach to Treasury bonds that should have zero CF betas
- They find that the variance of “CF news” is larger than that of “DR news”
- To see why CF news do not play a role we follow Campbell and Ammer (1993)
- Let $P_{n,t}$ be the price of an n-periods nominal bond at time t , and $p_{n,t} = \ln P_{n,t}$
- Then the holding period return from time t to time $t + 1$

$$r_{n,t+1} = p_{n-1,t+1} - p_{n,t}$$

Vector Auto regression Specification

- This equation can be thought as a difference equation in the log bond price
- Iterating forward and substitute $P_{0,t+n} = 1$ or $p_{0,t+n} = 0$, we obtain

$$p_{n,t} = -[r_{n,t+1} + r_{n-1,t+2} + \dots + r_{1,t+n}] = -\sum_{i=0}^{n-1} r_{n-i,t+1+i}$$

- This equation holds ex post, but it also holds ex ante
- Taking expectation conditional on information at time t we get

$$p_{n,t} = -E_t \sum_{i=0}^{n-1} r_{n-i,t+1+i}$$

- In the end we get

$$r_{n,t+1} - E_t(r_{n,t+1}) = -(E_{t+1} - E_t) \sum_{i=1}^{n-1} r_{n-i,t+1+i}$$

- This equation express the well-known fact that unexpected positive nominal returns today are always offset by decrease in expected future nominal returns.

Campbell(1993)

- Campbell (1993) implements similar concepts for representative agent dynamic budget constraint

$$W_{t+1} = R_{m,t+1}(W_t - C_t)$$

where $R_{m,t+1}$ is the gross simple return on wealth invested from period t to period $t + 1$.

- Then,

$$W_t = C_t + \frac{W_{t+1}}{R_{m,t+1}}$$

- Iterating plus transversality condition yields

$$W_t = C_t + \sum_{i=1}^{\infty} \frac{C_{t+i}}{(\prod_{j=1}^i R_{m,t+j})}$$

- Notice that, the budget constraint is highly nonlinear
- The log linear approximation is implemented on

$$\frac{W_{t+1}}{W_t} = R_{m,t+1} \left(1 - \frac{C_t}{W_t} \right)$$

Campbell(1993)

- ▶ Taking logs

$$\Delta w_{t+1} = r_{m,t+1} + \log(1 - \exp(c_t - w_t))$$

- ▶ First-order Taylor approximation yields

$$\log[1 - \exp(x_t)] \approx \log[1 - \exp(\bar{x})] - \frac{\exp(\bar{x})}{1 - \exp(\bar{x})} (x_t - \bar{x})$$

- ▶ If x_t is constant ($\frac{c_t}{w_t}$ is constant) then the approximation holds exactly
- ▶ The log-linear approximation is given by

$$\Delta w_{t+1} \approx r_{m,t+1} + k + \left(1 - \frac{1}{\rho}\right) (c_t - w_t)$$

where

$$\rho = 1 - \exp(\bar{x}) = \frac{W - C}{W}$$

Campbell(1993)

➤ Also

$$\Delta w_{t+1} = \Delta c_{t+1} + (c_t - w_t) - (c_{t+1} - w_{t+1})$$

➤ Then

$$\Delta c_{t+1} + (c_t - w_t) - (c_{t+1} - w_{t+1}) \approx r_{m,t+1} + k + \left(1 - \frac{1}{\rho}\right)(c_t - w_t)$$

$$\frac{1}{\rho}(c_t - w_t) = (c_{t+1} - w_{t+1}) + (r_{m,t+1} - \Delta c_{t+1}) + k$$

➤ Through iterating we get

$$c_t - w_t = \sum_{j=1}^{\infty} \rho^j (r_{m,t+j} - \Delta c_{t+j}) + \frac{\rho k}{1 - \rho}$$

➤ This is the log-linear version of

$$W_t = C_t + \sum_{i=1}^{\infty} \frac{C_{t+i}}{(\prod_{j=1}^i R_{m,t+j})}$$

➤ Taking expectations, we get

$$c_t - w_t = E_t \sum_{j=1}^{\infty} \rho^j (r_{m,t+j} - \Delta c_{t+j}) + \frac{\rho k}{1 - \rho}$$

Campbell(1993)

- Suggesting that revision in the consumption to wealth ratio predicts either return or consumption growth or both
- Rewriting the present-value relation we get

$$c_t - E_t(c_{t+1}) = \Delta E_{t+1} \sum_{j=0}^{\infty} \rho^j r_{m,t+1+j} - \Delta E_{t+1} \sum_{j=1}^{\infty} \rho^j \Delta c_{t+1+j}$$

- That is to say that upward surprise in consumption today must correspond to either upward revision in expected return on wealth today or downward revision in expected future consumption growth
- This is pretty much similar to what we get about the prediction where here consumption plays the role of dividends and aggregate wealth is associated with an asset whose dividends are equal to consumption

Binsbergen and Koijen (2010)

