

Topics in Asset Pricing

Lecture Notes

Doron Avramov

Professor of Finance

IDC, Israel

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 developing advanced mathematical, statistical, and econometric techniques to understand 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.
- This set of notes compliments notes on *Machine Learning Methods in the Cross Section of Asset Returns*
https://faculty.idc.ac.il/davramov/Machine_learning.pdf
- Required: prior exposure to matrix algebra, distribution theory, Ordinary Least Squares, as well as skills in computer programming.
- MATLAB and Python are the most recommended for this course.

Topics to be covered

- From CAPM to market anomalies
- Credit risk implications for the cross section of asset returns
- Rational versus behavioural attributes of stylized cross-sectional effects
- Are market anomalies pervasive?
- Conditional CAPM
- Conditional versus unconditional portfolio efficiency
- Multi-factor models
- Interpreting factor models
- Panel regressions with fixed effects and their association with market-timing and cross-section investment strategies
- Machine learning methods: Lasso, Ridge, elastic net, group Lasso, Neural Network, Random Forest, and adversarial GMM
- Stock return predictability by macro variables
- Finite sample bias in predictive regressions
- Lower bound on the equity premium
- The Campbell-Shiller log linearization
- 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

Topics to be covered

- Long-run risk
- Habit formation
- Prospect theory
- Time-series asset pricing tests
- Cross-section asset pricing tests
- 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 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, 2019) 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.
- Also, value weighting anomaly payoffs or excluding micro-cap stocks attenuate the strength of many prominent anomalies.
- 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 believe in the negative credit risk return relation, yet the contradicting findings asks 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.
- Remarkably, regressing gross momentum payoffs on the Fama-French factors tends to strengthen, rather than discount, momentum profitability.
- This is because momentum loads negatively on 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.
- Building on Avramov et al (2017), Eshani and Linnainmaa (2019) show that stock-level momentum emerges from momentum in factor returns.
- Thus, momentum is not a distinct risk factor; rather, it aggregates the autocorrelations found in all other factors.

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 (2019).
- ▶ They document overpricing in stocks and 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):

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

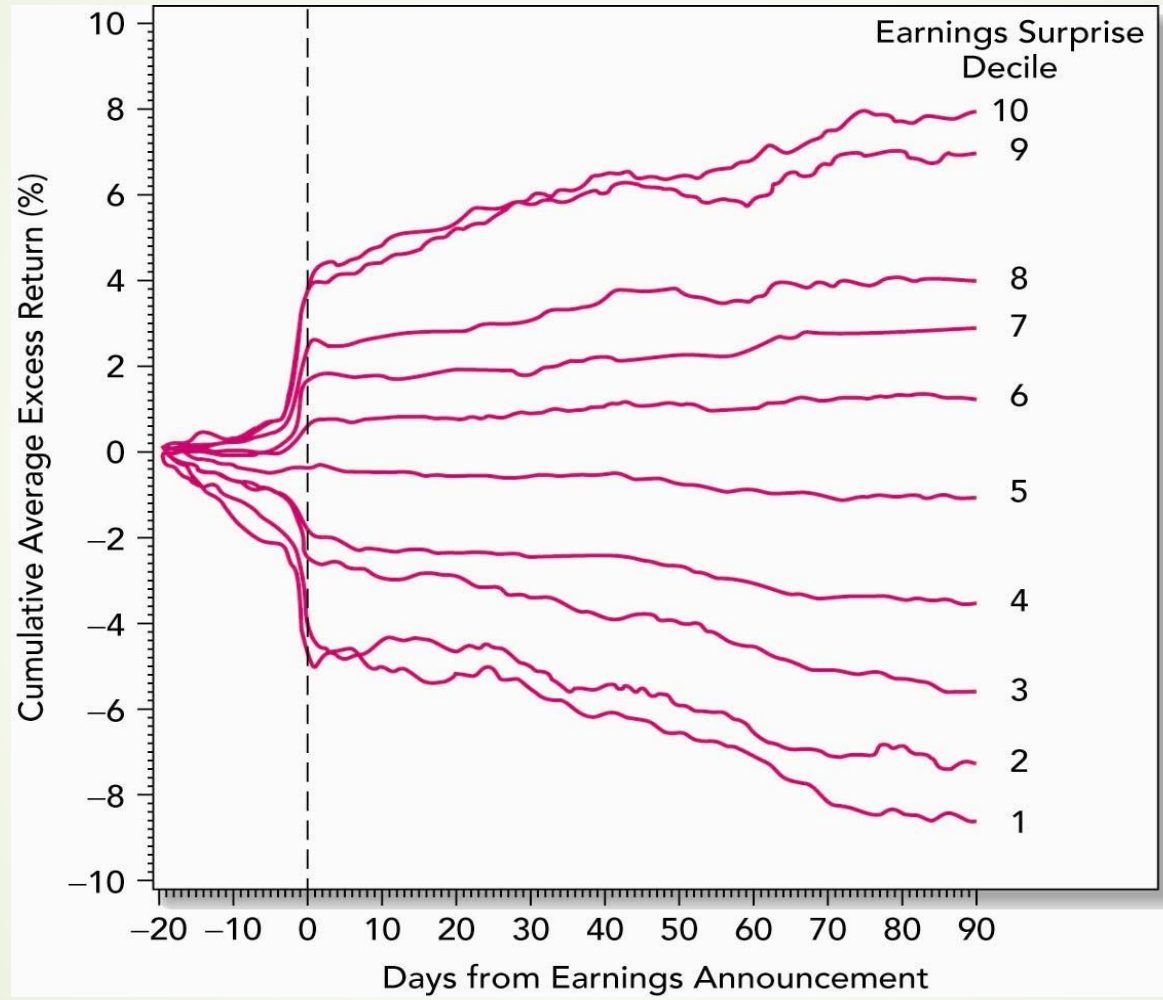
Earnings Momentum (see also next page)

- 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?

- The evidence tilts towards the NO answer. Albeit, the tilt is not decisive.
- 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, 2019) 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 (2019) integrate 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 (2019) 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 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 (2019) 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).

Sticky expectations and market anomalies

- 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, Kaplaniski, and Subrahmanyam (2019).
- While the former study points at anchoring as a potential rationale for mispricing, the sticky expectations (SE) concept somehow formalizes the same notion.
- The SE concept has been developed and studied by Mankiw and Reis (2002), Reis (2006), and Coibion and Gorodnichenko (2012, 2015).
- Bouchaus, Kruger, Landier, and Thesmar (2019) propose SE to explain the profitability anomaly along with price momentum and earnings momentum.
- The idea is straightforward.
- In particular, expectations about an economic quantity (π_{t+h}) are updated using the process

$$F_t \pi_{t+h} = (1 - \lambda) E_t \pi_{t+h} + \lambda F_{t-1} \pi_{t+h}$$

Sticky expectations and market anomalies

- The term $E_t \pi_{t+h}$ denotes the rational expectation of π_{t+h} conditional on information available at date t .
- The coefficient λ indicates the extent of expectation stickiness.
- When $\lambda = 0$, expectations are perfectly rational.
- Otherwise, new information is insufficiently accounted for in establishing forecasts.
- This framework accommodates patterns of both under-reaction ($0 < \lambda < 1$) and overreaction ($\lambda < 0$).
- As noted by Coibion and Gorodnichenko (2012, 2015), this structure gives rise to straightforward testable predictions that are independent of the process underlying π_{t+h} .
- This structure also provides a direct measure of the level of stickiness.

Sticky expectations and market anomalies

- In the first place, forecast errors should be predicted by past revisions:

$$E_t(\pi_{t+1} - F_t \pi_{t+1}) = \frac{\lambda}{1 - \lambda} (F_t \pi_{t+1} - F_{t-1} \pi_{t+1})$$

- Second, revisions are auto-correlated over time:

$$E_{t-1}(F_t \pi_{t+1} - F_{t-1} \pi_{t+1}) = \lambda (F_{t-1} \pi_{t+1} - F_{t-2} \pi_{t+1})$$

- These two relations can be readily tested on expectations data (including inflation, profitability, interest rate, future price) without further assumptions about the data-generating process of π .
- The intuition behind the first testable restriction is that forecast revisions contain some element of new information that is only partially incorporated into expectations.
- The second prediction pertains to the dynamics of forecast revisions.
- When expectations are sticky, information is slowly incorporated into forecasts, so that positive news generates positive forecast revisions over several periods.
- This generates momentum in forecasts.

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} is equal to

$$\text{cov}(R_{it}, R_{Mt}) = \text{cov}(\beta R_{Mt} + \eta_t R_{Mt} + \varepsilon_t, R_{Mt})$$

$$= \beta \sigma_M^2 + \text{cov}(\eta_t R_{Mt} + \varepsilon_t, R_{Mt})$$

$$= \beta \sigma_M^2 + E(\eta_t R_{Mt}^2) - E(\eta_t R_{Mt})E(R_{Mt})$$

$$= \beta \sigma_M^2 + \text{cov}(\eta_t, R_{Mt}^2) - \gamma \text{cov}(\eta_t, R_{Mt})$$

$$= \beta \sigma_M^2 + \text{cov}(\eta_t, \sigma_t^2) + \text{cov}(\eta_t, \gamma_t^2) - \gamma \text{cov}(\eta_t, \gamma_t)$$

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

Conditional CAPM

➤ Then,

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

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 when the conditional CAPM holds exactly we should expect to find deviations from the unconditional CAPM if **any** of the three covariance terms is nonzero.

➤ But if the conditional CAPM holds, α^u should be relatively small, at odds with market anomalies.

D. Perhaps Multifactor Models?

- The poor 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.
- However, 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, 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

- If you extract $K = 1, \dots, 15$ factors, the maximum squared SR based on the extracted factors rises with K .
- Out-of-sample --- things look 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.

Pricing Kernel with time-varying Parameters

- 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 $H \times 1$ vector.
- Plugging b_{t-1} into the conditional version of the pricing kernel yields

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

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 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 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.
- In response to Cochrane’s challenge and his call for different methods, there has been emerging literature that applies machine learning techniques in asset pricing.
- Before we delve into heavy duty machine learning methods, there is a short chapter, coming up, that tells you how to distinguish between time-series and cross-sectional effects in the relation between future stock return and the current value of predictive characteristic.

Panel regression slope coefficients and their association with trading strategies

Panel Regressions

- ▶ We study panel regressions with versus without fixed effects and their close association with average payoffs to time-series and cross-sectional strategies.
- ▶ The time-series dimension consists of T months altogether.
- ▶ The cross-section dimension consists of N firms altogether.
- ▶ We consider a panel that is not essentially balanced.
- ▶ In each month, there are $N_t \leq N$ firms, while each firm records $T_i \leq T$ monthly observations of returns and predictive characteristics, such as past return, volatility, investment, credit risk, and profitability.
- ▶ Let r_{it} denote return on stock i at month t and let z_{it} represent a single characteristic for stock i at month t .
- ▶ Consider the regression of future return on the current value of characteristic with stock fixed effects:

$$r_{it+1} = a_i + b_{TS}z_{it} + \varepsilon_{it+1}$$

- ▶ To increase the power of the inference the slope is assumed constant across stocks and through time.
- ▶ The slope has the TS subscript because the regression assesses only time-series predictability.
- ▶ That is, accounting for stock fixed-effects reflects only within-stock time-series variation in z_{it} .
- ▶ The slope can be estimated from a no-intercept regression of return on demeaned characteristic, where demeaning is along the time-series dimension.

Panel Regressions with stock fixed effects

- The estimated slope is thus given by

$$\hat{b}_{TS} = \frac{\sum_{i=1}^N T_i \hat{\sigma}_{zr,i}}{\sum_{n=1}^N T_n \hat{\sigma}_{z,n}^2}$$

where $\hat{\sigma}_{z,i}^2$ is the time-series variance of z_{it} and $\hat{\sigma}_{zr,i}$ is the time-series covariance between r_{it+1} and z_{it} .

- To understand the properties of the slope estimate and its significance, we can express the slope as

$$\hat{b}_{TS} = b_{TS} + \frac{\sum_i \sum_t (z_{it} - \bar{z}_i) \varepsilon_{it+1}}{\sum_{n=1}^N T_n \hat{\sigma}_{z,n}^2}$$

where \bar{z}_i is the firm i time-series mean of the predictive characteristic.

- Clustering is essential for estimating the standard error.
- Pastor, Stambaugh, and Taylor (2017) show that this slope can also be represented as

$$\hat{b}_{TS} = \sum_{i=1}^N w_i \hat{b}_i$$

where $w_i = \frac{T_i \hat{\sigma}_{z,i}^2}{\sum_{n=1}^N T_n \hat{\sigma}_{z,n}^2}$ and \hat{b}_i is the estimated slope coefficient in stock-level time-series regressions:

$$r_{it+1} = a_i + b_i z_{it} + e_{it+1}$$

- That is, the panel regression slope estimate is a value weighted average of estimated slopes from individual regressions.
- This weighting scheme places larger weights on time-series slopes of stocks with more observations as well as stocks whose predictive characteristic fluctuates more over time.

Time-series based investment strategies

- Pastor, Stambaugh, and Taylor (2017) also show that the panel regression coefficient is related to investment strategy payoff.

- To illustrate, consider a long-short trading strategy from a time-series perspective:

Long A: z_{it} in stock i at month t

Short B: \bar{z}_i in stock i at month t

- Strategy A is a market-timing strategy with time-varying weights.

- Strategy B is a static constant-weight strategy.

- Denote the total payoff for the long-short strategy by $\varphi_{TS,i}$

- It follows that:

$$\varphi_{TS,i} = \sum_{t=1}^{T_i} (z_{it} - \bar{z}_i) r_{it+1} = T_i \hat{\sigma}_{rz,i}$$

- That is, the payoff is proportional to the time-series covariance between future return and the current value of the predictive characteristic.

Panel Regressions with stock fixed effects

- Then, the total payoff that aggregates across all stocks is given by

$$\varphi_{TS} = \sum_{i=1}^N \varphi_{TS,i} = \sum_{i=1}^N T_i \hat{\sigma}_{rz,i}$$

- The total payoff can thus be represented as a function of the panel regression slope

$$\varphi_{TS} = \hat{b}_{TS} \left(\sum_{n=1}^N T_n \hat{\sigma}_{Z,n}^2 \right)$$

- Hence, the total (and average) payoff is proportional to the slope estimate in a panel regression with stock-fixed effects.
- We could make it equality $\varphi_{TS} = b_{TS}$ by scaling the investment.
- That is, rather than dollar long dollar short, invest $1/(\sum_{n=1}^N T_n \hat{\sigma}_{Z,n}^2)$ in both the long and the short.

Panel Regressions with month fixed effects

- Consider now a panel regression with month-fixed effects

$$r_{it+1} = a_t + b_{CS}z_{it} + v_{it+1}$$

- We use the CS subscript to reflect the notion that month fixed effects correspond to a cross-sectional analysis.
- Accounting for month fixed-effects reflects only cross-section variation in the predictive characteristic.
- The slope can be estimated through a no-intercept panel regression of future stock return on the demeaned characteristic, where demeaning is along the cross-section dimension.
- The slope of such demeaned regression is readily estimated as

$$\hat{b}_{CS} = \frac{\sum_{t=1}^T N_t \hat{\sigma}_{zr,t}}{\sum_{t=1}^T N_t \hat{\sigma}_{z,t}^2}$$

where $\hat{\sigma}_{z,t}^2$ is the cross-sectional variance of the characteristic in month t and $\hat{\sigma}_{zr,t}$ is the cross-sectional covariance between r_{it+1} and z_{it} .

- To understand the properties of the slope estimate and its significance, we can express the slope as

$$\hat{b}_{CS} = b_{CS} + \frac{\sum_i \sum_t (z_{it} - \bar{z}_t) \varepsilon_{it+1}}{\sum_{t=1}^T N_t \hat{\sigma}_{z,t}^2}$$

where \bar{z}_t is the time t cross-sectional mean of the predictive characteristic.

Panel Regressions with month fixed effects

➤ The slope can be estimated through $\hat{b}_{CS} = \sum_{t=1}^T w_t \hat{b}_t$ where $w_t = \frac{N_t \hat{\sigma}_{Z,t}^2}{\sum_{s=1}^T N_s \hat{\sigma}_{Z,s}^2}$.

➤ Thus, the slope is a value weighted average of slopes from monthly cross-section regressions:

$$r_{it+1} = a_t + b_t z_{it} + \eta_{it+1}$$

- Larger weights are placed on cross-sectional estimates from periods with more stocks and periods in which the independent variable exhibits more cross-sectional variation.
- Notice that the same slope obtains also through regressing demeaned return on demeaned characteristic (no month fixed effects), where demeaning is along the cross-section direction.
- With stock-fixed effects the weights depend on the number of time-series observations per stock, while with month fixed-effects, the weights depend on the number of stocks per month.
- Also, with stock-fixed effects the weights depend on the time-series variation of the independent variable, while with month fixed-effects, they depend on the cross-sectional variation.

Cross Section based investment strategies

- Let us now consider a long-short trading strategy from a cross-sectional perspective:

Long A: z_{it} in stock i at month t

Short B: \bar{z}_t in stock i at month t

where \bar{z}_t is the time t cross-sectional mean of z_{it}

$$\bar{z}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} z_{it}$$

- The total payoff for this long-short strategy is

$$\varphi_{CS,t} = \sum_{i=1}^N (z_{it} - \bar{z}_t) r_{it+1} = N_t \hat{\sigma}_{rz,t}$$

where $\hat{\sigma}_{rz,t}$ is the estimate of month t cross-sectional covariance between r_{it+1} and z_{it} .

Panel Regressions with month Fixed effects

- Aggregating through all months, the total payoff is given by

$$\varphi_{CS} = \sum_{t=1}^T \varphi_{CS,t} = \sum_{t=1}^T N_t \hat{\sigma}_{rz,t}$$

- The total payoff is thus proportional to the panel regression slope with month fixed effects.
- In particular

$$\varphi_{CS} = \hat{b}_{CS} \left(\sum_{t=1}^T N_t \hat{\sigma}_{rz,t} \right)$$

- Notice also that \hat{b}_{CS} is related to the commonly used Fama-MacBeth estimator

$$\hat{b}_{FM} = \frac{1}{T} \sum_{t=1}^T b_t$$

- The Fama-MacBeth estimator is a special case if the panel is balanced ($N_t = N$ for all t) and the cross-sectional variance of the dependent variable does not change over time.
- The addition of month fixed effects to stock fixed effects controls for any unobserved variables that change over time but not across stocks, such as macroeconomic variables, regulatory changes, and aggregate trading activity.
- Similarly, the addition of stock fixed effects to month fixed effects controls for any unobserved variables that change across stocks but not across time, such as managerial attributes, corporate governance, etc.

Unconditional Covariance between r_{it+1} and z_{it}

- ▶ While Pastor, Stambaugh, and Taylor (2017) have focused on conditional covariation between future return and current value of characteristic, it is also intriguing to understand the sources of unconditional covariation.
- ▶ The unconditional covariance between future stock return and the current value of firm characteristic is not conditioned on either stock fixed effects or time fixed effects.
- ▶ We show below that the unconditional covariance can also be expressed as a function of payoffs attributable to time-series and cross-sectional strategies.
- ▶ In particular, let $\bar{z} = \frac{1}{\sum_i T_i} \sum_i \sum_t z_{it}$ and $\bar{r} = \frac{1}{\sum_i T_i} \sum_i \sum_t r_{it+1}$.
- ▶ These quantities are the grand means (across months and stocks) of the characteristic and return.
- ▶ The estimated unconditional covariance is then given by

$$\begin{aligned} COV &= \sum_i \sum_t (z_{it} - \bar{z}) (r_{it+1} - \bar{r}) \\ &= \sum_i \sum_t (z_{it} - \bar{z}) r_{it+1} \end{aligned}$$

Unconditional Covariance between r_{it+1} and z_{it}

- As an intermediate stage, let us decompose the payoffs to the time-series and cross-sectional strategies as follows

$$\begin{aligned}\varphi_{TS} &= \sum_i \sum_t (z_{it} - \bar{z}_i) r_{it+1} = \underbrace{\sum_i \sum_t (\bar{z}_t - \bar{z}) r_{it+1}}_A + \underbrace{\sum_i \sum_t [(z_{it} - \bar{z}_t) - (\bar{z}_i - \bar{z})] r_{it+1}}_B \\ \varphi_{CS} &= \sum_i \sum_t (z_{it} - \bar{z}_t) r_{it+1} = \underbrace{\sum_i \sum_t (\bar{z}_i - \bar{z}) r_{it+1}}_C + \underbrace{\sum_i \sum_t [(z_{it} - \bar{z}_t) - (\bar{z}_i - \bar{z})] r_{it+1}}_B\end{aligned}$$

- Notice that the payoffs for the time-series and cross-sectional strategies do share a common component (B) reflecting through-time-across-stock payoff.
- Notice also that both strategies do have unique components A and C reflecting purely through-time and across-stock payoffs, respectively.
- Let us now decompose the difference $(z_{it} - \bar{z})$ such that

$$(z_{it} - \bar{z}) = (\bar{z}_t - \bar{z}) + [(z_{it} - \bar{z}_t) - (\bar{z}_i - \bar{z})] + (\bar{z}_i - \bar{z})$$

Unconditional Covariance between r_{it+1} and z_{it}

- Thus, the unconditional covariance between future stock return and the current value of predictive characteristic is given by

$$COV(r_{it+1}, z_{it}) = A + B + C$$

- In words, the estimated unconditional covariance is equal to the sum of the unique time-series payoff (A), the unique cross-sectional payoff (C), and the common component (B).
- Now, let γ denote the slope in the unconditional (no fixed effects) panel regression.
- The slope is estimated as

$$\hat{\gamma} = \frac{\sum_i \sum_t (z_{it} - \bar{z})(r_{it+1} - \bar{r})}{\sum_i \sum_t (z_{it} - \bar{z})^2}$$

- The slope can then be decomposed into its three components reflecting the contribution of A, B, and C in explaining the total variation (through time and across stocks) in return:

$$\hat{\gamma} = \hat{\gamma}_A + \hat{\gamma}_B + \hat{\gamma}_C.$$

- For instance,

$$\hat{\gamma}_A = \frac{A}{\sum_i \sum_t (z_{it} - \bar{z})^2}$$

- Avramov and Xu (2019) implement such slope decomposition in the context of predicting future currency return by the foreign interest rate for various economies.