- Binsbergen and Koijen (2010) propose a latent variable approach within a present-value model to estimate the expected returns and expected dividend growth rates of the aggregate stock market.
- This approach aggregates information contained in the history of price-dividend ratios and dividend growth rates to predict future returns and dividend growth rates.
- They find that returns and dividend growth rates are predictable with R -Square values ranging from 8.2% to 8.9% for returns and 13.9% to 31.6% for dividend growth rates.
- Both expected returns and expected dividend growth rates have a persistent component, but expected returns are more persistent than expected dividend growth rates.
- In particular expected returns and expected dividend growth rates are modeled as AR(1) process:

$$\mu_{t+1} = \delta_0 + \delta_1(\mu_t - \delta_0) + \varepsilon_{t+1}^{\mu}$$

$$g_{t+1} = \gamma_0 + \gamma_1(g_t - \gamma_0) + \varepsilon_{t+1}^g$$

where

$$\mu_t \equiv E_t[r_{t+1}]$$

$$g_t \equiv E_t[\Delta d_{t+1}]$$

Binsbergen and Koijen (2010)

- Then

$$\Delta d_{t+1} = g_t + \varepsilon_{t+1}^d$$

- Plugging these equations into

$$dp_t = d_t - p_t = \frac{k}{1-\rho} + E_t \left[\sum_{j=0}^{\infty} [\Delta d_{t+1+j} - r_{t+1+j}] \right]$$

- They get the following expression for pd_t

$$pd_t = \frac{k}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} - \frac{\mu_t - \delta_0}{1-\rho\delta_1} + \frac{g_t - \gamma_0}{1-\rho\gamma_1}$$

- So altogether we can write the implied dynamics of the dividend-to-price ratio as:

$$pd_t = A - B_1(\mu_t - \delta_0) + B_2(g_t - \gamma_0)$$

where

$$A = \frac{k}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho}$$
$$B_1 = \frac{1}{1-\rho\delta_1}$$
$$B_2 = \frac{1}{1-\rho\gamma_1}$$

Binsbergen and Koijen (2010)

- Suggesting that the log dividend-price ratio is linear in μ_t and g_t .
- Further, the loadings of pd_t on μ_t and g_t depend on the relative persistence of these two variables as reflected through δ_1 and γ_1
- Their model has two transition equations,

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \varepsilon_{t+1}^g$$

$$\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \varepsilon_{t+1}^\mu$$

and two measurement equations,

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon_{t+1}^d$$

$$pd_t = A - B_1 \hat{\mu}_t + B_2 \hat{g}_t$$

where

$$\hat{\mu}_t = \mu_t - \delta_0$$

$$\hat{g}_t = g_t - \gamma_0$$

- It may be surprising that there is no measurement equation for returns.
- However, the measurement equation for dividend growth rates and the price-dividend ratio together imply the measurement equation for returns.

Avramov, Cederburg, and Lucivjanska (2017)

- ▶ They study the present value relation from a long horizon perspective.

- ▶ They show that cumulative log return can be approximated as

$$r_{T,T+k} = (p_{T+k} - d_{T+k}) - (p_T - d_T) + \sum_{i=1}^k (1 - \rho) (d_{T+i} - p_{T+i}) + \Delta d_{T,T+k} + kq$$

- ▶ Returns have three components that constitute the uncertainty

- ▶ The difference between beginning and terminal *price-to-dividend* ratio $(p_{T+k} - d_{T+k}) - (p_T - d_T)$

- ▶ The $\sum_{i=1}^k (1 - \rho) (d_{T+i} - p_{T+i})$ term captures the cumulative effect of dividend income during the holding period

- ▶ The $\Delta d_{T,T+k}$ term is the cumulative dividend growth that is realized over the horizon.

- ▶ Notice that the present value relation implies restriction on the predictive framework

$$\begin{bmatrix} r_{t+1} \\ \Delta d_{t+1} \\ p_{t+1} - d_{t+1} \\ z_{t+1} \end{bmatrix} = a + B \begin{bmatrix} p_t - d_t \\ z_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{r,t+1} \\ \varepsilon_{d,t+1} \\ \varepsilon_{pd,t+1} \\ \varepsilon_{z,t+1} \end{bmatrix}$$

- In particular

$$\varepsilon_{r,t+1} = \rho \varepsilon_{pd,t+1} + \varepsilon_{d,t+1}$$

- The VAR thus must be estimated with observation equations for only two of the r_{t+1} , Δd_{t+1} , and $p_{t+1} - d_{t+1}$ variables to ensure that the covariance matrix is nonsingular.