Unconditional Covariance between r_{it+1} and z_{it}

- To understand the quantity $\hat{\gamma}_A$, consider the regression of return r_{it+1} on \bar{z}_t .
- The estimated slope in that regression is denoted by $\hat{\delta}_A$.
- Then, it follows that

$$\hat{\gamma}_A = \hat{\delta}_A \frac{VAR(\bar{z}_t)}{VAR(z_{it})}$$

- Similarly, consider the regression of return r_{it+1} on \bar{z}_i .
- The estimated slope in this regression is denoted by $\hat{\delta}_C$.
- Then, it follows that

$$\hat{\gamma}_C = \hat{\delta}_C \frac{VAR(\bar{z}_t)}{VAR(z_{it})}$$

- Notice also that $\hat{\delta}_A = \frac{A}{VAR(\bar{z}_t)}$ and $\hat{\delta}_C = \frac{C}{VAR(\bar{z}_i)}$.
- Significance of $\hat{\gamma}$ is easily inferred from the unconditional regression while significance of $\hat{\gamma}_A$, $\hat{\gamma}_B$, or $\hat{\gamma}_C$ or significance of ratios such as A/C can be assessed through Jackknife or Bootstrap (see Avramov and Xu (2019)).
- The ratio A/C reflects the strength of time-series versus cross-sectional strategies.

Multiple Panel regressions with fixed effects

➤ Thus far, we have studied univariate panel regressions.

➤ Consider now a multiple panel regression with stock fixed effects

$$r_{it+1} = a_i + b_{TS}' z_{it} + \varepsilon_{it+1}$$

➤ z_{it} is an M -vector of characteristics for firm i at month t and b_{TS} is an M -vector of slope coefficients.

➤ We thus have to estimate N intercepts (a_1, a_2, \dots, a_N) along with M slopes, altogether $(N+M)$ parameters.

➤ OLS coefficients are given by

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_N \\ \hat{b}_{TS} \end{bmatrix} = (X'X)^{-1} X'R$$

➤ Let $\tilde{T} = \sum_{i=1}^N T_i$, then R is a \tilde{T} -vector, $R' = [r'_1, r'_2, \dots, r'_N]$, with r_i being a T_i -vector of returns for firm i .

Multiple Panel regressions with fixed effects

- In addition, X is a $\tilde{T} \times (M + N)$ matrix:

$$X = \begin{pmatrix} \iota_1 & 0 & \cdots & 0 & z_1 \\ 0 & \iota_2 & \cdots & 0 & z_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \iota_N & z_N \end{pmatrix}$$

- Then, it can be shown that

$$\hat{b}_{TS} = \sum_{i=1}^N w_i \hat{b}_i$$

where $w_i = T_i (\sum_{s=1}^N T_s \hat{\Sigma}_{z_s})^{-1} \hat{\Sigma}_{z_i}$, $\hat{\Sigma}_{z_i}$ is the time-series covariance matrix of firm i characteristics, and \hat{b}_i is an M -vector of slope estimates from individual predictive regressions of future returns on current values of firm characteristics.

- As in the univariate case, the slope is a value weighted average of firm-level slopes but now weights are formulated through an $M \times M$ matrix.
- The slope can also be represented as

$$\hat{b}_{TS} = \left(\sum_{s=1}^N T_s \hat{\Sigma}_{z_s} \right)^{-1} \sum_{i=1}^N T_i \hat{\Sigma}_{zr_i}, \text{ where } \hat{\Sigma}_{zr_i} \text{ is a covariance vector between future return and } M \text{ characteristics.}$$

Multiple Panel regressions with fixed effects

- Similarly, consider a multiple regression with month fixed-effects:

$$r_{it+1} = a_t + b_{CS}' z_{it} + \varepsilon_{it+1}$$

- We estimate T intercepts and M slopes coefficients, altogether $T+M$ parameters.
- The slope estimate is given by:

$$\hat{b}_{CS} = \sum_{t=1}^T w_t \hat{b}_t$$

where \hat{b}_t is an M -vector of slopes estimated from monthly cross section regressions, $w_t = N_t \left(\sum_{s=1}^T N_s \hat{\Sigma}_{z_s} \right)^{-1} \hat{\Sigma}_{z_t}$, and $\hat{\Sigma}_{z_t}$ is the month- t cross-sectional covariance matrix of firm characteristics.

Panel regressions with Common Factors

- Let us now extend the panel regression setup to account for common factors.
- To keep it simple, let us consider a single factor model.
- Then, the panel regression is formulated as

$$r_{it+1} = a_i + r_{ft} + \beta_i f_{t+1} + \gamma z_{it} + \varepsilon_{it+1}$$

- r_{ft} is the risk-free rate for period $t+1$, observed at time t .
- Clearly, r_{ft} and f_{t+1} vary only through time but not across stocks.
- In the time-invariant beta setup, beta varies only across stocks.
- Essentially, we ask what does the $\beta_i f_{t+1}$ component capture?
- It can capture cross-sectional predictability if β_i is correlated, in the cross section, with the characteristic.
- It can also capture time-series predictability if, for at least one stock, f_{t+1} is correlated, in the time series, with z_{it} .
- We formalize these concepts below.

Panel regressions with Common Factors

- Let us define risk adjusted excess return as $\tilde{r}_{it+1} = r_{it+1} - r_{ft} - \beta_i f_{t+1}$.
- The corresponding payoffs to the time-series and cross-sectional strategies are given by

$$\begin{aligned}\tilde{\varphi}_{TS} &= \sum_i \sum_t (z_{it} - \bar{z}_i) \tilde{r}_{it+1} = \tilde{A} + \tilde{B} \\ \tilde{\varphi}_{CS} &= \sum_i \sum_t (z_{it} - \bar{z}_t) \tilde{r}_{it+1} = \tilde{C} + \tilde{B}\end{aligned}$$

where \tilde{A} , \tilde{B} , and \tilde{C} are similar to A , B , and C except that \tilde{r}_{it+1} is replacing r_{it+1} .

- Then, it can be shown that

$$\begin{aligned}\tilde{A} &= A - \sum_i \sum_t (z_{it} - \bar{z}_i) \beta_i f_{t+1} \\ \tilde{A} &= A - \sum_i \beta_i \sum_t (z_{it} - \bar{z}_i) f_{t+1} \\ \tilde{A} &= A - \sum_i \beta_i T_i \hat{\sigma}_{zf,i}\end{aligned}$$

where $\hat{\sigma}_{zf,i}$ is the time-series covariance between f_{t+1} and z_{it} and T_i is the number of time-series observations per stock i .

Panel regressions with Common Factors

- ▶ The last equation formalizes the notion that including common factors can capture time-series predictability by firm characteristics as long as the factor is correlated with the lagged characteristic for at least one stock.
- ▶ Clearly, if $\hat{\sigma}_{zf, i}$ is equal to zero for each of the stocks then asset pricing factors do not explain any time-series predictability.
- ▶ From a cross-sectional perspective, it follows that

$$\begin{aligned}\tilde{C} &= C - \sum_i \sum_t (\bar{z}_{it} - \bar{z}_t) \beta_i f_{t+1} \\ &= C - \sum_t f_{t+1} \sum_i (\bar{z}_{it} - \bar{z}_t) \beta_i \\ &= C - \sum_{t=1}^T N_t f_{t+1} \hat{\sigma}_{\beta z, t}\end{aligned}$$

where $\hat{\sigma}_{\beta z, t}$ is the cross-section covariance between β_i and z_{it} while N_t is the number of stocks at month t .

- ▶ Asset pricing factors could explain cross-sectional predictability as long as factor loadings are correlated with firm characteristics, for at least one period.

Predictive Regressions: statistical evidence and economic restrictions

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 system
$$r_t = a + \beta z_{t-1} + u_t$$
$$z_t = c + \rho z_{t-1} + v_t$$
- Statistically, predictability means that the β coefficient 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.

Predictive regressions- Finite sample bias in the slope coefficients

- Re-writing 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.

Stock Return Predictability - Is The Evidence on Predictability Robust?

- Predictability based on macro variables is still a research controversy:
 - Asset pricing theories often do not identify variables that predict asset returns. For instance, Menzly, Santos, and Veronesi (2006), just like other studies cited on the previous page, provide theoretical validity for predictability – but ex post. There are two exceptions. The present value formula clearly identifies the dividend-to-price (or consumption to wealth) as a potential predictor. The lower bound identifies SVIX.
 - 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.
- 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 essentially biased in favor of detecting predictability.
- The out-of-sample performance of the selected model is always a disaster.
- The Bayesian approach of model combination could improve out-of-sample performance (see Avramov (2002)). In fact, Avramov (2002) warns against using model selection criteria.
- 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.

Economic restrictions on predictive regressions

- In the presence of biases in predictive regressions, data mining concerns, as well as dismal out of sample predictive power, financial economists have seriously questioned the notion that returns are really predictable.
- There have been several major responses to that concern.
- Martin derives a lower bound on the equity premium and shows that this bound is an adequate predictor of future return.
- Cochrane is using the log linearization to tell us that if dividend growth is unpredictable by the dividend-to-price ratio then it must be the case that returns are predictable.
- Some other economic restrictions include non negative expected return (e.g., Campbell and Thompson (2008)).
- More details are coming up next.

Predictability and Lower bound on the equity premium

- Martin (2017) contributes to the literature on return predictability through formulating a lower bound on the equity premium under sensible assumptions on investor preferences.
- The lower bound is based on option prices.
- This is perhaps surprising because option pricing formulas (e.g., Black and Sholes) do not display the mean return, or the drift.
- Here are formal details.
- Let X_T be a payoff to be paid at time T and let M_T be the corresponding pricing kernel.
- The time t price of the future payoff can be described in two different ways

$$P_t = E_t(M_T X_T) = \frac{1}{R_{ft}} E_t^* X_T$$

- E_t is the conditional expectations operator under the physical measure, while E_t^* is the corresponding operator under the risk-neutral measure.
- The second equality reflects the notion that in a risk neutral world, an asset price is equal to the present value of expected cash flows, where discounting is based on the risk free rate.
- Letting $X_T = R_T^2$, it follows that

$$E_t(M_T R_T^2) = \frac{1}{R_{ft}} E_t^* R_T^2$$

Lower bound on the equity premium

- R_{ft} is one plus the risk free rate for the period starting at time t ending at T .

- The risk-neutral variance of future return is

$$VaR_t^*(R_T) = E_t^*(R_T^2) - [E_t^*(R_T)]^2 = R_{ft}E_t(M_T R_T^2) - R_{ft}^2$$

- Now, the equity premium is, by definition,

$$\begin{aligned} E_t(R_T) - R_{ft} &= [E_t(R_T) - E_t(M_T R_T^2)] - [R_{ft} - E_t(M_T R_T^2)] \\ &= \frac{1}{R_{ft}} VaR_t^*(R_T) - cov_t(M_T R_T, R_T) \end{aligned}$$

- Notice that

$$cov_t(M_T R_T, R_T) = E_t(M_T R_T^2) - E_t(M_T R_T)E_t(R_T) = E_t(M_T R_T^2) - E_t(R_T)$$

where the last equality follows because the asset pricing relation is about $E_t(M_T R_T) = 1$

- As long as $cov_t(M_T R_T, R_T)$ is non-positive for any time period, the lower bound applies

$$E_t(R_T) - R_{ft} \geq \frac{1}{R_{ft}} VaR_t^*(R_T)$$

- The negative correlation condition (NCC) holds for flexible set of preferences.

Lower bound on the equity premium

- For example, consider the power preferences: $U(W_T) \propto W_T^{1-\gamma}$, then $M_T \propto U'(W_T)$.
- The covariance is negative because the expression $R_T U'(W_T)$ is decreasing in R_T
- The essential step now is to measure the risk-neutral variance

$$\frac{1}{R_{ft}} \text{VaR}_t^*(R_T)$$

- To understand how to solve for $\text{VaR}_t^*(R_T)$, we start with the Carr-Madan (CM) formula.
- In particular, let $g(S_T)$ be a payoff that depends on S_T , the stock price at time T
- In addition, $g(x)$ is a continuously twice differentiable function.
- The CM formula states that

$$g(S_T) = g(F_{t,T}) + g'(F_{t,T})(S_T - F_{t,T}) + \int_0^{F_{t,T}} g''(K)(K - S_T)^+ dK + \int_{F_{t,T}}^{\infty} g''(K)(S_T - K)^+ dK$$

Lower bound on the equity premium

where,

❖ $F_{t,T}$ is the future price on the stock with delivery at time T , thus $F_{t,T} = S_T R_{ft}$ in the absence of dividend payments.

❖ $(S_T - K)^+ = \max[S_T - K, 0]$

❖ $(K - S_T)^+ = \max[K - S_T, 0]$

➡ Now, let $g(S_T) = \left(\frac{S_T}{S_t}\right)^2$

➡ Applying the CM formula to $\left(\frac{S_T}{S_t}\right)^2$ we get

$$\left(\frac{S_T}{S_t}\right)^2 = \left(\frac{F_{t,T}}{S_t}\right)^2 + \frac{2F_{t,T}}{S_t^2} (S_T - F_{t,T}) + \frac{2}{S_t^2} \left[\int_0^{F_{t,T}} (K - S_T)^+ dK + \int_{F_{t,T}}^{\infty} (S_T - K)^+ dK \right]$$

➡ Notice that $E_t^*(S_T) = F_{t,T}$

Lower bound on the equity premium

- Then, taking risk-neutral expectations from both sides of the equation yields

$$E_t^* \left(\frac{S_T}{S_t} \right)^2 = R_{ft}^2 + \frac{2R_{ft}}{S_t^2} \left[\int_0^{F_{t,T}} P_{t,T}(K) dK + \int_{F_{t,T}}^{\infty} C_{t,T}(K) dK \right]$$

where $C_{t,T}(K)$ and $P_{t,T}(K)$ are current prices of European call and put options on the underlying stock with maturity at time T and strike price K .

- Having at hand the expression for $E_t^* \left(\frac{S_T}{S_t} \right)^2$, the risk-neutral variance follows.
- In particular,

$$VaR_t^*(R_T) = E_t^* \left(\frac{S_T}{S_t} \right)^2 - R_{ft}^2$$

- Thus,

$$\frac{VaR_t^*(R_T)}{R_{ft}} = \frac{2}{S_t^2} \left[\int_0^{F_{t,T}} P_{t,T}(K) dK + \int_{F_{t,T}}^{\infty} C_{t,T}(K) dK \right]$$

Lower bound on the equity premium

- Martin defines the simple VIX (SVIX) to be equal to the risk natural variance.
- Using the property $F_t = S_t R_{ft}$, we can also express $SVIX_t^2$ as

$$SVIX_t^2 = \frac{2R_{ft}}{(T-t)F_{t,T}^2} \left\{ \int_0^{F_{t,T}} P_{t,T}(K) dK + \int_{F_{t,T}}^{\infty} C_{t,T}(K) dK \right\}$$

- VIX is similar to SVIX, but is more sensitive to left tail events.
- In particular, VIX_t^2 is given by

$$VIX_t^2 = \frac{2R_{ft}}{(T-t)} \left\{ \int_0^{F_{t,T}} \frac{1}{K^2} P_{t,T}(K) dK + \int_{F_{t,T}}^{\infty} \frac{1}{K^2} C_{t,T}(K) dK \right\}$$

- The idea that risk-neutral variance of a stock return can be computed the way suggested here goes back to Bakshi and Madan (2000), Carr and Madan (2001), and Bakshi, Kapadia, and Madan (2003).

Measuring individual stock expected return

- Option prices can also be used to measure individual stock expected returns.
- To illustrate, the risk-neutral covariance between $\frac{1}{m}$ and R_i and the variance of $\frac{1}{m}$ are given by

$$\begin{aligned}Cov^*\left(\frac{1}{m}, R_i\right) &= E^*\left(\frac{1}{m}, R_i\right) - E^*\frac{1}{m}E^*R_i \\&= R_{ft}(ER_i - R_{ft}) \\Var^*\left(\frac{1}{m}\right) &= E^*\left(\frac{1}{m}\right)^2 - \left(E^*\frac{1}{m}\right)^2 \\&= R_{ft}E\left(\frac{1}{m}\right) - R_{ft}^2\end{aligned}$$

- Dividing both equations, we get

$$E_t R_{i, t+1} - R_{ft} = \beta_{i, t}^* \left(E_t \frac{1}{m_{t+1}} - R_{ft} \right)$$

where,

$$\beta_{i, t}^* = \frac{Cov_t^*\left(\frac{1}{m_{t+1}}, R_{i, t+1}\right)}{Var_t^*\left(\frac{1}{m_{t+1}}\right)}$$

Measuring individual stock expected return

- This equation establishes a single factor model that only requires the existence of SDF.
- The basic pricing equation holds for any asset, be it stock, bond, option, and real investment, and it holds for any two periods in a dynamic setup.
- It does not assume that
 - markets are complete or the existence of a representative investor
 - asset returns are normally distributed or IID through time
 - Setup is static
 - particular form of preferences (such as separable utility)
 - there is no human capital or any other source of non-asset income
 - markets have reached an equilibrium.
- Martin and Wagner (2019) formulate individual excess expected return as

$$E_t \frac{R_{i,t+1} - R_{ft}}{R_{ft}} = SVIX_t^2 + \frac{1}{2} (SVIX_{i,t}^2 - \overline{SVIX_t^2})$$

- market volatility: $SVIX_t$
- volatility of stock i : $SVIX_{i,t}$
- average stock volatility: $\overline{SVIX_t}$
- The formula requires observations of option prices but no estimation.
- But it is based on three nontrivial approximations.

Using options to measure credit spread

- Culp, Nozowa, and Veronesi (2018) analyze credit risk using “pseudo firms” that purchase traded assets financed with equity and zero-coupon bonds.
- Pseudo bonds are equivalent to Treasuries minus put options on pseudo firm assets.
- They find that pseudo bond spreads are large, countercyclical, and predict lower economic growth, just like real corporate spreads.
- Using that framework, they show that bond market illiquidity, investors’ overestimation of default risks, and corporate frictions do not seem to explain excessive observed credit spreads.
- Instead, a risk premium for tail and idiosyncratic asset risks is the primary determinant of corporate spreads.
- Below, I provide the intuition for their work.
- Consider a firm that has no dividend-paying equity outstanding, and a single zero-coupon debt issue.
- The time t values of the assets of the firm, the debt, and the equity are A_t , B_t , and E_t , respectively. The debt matures at time T .
- The value of the equity at time T is

$$E_T = \max(0, A_T - \bar{B})$$

- This is the payoff to a call option.
- What is the underlying asset?
- What is the strike price?

Interpreting Default-able Bonds

- The value of the debt is

$$B_T = \min(A_T, \bar{B})$$

or

$$B_T = A_T + \min(0, \bar{B} - A_T) = A_T - \max(0, A_T - \bar{B})$$

- This implies that corporate bondholders could be viewed as those owning the firm assets, but have written a call option on the firm assets to the equity-holders.
- Or the bondholders own a default free bond and have written a put option on the firm assets.
- You can use the call-put parity to verify that both perspectives are indeed equivalent.
- Thus, we can compute the value of debt and equity prior to time T using option pricing methods, with the value of assets taking the place of the stock price and the face value of the debt taking the place of the strike price
- The equity value at time t is the value of a call option on the firm assets. The value of the debt is then $B_t = A_t - E_t$.

Pricing Zero Coupon Bonds with Default Risk using the B&S Formula

- Suppose that a non dividend paying firm issues a zero coupon bond maturing in five years.
- The bond's face value is \$100, the current value of the assets is \$90, the risk-free rate (cc) is 6%, and the volatility of the underlying assets is 25%.
- What is the equity and debt value?
- What is the bond's yield to maturity (ytm)?
- The Black-Scholes Formula Revisited

Call Option price:

$$C(S, K, \sigma, r, T, \delta) = Se^{-\delta T} N(d_1) - Ke^{-rT} N(d_2)$$

Put Option price:

$$P(S, K, \sigma, r, T, \delta) = Ke^{-rT} N(-d_2) - Se^{-\delta T} N(-d_1)$$

Default Risk Premium

- The equity value solves the BSCall:

$$\text{Equity} = \text{BSCall}(90, 100, 0.25, 0.06, 5, 0) = 27.07$$

- The debt value is thus $90 - 27.07 = 62.93$.
- The debt cc ytm $= 1/5 \times \ln(100/62.93) = 9.26\%$.
- The ytm is greater than the risk free rate due to default risk premium.
- The default risk premium is equal to

$$\exp(0.0926) - \exp(0.06) = 3.52\%$$

The Campbell-Shiller (CS) present value formula

- The CS decomposition is yet another economic based response to the concern that returns are unpredictable.
- The notion is to identify the dividend-to-price ratio as a return predictor.
- The setup is developed below.
- 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

$$\begin{aligned} \log[1 + \exp(d_{t+1} - p_{t+1})] &\approx \\ -\log \rho - (1 - \rho) \log\left(\frac{1}{\rho} - 1\right) &+ (1 - \rho)(d_{t+1} - p_{t+1}) \end{aligned}$$

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

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

- 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 .