Pastor, Sinha, and Swaminathan (2008)

- **The Implied Cost of Capital (ICC)** is the discount rate that equates the present value of expected future dividends to the current stock price
- One common approach is to define the ICC as the value of r_e that solves

$$P_t = \sum_{k=1}^{\infty} \frac{E_t(D_{t+k})}{(1 + r_e)^k}$$

- Recall that CS develop a useful approximation for the stock price which expresses the log price $p_t = \log(P_t)$ as

$$p_t = \frac{k}{1 - \rho} + (1 - \rho) \sum_{j=0}^{\infty} \rho^j E_t(d_{t+1+j}) - \sum_{j=0}^{\infty} \rho^j E_t(r_{t+1+j})$$

Pastor, Sinha, and Swaminathan (2008)

- In this framework, it is natural to define the ICC as the value of $r_{e,t}$ that solves

$$p_t = \frac{k}{1-\rho} + (1-\rho) \sum_{j=0}^{\infty} \rho^j E_t(d_{t+1+j}) - r_{e,t} \sum_{j=0}^{\infty} \rho^j$$

- To provide some insight into the ICC, it is convenient to assume that log dividend growth

$g_{t+1} \equiv d_{t+1} - d_t$ follows a stationary AR(1) process

$$g_{t+1} = \gamma + \delta g_t + v_{t+1}, \quad 0 < \delta < 1, \quad v_{t+1} \sim N(0, \sigma_v^2)$$

- Given these dynamics of g_t

$$\begin{aligned} \sum_{j=0}^{\infty} \rho^j E_t(d_{t+1+j}) = \\ \frac{d_t}{1-\rho} + \frac{\gamma}{(1-\delta)(1-\rho)^2} - \frac{\gamma\delta}{(1-\delta)(1-\rho)(1-\rho\delta)} + \frac{\delta g_t}{(1-\rho)(1-\rho\delta)} \end{aligned}$$

- Substituting this equation into p_t we obtain

$$p_t = \frac{k}{1-\rho} + d_t + \frac{\gamma}{(1-\delta)(1-\rho)} - \frac{\gamma\delta}{(1-\delta)(1-\rho\delta)} + g_t \frac{\delta}{(1-\rho\delta)} - \frac{r_{e,t}}{(1-\rho)}$$

Pastor, Sinha and, Swaminathan (2008)

which can be rearranged into

$$r_{e,t} = k + \frac{\gamma}{(1-\delta)} + (d_t - p_t)(1 - \rho) + \left(g_t - \frac{\gamma}{1-\delta}\right) \frac{\delta(1-\rho)}{1-\rho\delta}$$

- The ICC, $r_{e,t}$, is a simple linear function of the log dividend-price ratio, $d_t - p_t$, and log dividend growth, g_t .
- Further insight into the ICC can be obtained by assuming that the conditional expected return, $\mu_t \equiv E_t(r_{t+1})$ also follows a stationary AR(1) process

$$\mu_{t+1} = \alpha + \beta\mu_t + u_{t+1}, \quad 0 < \beta < 1, \quad u_{t+1} \sim N(0, \sigma_u^2)$$
$$\sum_{j=0}^{\infty} \rho^j E_t(r_{t+1+j}) = \frac{\alpha}{(1-\beta)(1-\rho)} + \left(\mu_t - \frac{\alpha}{1-\beta}\right) \frac{1}{1-\rho\beta}$$

- Plugging $\sum_{j=0}^{\infty} \rho^j E_t(d_{t+1+j})$ and $\sum_{j=0}^{\infty} \rho^j E_t(r_{t+1+j})$ into p_t , we obtain

$$p_t = \frac{k}{1-\rho} + \frac{\gamma}{(1-\delta)(1-\rho)} - \frac{\alpha}{(1-\beta)(1-\rho)}$$
$$+ d_t + \left(g_t - \frac{\gamma}{1-\delta}\right) \frac{\delta}{(1-\rho\delta)} - \left(\mu_t - \frac{\alpha}{1-\beta}\right) \frac{1}{1-\rho\beta}$$

Pastor, Sinha and, Swaminathan (2008)

- The log stock price p_t is a simple function of d_t , g_t and μ_t .
- The stock price increases with dividends d_t and dividend growth g_t , and it decreases with expected return μ_t .
- Note that p_t depends on the deviations of μ_t and g_t from their unconditional means of $\frac{\alpha}{1-\beta}$ and $\frac{\gamma}{1-\delta}$ respectively
- Comparing the equations of p_t and $r_{e,t}$ we have

$$r_{e,t} = \frac{\alpha}{1-\beta} + \left(\mu_t - \frac{\alpha}{1-\beta} \right) \frac{1-\rho}{1-\rho\beta}$$

which implies that $r_{e,t}$ and μ_t are perfectly correlated.

- Thus, the ICC is a perfect proxy for the conditional expected return in the time series in an AR(1) framework
- They also consider a modified version of the ICC,

$$r_{e2,t} = k + \frac{\gamma}{1-\delta} + (d_t - p_t)(1-\rho)$$

Pastor, Sinha and, Swaminathan (2008)

- This expression is obtained from the equation of $r_{e,t}$ by setting g_t equal to its unconditional mean $\frac{\gamma}{1-\delta}$.
- This definition of $r_{e2,t}$ captures the idea that our information about dividend growth is often limited in practice.
- Note that $r_{e2,t}$ is perfectly correlated with the dividend-to-price ratio, which is commonly used to proxy for expected return.
- Since dividends tend to vary less than prices, the time variation in $r_{e2,t}$ is driven mostly by the time variation in p_t .