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,

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

- 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:

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

- ▶ 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:

$$\begin{array}{l} r_t^e \\ \vdots \\ r_N^e \end{array} = \begin{array}{l} X \\ \vdots \\ X \end{array} \begin{array}{l} \theta_1 \\ \vdots \\ \theta_N \end{array} + \begin{array}{l} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{array} \quad \text{where} \quad \begin{array}{l} X \\ \vdots \\ X \end{array} = \begin{array}{l} [1, r_{m1}^e] \\ \vdots \\ [1, r_{mT}^e] \end{array} \\ \begin{array}{l} T_{x1} \\ \vdots \\ T_{x1} \end{array} \quad \begin{array}{l} 2_{x1} \\ \vdots \\ 2_{x1} \end{array} \quad \begin{array}{l} T_{x1} \\ \vdots \\ T_{x1} \end{array} \quad \theta_i = [\alpha_i, \beta_i]'$$

- Retain the estimated regression intercepts

$$\hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_N]' \quad \text{and} \quad \begin{array}{l} \hat{\varepsilon} \\ T_{xN} \end{array} = [\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N]$$

- 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

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

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

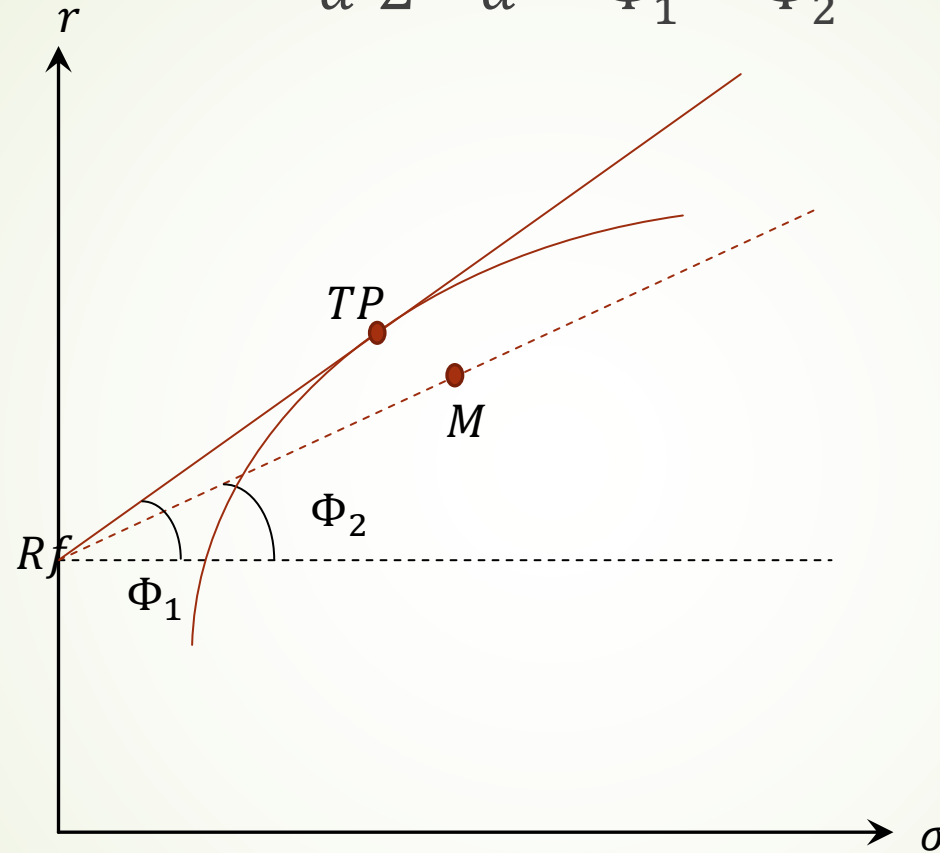
or

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

- 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 GMM: Econometrics Setup and Applications

GMM

- We can test theories in financial economics by the GMM of Hansen (1982).
- Let us describe the basic concepts of GMM and propose some applications.
- Let Θ be an $m \times 1$ vector of parameters to be estimated from a sample of observations x_1, x_2, \dots, x_T .
- One drawback in the maximum likelihood principle is that it requires specifying the joint density of the observations.
- The ML principle is indeed a parametric one.
- ML typically makes the IID, Normal, and homoscedastic assumptions.
- All these assumptions can be relaxed in the GMM framework.
- The GMM only requires specification of certain moment conditions (often referred as orthogonality conditions) rather than the full density.

- It is therefore considered a nonparametric approach.
- Do not get it wrong: The GMM is not ideal.
- First, it may not make efficient use of all the information in the sample.
- Second, non parametric approaches typically have low power in out-of-sample tests possibly due to over-fitting.
- Also the GMM is asymptotic and can deliver poor, even measurable, final sample properties.
- Let $f_t(\Theta)$ be an $r \times 1$ vector of moment conditions.
- Note that f_t is not necessarily linear in the data or the parameters, and it can be heteroskedastic and serially correlated.
- If $r = m$, i.e., if there is the same number of parameters as there are moments, then the system is exactly identified.
- In this case, one could find the GMM estimate $\hat{\Theta}$, which satisfies

$$E(f_t(\hat{\Theta})) = 0.$$

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

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

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

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

- Otherwise, the GMM estimator minimizes the quadratic form

$$J_T(\Theta) = g_T(\Theta)' W_T g_T(\Theta),$$

- Here, W_T is some $r \times r$ weighing matrix to be discussed later.

- Differentiating with respect to Θ yields

$$D_T(\Theta)'W_Tg_T(\Theta),$$

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

$$D_T(\hat{\Theta})'W_Tg_T(\hat{\Theta}) = 0.$$

- Observe that the left (and obviously the right) hand side is an $m \times 1$ vector. Therefore, as $r > m$ only a linear combination of the moments, given by $D_T(\hat{\Theta})'W_T$, is set to zero.
- Hansen (1982, theorem 3.1) tells us the asymptotic distribution of the GMM estimate is

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

where

$$V = (D_0'WD_0)^{-1}D_0'WSWD_0(D_0'WD_0)^{-1},$$

$$S = \lim_{T \rightarrow \infty} \text{var}[\sqrt{T}g_T(\Theta)],$$

$$= \sum_{j=-\infty}^{\infty} E[f_t(\Theta)f_{t-j}(\Theta)'],$$

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

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

- If the moment conditions are serially uncorrelated then

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

- We have not yet addressed the issue of how to choose the optimal weighting matrix.
- Hansen shows that optimally $W = S^{-1}$.

- The optimal V matrix is therefore

$$V^* = (D_0' S^{-1} D_0)^{-1}.$$

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

$$TJ_T(\hat{\Theta}) \sim \chi_{r-m}^2.$$

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

$$g_T(\hat{\Theta})' \{var[g_T(\hat{\Theta})]\}^{-1} g_T(\hat{\Theta}) \sim \chi_{r-m}^2.$$

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

Application# 1: Estimating the mean of a time series

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

$$f_t(\Theta) = (x_t - \mu).$$

- The system is exactly identified.
- Notice that

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

- Setting

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

- Then the GMM estimate for μ is the sample mean

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

- Moreover, if x_t 's are uncorrelated then

$$S = E[f_t(\Theta)f_t(\Theta)'],$$

estimated using

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

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

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

- The optimal V matrix is then given by

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

- Use S_T as a consistent estimator.

- Asymptotically, we get

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

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

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

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

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

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

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

$$f_t(\Theta) = x_t(r_t - x_t' \beta),$$

where

$$r_t = r_{i,t},$$

$$x_t = [1, r_{m,t}]',$$

$$\epsilon_t = \epsilon_{i,t}, \text{ and } \beta = [\alpha_i, \beta_i]'.$$

- Let us now compute the sample moment

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

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

$$\begin{aligned} \hat{\beta} &= (\sum_{t=1}^T x_t x_t')^{-1} (\sum_{t=1}^T x_t r_t), \\ &= (X'X)^{-1} X'R, \end{aligned}$$

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

- The GMM estimator for β is the usual OLS estimator.
- To find V first compute

$$\begin{aligned} \frac{\partial f_t(\theta)}{\partial \theta'} &= -x_t x_t', \\ \frac{\partial g_T(\theta)}{\partial \theta'} &= D_T(\Theta) = -\frac{1}{T} \sum_{t=1}^T x_t x_t' = -\frac{X'X}{T}. \end{aligned}$$

- Moreover, if ϵ 's are serially uncorrelated then

$$\begin{aligned} S_T &= \frac{1}{T} \sum_{t=1}^T f_t(\hat{\Theta}) f_t(\hat{\Theta})', \\ &= \frac{1}{T} \sum_{t=1}^T x_t \hat{\epsilon}_t \hat{\epsilon}_t' x_t', \\ &= \frac{1}{T} \sum_{t=1}^T x_t x_t' \hat{\epsilon}_t^2. \end{aligned}$$

- Using the optimal weighting matrix, it follows that

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

estimated by

$$\begin{aligned} \widehat{\text{Var}}(\hat{\beta}) &= \frac{1}{T} (D_T' S_T^{-1} D_T)^{-1}, \\ &= (X'X)^{-1} \left(\sum_{t=1}^T x_t x_t' \hat{\epsilon}_t^2 \right) (X'X)^{-1}. \end{aligned}$$

Application # 3: Testing the CAPM

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

$$r_t^e = \alpha + \beta r_{mt}^e + \epsilon_t.$$

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

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

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

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

2. Compute D_0 .

$$\begin{aligned}\frac{\partial f_t(\theta)}{\partial \theta'} &= x_t \otimes -(x_t' \otimes I_N), \\ &= -\begin{bmatrix} 1 & r_{m,t}^e \\ r_{m,t}^e & r_{m,t}^{e2} \end{bmatrix} \otimes I_N.\end{aligned}$$

Moreover,

$$\begin{aligned}D_0 &= E \left[\frac{\partial g_T(\theta)}{\partial \theta'} \right] \\ &= E \left[\frac{\partial f_t(\theta)}{\partial \theta'} \right] \\ &= -\begin{bmatrix} 1 & \mu_m \\ \mu_m & \mu_m^2 + \sigma_m^2 \end{bmatrix} \otimes I_N,\end{aligned}$$

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

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

4. That is,

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

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

$$\begin{aligned} g_T(\Theta) &= \frac{1}{T} \sum_{t=1}^T f_t(\Theta), \\ &= \frac{1}{T} \sum_{t=1}^T (x_t \otimes \epsilon_t). \end{aligned}$$

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

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

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

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

The GMM estimator is thus given by

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

These are the OLS estimators for the CAPM parameters.

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

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

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

$$V_T = (D_T' S_T^{-1} D_T)^{-1}$$

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

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

so you should substitute V_T for V .

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

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

12. The statistic J_7 in CLM is derived using the Wald statistic

$$\begin{aligned} J_7 &= \hat{\Theta}' R' \left(R \left(\frac{1}{T} V_T \right) R' \right)^{-1} R \hat{\Theta}, \\ &= T \hat{\alpha}' (R (D_T' S_T^{-1} D_T)^{-1} R')^{-1} \hat{\alpha}. \end{aligned}$$

13. Under the null hypothesis $J_7 \sim \chi_N^2$.

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

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

- Thus J_7 becomes the $J1$ Wald statistic derived earlier
- So, the GMM test statistic is a generalized version of Wald correcting for heteroscedasticity.
- You can also relax the non serial correlation assumption.

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

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

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

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

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

$$\begin{aligned}\mathbb{E}[r_t | z_{t-1}] &= \delta_r x_{t-1}, \\ \mathbb{E}[r_{mt} | z_{t-1}] &= \delta_m x_{t-1}.\end{aligned}$$

- Then, let us define several residuals

$$u_{rt} = r_t - \delta_r x_{t-1},$$

$$u_{mt} = r_{mt} - \delta_m x_{t-1},$$

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

- Collecting the residuals into one vector yields

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

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

- That is, there are $(M + 1)(N + 1) + 1$ parameters.
- How many moment conditions do we have? More than you think!

- Note that

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

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

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

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

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

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

Application # 5: Estimating Standard Errors in the presence of correlation among firms

- This application builds on Avramov, Chordia, and Goyal (2006b).
- Consider N stocks with T observations.
- The dependant variable is denoted by y and the set of K independent variables by x .
- For instance, y could denote volatility and x could include lags of volatility, day of week dummies, and trading-related variables.
- The regression equation is as follows:

$$y_{it} = x'_{it}\beta_i + \epsilon_{it}$$

- Let us introduce some notation. Let $\beta = (\beta'_1, \dots, \beta'_N)'$, let $x_t = (x'_{1t}, \dots, x'_{Nt})'$, $X_i = (x_{i1}, \dots, x_{iT})'$, and let $Y_i = (y_{i1}, \dots, y_{iT})'$. Notice that β is an NK vector.
- Moment conditions are written as $E(f(x_t, \beta)) = 0$, where $f(x_t, \beta)$ is an NK valued function given by

$$f(x_t, \beta) = \begin{pmatrix} x_{1t}\epsilon_{1t} \\ \vdots \\ x_{Nt}\epsilon_{Nt} \end{pmatrix} \equiv \begin{pmatrix} x_{1t}(y_{1t} - x'_{1t}\beta_1) \\ \vdots \\ x_{Nt}(y_{Nt} - x'_{Nt}\beta_N) \end{pmatrix}$$

Calculation of Standard Errors

- Since the number of moment conditions is exactly equal to the number of parameters, the system is exactly identified.
- Thus, we do not need the usual weighting matrix to carry out the optimization.
- The solution is, of course, given by the usual OLS

$$\hat{\beta}_i = (X_i' X_i)^{-1} X_i' Y_i$$

- The variance of the estimator is given from the GMM formula as

$$T \times \text{cov}(\hat{\beta}) = D^{-1} S D^{-1'}$$

where the score matrix D and the spectral density matrix S are given by

$$D = E \left(\frac{\partial f(x_t, \beta)}{\partial \beta'} \right)$$

$$S = \sum_{s=-\infty}^{\infty} E[f(x_t, \beta) f(x_{t-s}, \beta)']$$

Calculation of Standard Errors

- The special structure of f and β makes the computation of D especially straightforward.
- In particular, we have

$$\frac{\partial f(x_t, \beta)}{\partial \beta'} = \begin{bmatrix} -x_{1t}x'_{1t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & -x_{Nt}x'_{Nt} \end{bmatrix}$$

- Thus,

$$\hat{D} = \hat{E} \left(\frac{\partial f(x_t, \beta)}{\partial \beta'} \right) = \frac{1}{T} \begin{bmatrix} -X'_1X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & -X'_NX_N \end{bmatrix}$$

where 0 is suitably defined $K \times K$ matrices.

- Thus, the matrix D consists of K^2 blocks with the (i, j) th block equal to zero if $i \neq j$ and equal to X'_iX_i when $i = j$.
- The diagonal nature of this matrix also makes the computation of the inverse of D trivial.

Calculation of Standard Errors

- The matrix S consists of K^2 blocks, where the (i, j) th block is given by

$$S_{ij} = \sum_{s=-\infty}^{\infty} E[x_{it}x'_{jt-s}\epsilon_{it}\epsilon'_{jt-s}]$$

and is estimated using the Newey–West estimation technique with L lags as

$$\widehat{S}_{ij} = \frac{1}{T} \left[\sum_{t=1}^T (x_{it}x'_{jt}e_{it}e'_{jt}) + \sum_{s=1}^L \frac{L-s}{L} \sum_{t=s+1}^T (x_{it}x'_{jt-s}e_{it}e'_{jt-s} + x_{jt}x'_{it-s}e_{jt}e'_{it-s}) \right]$$

- Combining all the pieces together, we get

$$\widehat{cov}(\widehat{\beta}_i, \widehat{\beta}_j) = (X_i'X_i)^{-1}T\widehat{S}_{ij}(X_j'X_j)^{-1}$$

- Finally, can consider the average β coefficient across all stocks.
- This is estimated as

$$\widehat{\beta} = \frac{1}{N} \sum_{i=1}^N \widehat{\beta}_i$$

with the variance given by

$$\widehat{cov}(\widehat{\beta}) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \widehat{cov}(\widehat{\beta}_i, \widehat{\beta}_j)$$

Application # 6: Testing return predictability over long horizon

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

- The model estimated is

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

where $R_{t+k} = \sum_{i=1}^k r_{t+i}$, with r_{t+i} being log return at time $t + i$.

- Fama and French (1989) observe a dramatic increase in the sample R^2 as the return horizon grows from one month to four years.
- Kirby (1997) challenges the Fama-French findings using a GMM framework that accounts for serial correlation in the residuals.
- Let us formalize his test statistic.
- There are $M + 1$ parameters $\Theta = (\alpha, \beta')'$, where β is a vector of dimension M .

- From Hansen (1982)

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

where

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

- There are $M + 1$ moment conditions:

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

- Compute D_0 :

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

- It would be useful to take the inverse of D_0

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

- And the matrix S is given by

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E} \begin{bmatrix} \epsilon_{t+k} \epsilon_{t+k-j} & \epsilon_{t+k} \epsilon_{t+k-j} z_{t-j}' \\ \epsilon_{t+k} \epsilon_{t+k-j} z_t & \epsilon_{t+k} \epsilon_{t+k-j} z_t z_{t-j}' \end{bmatrix}.$$

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

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

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

- It follows that

$$\begin{aligned} \tilde{V} &= \sum_{j=-\infty}^{\infty} \mathbb{E} \begin{bmatrix} \mu_z' \Sigma_z^{-1} \\ -\Sigma_z^{-1} \end{bmatrix}' \begin{bmatrix} \epsilon_{t+k} \epsilon_{t+k-j} & \epsilon_{t+k} \epsilon_{t+k-j} Z_{t-j}' \\ \epsilon_{t+k} \epsilon_{t+k-j} Z_t & \epsilon_{t+k} \epsilon_{t+k-j} Z_t Z_{t-j}' \end{bmatrix} \begin{bmatrix} \mu_z' \Sigma_z^{-1} \\ -\Sigma_z^{-1} \end{bmatrix}, \\ &= \Sigma_z^{-1} \left[\sum_{j=-\infty}^{\infty} \mathbb{E} (\epsilon_{t+k} \epsilon_{t+k-j}) (Z_t - \mu_z) (Z_{t-j} - \mu_z)' \right] \Sigma_z^{-1}, \\ &= \Sigma_z^{-1} \left[\sum_{j=-\infty}^{\infty} \mathbb{E} (\delta_{t+k} \delta_{t+k-j}') \right] \Sigma_z^{-1}, \end{aligned}$$

where

$$\begin{aligned} \delta_{t+k} &= \epsilon_{t+k} Z_t, \\ \delta_{t+k-j} &= \epsilon_{t+k-j} Z_{t-j}. \end{aligned}$$

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

► Then,

$$\tilde{V} = \Sigma_Z^{-1} [\mathbb{E}(\delta_{t+k} \delta_{t+k}')] \Sigma_Z^{-1},$$

which can be estimated by

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

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

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

$$\tilde{V} = \Sigma_Z^{-1} [\mathbb{E}(\epsilon_{t+k} \epsilon_{t+k-j}) \mathbb{E}(z_t - \mu_Z)(z_{t-j} - \mu_Z)'] \Sigma_Z^{-1},$$

► In that case

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

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

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

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

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

suggesting that

$$TR^2 \sim \chi^2(M).$$

- So we are able to derive a limiting distribution for the regression R^2 .
- Kirby considers cases with heteroscedasticity and serial correlation.
- Then the distribution of the regression slope coefficient and the R^2 are much more complex.
- His conclusion: the R^2 in a predictive regression does not increase with the investment horizon
- That is almost a no brainer.

Application # 7: Yet, another Testing return predictability over long horizon

- Boudoukh, Richardson, and Whitelaw (2006) is another interesting application of the GMM in the context of long-horizon return predictability
- Like Kirby (1997), they show that the sample evidence does not support predictability
- Further, they show that for persistent predictors the estimates of slope coefficients are almost perfectly correlated across horizons under the null hypothesis of no predictability
- They consider regression systems of the following type:

$$\begin{aligned} R_{t,t+1} &= \alpha_1 + \beta_1 Z_t + \varepsilon_{t,t+1} \\ &\vdots \\ R_{t,t+j} &= \alpha_j + \beta_j Z_t + \varepsilon_{t,t+j} \\ &\vdots \\ R_{t,t+K} &= \alpha_K + \beta_K Z_t + \varepsilon_{t,t+K} \end{aligned}$$

- Under the null hypothesis of no predictability

$$\beta_1 = \dots = \beta_j = \dots = \beta_K = 0$$

- And the same number of moment conditions corresponding to the regression system as

$$f_t(\theta) = \begin{bmatrix} (R_{t,t+1} - \alpha_1 - \beta_1 Z_t) \\ (R_{t,t+1} - \alpha_1 - \beta_1 Z_t) Z_t \\ \vdots \\ (R_{t,t+j} - j\alpha_1 - \beta_j Z_t) Z_t \\ \vdots \\ (R_{t,t+K} - K\alpha_1 - \beta_K Z_t) Z_t \end{bmatrix}$$

where $\theta = (\alpha_1, \beta_1, \dots, \beta_K)'$

- There are $k+1$ parameters and moment conditions
- Under the null, the regression estimate θ has an asymptotic normal distribution with mean $(\alpha_1, 0)'$ and covariance matrix $(D_0' S_0^{-1} D_0)^{-1}$ where $D_0 = E \left(\frac{\partial f_t}{\partial \theta} \right)$ and $S_0 = \sum_{j=-\infty}^{\infty} E(f_t f_{t-j})$
- D_0 is easily calculated as

$$D_0 = - \begin{pmatrix} 1 & \mu_Z & 0 & \cdots & \cdots & \cdots \\ \mu_Z & \mu_Z^2 + \sigma_Z^2 & 0 & \cdots & \cdots & \cdots \\ \vdots & 0 & \ddots & 0 & \cdots & \cdots \\ j\mu_Z & \vdots & 0 & \mu_Z^2 + \sigma_Z^2 & 0 & \cdots \\ \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ J\mu_Z & \vdots & \vdots & \vdots & 0 & \mu_Z^2 + \sigma_Z^2 \end{pmatrix}$$

$$S_0 = \begin{bmatrix} \sigma_R^2 & \sigma_R^2 \mu_Z & \dots & j \sigma_R^2 \mu_Z & \dots & k \sigma_R^2 \mu_Z \\ \sigma_R^2 \mu_Z & \sigma_R^2 (j \mu_Z^2 + \sigma_Z^2) & \dots & \sigma_R^2 (j \mu_Z^2 + \sigma_Z^2 [1 + \sum_{l=1}^{j-1} \rho_l]) & \dots & \sigma_R^2 (k \mu_Z^2 + \sigma_Z^2 [1 + \sum_{l=1}^{j-1} \rho_l]) \\ \vdots & \vdots & \ddots & \vdots & \dots & \dots \\ j \sigma_R^2 \mu_Z & \sigma_R^2 (j \mu_Z^2 + \sigma_Z^2 [1 + \sum_{l=1}^{j-1} \rho_l]) & \vdots & \sigma_R^2 (j^2 \mu_Z^2 + \sigma_Z^2 [j + 2 \sum_{l=1}^{j-1} (j-l) \rho_l]) & \dots & \sigma_R^2 \left(j^K \mu_Z^2 + \sigma_Z^2 \left[\begin{matrix} j + \sum_{l=1}^{j-1} (j-l) \\ \times [\rho_l + \rho_{j-j+l}] + \sum_{l=1}^{j-j} j \rho_l \end{matrix} \right] \right) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ k \sigma_R^2 \mu_Z & \sigma_R^2 (k \mu_Z^2 + \sigma_Z^2 [1 + \sum_{l=1}^{j-1} \rho_l]) & \vdots & \sigma_R^2 \left(j^K \mu_Z^2 + \sigma_Z^2 \left[\begin{matrix} j + \sum_{l=1}^{j-1} (j-l) \\ \times [\rho_l + \rho_{j-j+l}] + \sum_{l=1}^{j-j} j \rho_l \end{matrix} \right] \right) & \vdots & \sigma_R^2 (k^2 \mu_Z^2 + \sigma_Z^2 [k + 2 \sum_{l=1}^{j-1} (j-l) \rho_l]) \end{bmatrix}$$

- Long story short, for the typical 1- through 5-year horizons and for $\rho = 0.953$ the covariance matrix of slope estimates under the null of no predictability is given by

$$Tcov(\hat{\beta}_1, \dots, \hat{\beta}_5) = \frac{\sigma_R^2}{\sigma_Z^2} \begin{pmatrix} 1 & 0.988 & 0.974 & 0.959 & 0.945 \\ & 1 & 0.993 & 0.982 & 0.970 \\ & & 1 & 0.995 & 0.986 \\ & & & 1 & 0.996 \\ & & & & 1 \end{pmatrix}$$

Incorporating serial correlation

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

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

- Therefore, we must limit the number of terms.
- More terms means more ability to pick up autocorrelation if there is any.
- But this comes at the cost of losing efficiency in finite samples.

The covariance matrix of the slope estimates

- Recall the regression estimates based on either the GMM or OLS are given by

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

- Considering the base case of no serial correlation (SC) and no heteroscedasticity (HS), we have

$$\text{Var}(\hat{\beta}|X) = (X'X)^{-1}X' \text{Var}[U|X] X(X'X)^{-1}$$

$$\text{Var}[U|X] = \sigma^2 I_T$$

- And the variance is estimated by:

$$\widehat{\text{Var}}(\hat{\beta}|X) = (X'X)^{-1}\hat{\sigma}^2$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2$$

- Now let us assume the presence of HS

$$\widehat{\text{Var}}(\hat{\beta}|X) = (X'X)^{-1}\widehat{\text{Var}}[X'U|X] (X'X)^{-1}$$

$$\widehat{\text{Var}}[X'U|X] = \sum_{t=1}^T x_t x_t' \hat{u}_t^2$$

$$\widehat{\text{Var}}(\hat{\beta}|X) = (X'X)^{-1}(\sum_{t=1}^T x_t x_t' \hat{u}_t^2) (X'X)^{-1}$$

- Let us assume now that there is serial correlation only
- In particular the error term obeys the $AR(1)$ process

$$y_t = x_t' \beta + u_t$$

$$u_t = \rho u_{t-1} + e_t$$

where

$$\text{Var}(e_t) = \sigma^2 \quad \forall t$$

- Therefore

$$\text{Var}(u_t) = \frac{\sigma^2}{1-\rho^2} \quad \forall t$$

- Then we get:

$$\text{Var}(\hat{\beta}) = [(X'X)^{-1} X' E(UU') X (X'X)^{-1}]$$

$$E(UU') = E \begin{bmatrix} u_1^2 & u_1 u_2 & \dots & u_1 u_T \\ u_1 u_2 & u_2^2 & \dots & u_2 u_T \\ \vdots & \vdots & & \vdots \\ u_1 u_T & \dots & & u_T^2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma^2}{1-\rho^2} & \frac{\rho\sigma^2}{1-\rho^2} & \dots & \dots \\ \frac{\rho\sigma^2}{1-\rho^2} & \frac{\sigma^2}{1-\rho^2} & \frac{\rho\sigma^2}{1-\rho^2} & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\rho^{T-1}\sigma^2}{1-\rho^2} & \vdots & \dots & \dots \end{bmatrix}$$

$$= \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(T-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(T-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(T-1) & \gamma(T-2) & \dots & \dots & \gamma(0) \end{bmatrix}$$

$$= \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & \dots & 1 \end{bmatrix}$$

$$= \frac{\sigma^2}{1-\rho^2} I_T + \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 0 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 0 & \rho & \dots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & \dots & 0 \end{bmatrix}$$

► Let

$$\bar{\rho} \equiv \begin{bmatrix} 0 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 0 & \rho & \dots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & \dots & 0 \end{bmatrix}$$

► The co-variance matrix is eventually given by

$$Var(\hat{\beta}) = (X'X)^{-1} \frac{\sigma^2}{1-\rho^2} + (X'X)^{-1} X' \left(\frac{\sigma^2}{1-\rho^2} \bar{\rho} \right) X (X'X)^{-1}$$

- Notice that with zero autocorrelation we are back with the base-case covariance matrix.
- Assume now that the error term obeys the $AR(2)$ process

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + e_t$$

$$Var(e_t) = \sigma^2 \quad \forall t$$

$$E(UU') = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(T-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(T-2) \\ \vdots & \vdots & & & \vdots \\ \gamma(T-1) & \gamma(T-2) & & \dots & \gamma(0) \end{bmatrix}$$

- $\gamma(0) = \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)(1-\phi_1-\phi_2)(1+\phi_1-\phi_2)}$

- $\gamma(1) = \gamma(0) \frac{\phi_1}{1-\phi_2}$

- $\gamma(2) = \gamma(0) \left(\frac{\phi_1^2}{1-\phi_2} + \phi_2 \right)$

Newey and West (1987)

- If both HS and SC, Andrews (1991) develops a complex, albeit useful, estimator.
- We focus here on Newey and West (1987) who propose the following covariance matrix

$$\widehat{Var}[X'U|X] = \sum_{t=1}^T x_t x_t' \hat{u}_t^2 + \sum_{j=1}^k \frac{k-j}{k} \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j} (x_t x_{t-j}' + x_{t-j} x_t')$$

- E.g.,

- $k=1$

$$\widehat{Var}[X'U|X] = \sum_{t=1}^T x_t x_t' \hat{u}_t^2$$

- $k=2$

$$\widehat{Var}[X'U|X] = \sum_{t=1}^T x_t x_t' \hat{u}_t^2 + \frac{1}{2} \sum_{t=2}^T \hat{u}_t \hat{u}_{t-1} (x_t x_{t-1}' + x_{t-1} x_t')$$

- $k=3$

$$\widehat{Var}[X'U|X] = \sum_{t=1}^T x_t x_t' \hat{u}_t^2 + \frac{2}{3} \sum_{t=2}^T \hat{u}_t \hat{u}_{t-1} (x_t x_{t-1}' + x_{t-1} x_t') + \frac{1}{3} \sum_{t=3}^T \hat{u}_t \hat{u}_{t-2} (x_t x_{t-2}' + x_{t-2} x_t')$$

Bayesian Econometrics

Bayes Rule

- Let x and y be two random variables
- Let $P(x)$ and $P(y)$ be the two *marginal* probability distribution functions of x and y
- Let $P(x|y)$ and $P(y|x)$ denote the corresponding *conditional* pdfs
- Let $P(x, y)$ denote the *joint* pdf of x and y
- It is known from the law of total probability that the joint pdf can be decomposed as

$$P(x, y) = P(x)P(y|x) = P(y)P(x|y)$$

- Therefore

$$\begin{aligned} P(y|x) &= \frac{P(y)P(x|y)}{P(x)} \\ &= cP(y)P(x|y) \end{aligned}$$

where c is the constant of integration (see next page)

- The Bayes Rule is described by the following proportion

$$P(y|x) \propto P(y)P(x|y)$$

Bayes Rule

- Notice that the right hand side retains only factors related to y , thereby excluding $P(x)$
- $P(x)$, termed the marginal likelihood function, is

$$\begin{aligned} P(x) &= \int P(y)P(x|y)dy \\ &= \int P(x, y)dy \end{aligned}$$

as the conditional distribution $P(y|x)$ integrates to unity.

- The marginal likelihood $P(x)$ is an essential ingredient in computing an important quantity - model posterior probability.
- Notice from the second equation above that the marginal likelihood obtains by integrating out y from the joint density $P(x, y)$.
- Similarly, if the joint distribution is $P(x, y, z)$ and the pdf of interest is $P(x, y)$ one integrates $P(x, y, z)$ with respect to z .

Bayes Rule

- The essence of Bayesian econometrics is the Bayes Rule.
- Ingredients of Bayesian econometrics are parameters underlying a given model, the sample data, the prior density of the parameters, the likelihood function describing the data, and the posterior distribution of the parameters.
- A predictive distribution could also be involved.
- In the Bayesian setup, parameters are stochastic while in the classical (non Bayesian) approach parameters are unknown constants.
- Decision making is based on the posterior distribution of the parameters or the predictive distribution of next period quantities as described below.
- On the basis of the Bayes rule, in what follows, y stands for unknown stochastic parameters, x for the data, $P(y|x)$ for the posterior distribution, $P(y)$ for the prior, and $P(x|y)$ for the likelihood.
- The Bayes rule describes the relation between the prior, the likelihood, and the posterior, or put differently it shows how prior beliefs are updated to produce posterior beliefs:
$$P(y|x) \propto P(y)P(x|y)$$
- Zellner (1971) is an excellent source of reference.

Bayes Econometrics in Financial Economics

- You observe the returns on the market index over T months: r_1, \dots, r_T
- Let $R: [r_1, \dots, r_T]'$ denote the $T \times 1$ vector of all return realizations
- Assume that $r_t \sim N(\mu, \sigma_0^2)$ for $t = 1, \dots, T$

where

μ is a stochastic random variable denoting the mean return

σ_0^2 is the variance which, at this stage, is assumed to be a known constant

and returns are IID (independently and identically distributed) through time.

- By Bayes rule

$$P(\mu|R, \sigma_0^2) \propto P(\mu)P(R|\mu, \sigma_0^2)$$

where

$P(\mu|R, \sigma_0^2)$ is the posterior distribution of μ

$P(\mu)$ is the prior distribution of μ

and $P(R|\mu, \sigma_0^2)$ is the joint likelihood of all return realizations.

Bayes Econometrics: Likelihood

- The likelihood function of a normally distributed return realization is given by

$$P(r_t|\mu, \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2}(r_t - \mu)^2\right)$$

- Since returns are assumed to be IID, the joint likelihood of all realized returns is

$$P(R|\mu, \sigma_0^2) = (2\pi\sigma_0^2)^{-\frac{T}{2}} \exp\left(-\frac{1}{2\sigma_0^2} \sum_{t=1}^T (r_t - \mu)^2\right)$$

- Notice:

$$\begin{aligned} \sum (r_t - \mu)^2 &= \sum [(r_t - \hat{\mu}) + (\hat{\mu} - \mu)]^2 \\ &= v s^2 + T(\mu - \hat{\mu})^2 \end{aligned}$$

since the cross product is zero, and

$$\begin{aligned} v &= T - 1 \\ s^2 &= \frac{1}{T - 1} \sum (r_t - \hat{\mu})^2 \end{aligned}$$

$$\hat{\mu} = \frac{1}{T} \sum r_t$$

Prior

- The prior is specified by the researcher based on economic theory, past experience, past data, current similar data, etc. Often, the prior is diffuse or non-informative
- For the next illustration, it is assumed that $P(\mu) \propto c$, that is, the prior is diffuse, non-informative, in that it *apparently* conveys no information on the parameters of interest.
- I emphasize “apparently” since innocent diffuse priors could exert substantial amount of information about quantities of interest which are non-linear functions of the parameters.
- Informative priors with sound economic appeal are well perceived in financial economics.
- For instance, Kandel and Stambaugh (1996), who study asset allocation when stock returns are predictable, entertain informative prior beliefs weighted against predictability. Pastor and Stambaugh (1999) introduce prior beliefs about expected stock returns which consider factor model restrictions. Avramov, Cederburg, and Kvasnakova (2017) study prior beliefs about predictive regression parameters which are disciplined by consumption based asset pricing models including habit formation, prospect theory, and long run risk.
- Computing posterior probabilities (as opposed to posterior densities) of competing models (e.g., Avramov (2002)) necessitates the use of informative priors. Diffuse priors won't fit.

The Posterior Distribution of Mean Return

- ▶ With diffuse prior and normal likelihood, the posterior is proportional to the likelihood function:

$$\begin{aligned} P(\mu|R, \sigma_0^2) &\propto \exp\left(-\frac{1}{2\sigma_0^2} [vs^2 + T(\mu - \hat{\mu})]^2\right) \\ &\propto \exp\left(-\frac{T}{2\sigma_0^2} (\mu - \hat{\mu})^2\right) \end{aligned}$$

- ▶ The bottom relation follows since only factors related to μ are retained
- ▶ The posterior distribution of the mean return is given by

$$\mu|R, \sigma_0^2 \sim N\left(\hat{\mu}, \sigma_0^2/T\right)$$

- ▶ In classical econometrics:

$$\hat{\mu}|R, \sigma_0^2 \sim N\left(\mu, \sigma_0^2/T\right)$$

- ▶ That is, in classical econometrics, the sample estimate of μ is stochastic while μ itself is an unknown non-stochastic parameter.

Informative Prior

- The prior on the mean return is often modeled as

$$P(\mu) \propto (\sigma_a)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_a^2}(\mu - \mu_a)^2\right)$$

where μ_a and σ_a are prior parameters to be specified by the researcher

- The posterior obtains by combining the prior and the likelihood:

$$\begin{aligned} P(\mu|R, \sigma_0^2) &\propto P(\mu)P(R|\mu, \sigma_0^2) \\ &\propto \exp\left[-\frac{(\mu - \mu_a)^2}{2\sigma_a^2} - \frac{T(\mu - \hat{\mu})^2}{2\sigma_0^2}\right] \\ &\propto \exp\left(-\frac{1}{2} \frac{(\mu - \tilde{\mu})^2}{\tilde{\sigma}^2}\right) \end{aligned}$$

- The bottom relation obtains by **completing the square** on μ
- Notice, in particular,

$$\begin{aligned} \frac{\mu^2}{\sigma_a^2} + \frac{T}{\sigma_0^2} \mu^2 &= \frac{\mu^2}{\tilde{\sigma}^2} \\ \left(\frac{1}{\sigma_a^2} + \frac{T}{\sigma_0^2}\right) &= \frac{1}{\tilde{\sigma}^2} \end{aligned}$$

The Posterior Mean

- ▶ Hence, the posterior variance of the mean is

$$\begin{aligned}\tilde{\sigma}^2 &= \left[\frac{1}{\sigma_a^2} + \frac{1}{\sigma_0^2/T} \right]^{-1} \\ &= (\text{prior precision} + \text{likelihood precision})^{-1}\end{aligned}$$

- ▶ Similarly, the posterior mean of μ is

$$\begin{aligned}\tilde{\mu} &= \tilde{\sigma}^2 \left[\frac{\mu_a}{\sigma_a^2} + \frac{T\hat{\mu}}{\sigma_0^2} \right] \\ &= w_1\mu_a + w_2\hat{\mu}\end{aligned}$$

where

$$\begin{aligned}w_1 &= \frac{\frac{1}{\sigma_a^2}}{\frac{1}{\sigma_a^2} + \frac{1}{\sigma_0^2/T}} = \frac{\text{prior precision}}{\text{prior precision} + \text{likelihood precision}} \\ w_2 &= 1 - w_1\end{aligned}$$

- ▶ Intuitively, the posterior mean of μ is the weighted average of the prior mean and the sample mean with weights depending on prior and likelihood precisions, respectively.

What if σ is unknown? – The case of Diffuse Prior

► Bayes: $P(\mu, \sigma|R) \propto P(\mu, \sigma)P(R|\mu, \sigma)$

► The non-informative prior is typically modeled as

$$P(\mu, \sigma) \propto P(\mu)P(\sigma)$$

$$P(\mu) \propto c$$

$$P(\sigma) \propto \sigma^{-1}$$

► Thus, the joint posterior of μ and σ is

$$P(\mu, \sigma|R) \propto \sigma^{-(T+1)} \exp\left(-\frac{1}{2\sigma^2} [vs^2 + T(\mu - \hat{\mu})^2]\right)$$

► The conditional distribution of the mean follows straightforwardly

$$P(\mu|\sigma, R) \text{ is } N\left(\hat{\mu}, \sigma^2/T\right)$$

► More challenging is to uncover the marginal distributions, which are obtained as

$$P(\mu|R) = \int P(\mu, \sigma|R) d\sigma$$

$$P(\sigma|R) = \int P(\mu, \sigma|R) d\mu$$

Solving the Integrals: Posterior of μ

➤ Let $\alpha = \nu s^2 + T(\mu - \hat{\mu})^2$

➤ Then,

$$P(\mu|R) \propto \int_{\sigma=0}^{\infty} \sigma^{-(T+1)} \exp\left(-\frac{\alpha}{2\sigma^2}\right) d\sigma$$

➤ We do a change of variable

$$\begin{aligned} x &= \frac{\alpha}{2\sigma^2} \\ \frac{d\sigma}{dx} &= -2^{-1} \frac{1}{2} \alpha^{\frac{1}{2}} x^{-1\frac{1}{2}} \\ \sigma^{-T+1} &= \left(\frac{\alpha}{2x}\right)^{-\frac{T+1}{2}} \end{aligned}$$

➤ Then

$$P(\mu|R) \propto 2^{\frac{T-2}{2}} \alpha^{-\frac{T}{2}} \int_{x=0}^{\infty} x^{\frac{T}{2}-1} \exp(-x) dx$$

➤ Notice

$$\int_{x=0}^{\infty} x^{\frac{T}{2}-1} \exp(-x) dx = \Gamma\left(\frac{T}{2}\right)$$

➤ Therefore,

$$\begin{aligned} P(\mu|R) &\propto 2^{\frac{T-2}{2}} \Gamma\left(\frac{T}{2}\right) \alpha^{-\frac{T}{2}} \\ &\propto [\nu s^2 + T(\mu - \hat{\mu})^2]^{-\frac{\nu+1}{2}} \end{aligned}$$

➤ We get $t = \frac{\mu - \hat{\mu}}{s/\sqrt{T}} \sim t(\nu)$, corresponding to the Student t distribution with ν degrees of freedom.

The Marginal Posterior of σ

- The posterior on σ

$$\begin{aligned} P(\sigma|R) &\propto \int \sigma^{-(T+1)} \exp\left(-\frac{1}{2\sigma^2} [vs^2 + T(\mu - \hat{\mu})]^2\right) d\mu \\ &\propto \sigma^{-(T+1)} \exp\left(-\frac{vs^2}{2\sigma^2}\right) \int \exp\left(-\frac{T}{2\sigma^2} (\mu - \hat{\mu})^2\right) d\mu \end{aligned}$$

- Let $z = \frac{\sqrt{T}(\mu - \hat{\mu})}{\sigma}$, then

$$\frac{dz}{d\mu} = \sqrt{T} \frac{1}{\sigma}$$

- $$\begin{aligned} P(\sigma|R) &\propto \sigma^{-T} \exp\left(-\frac{vs^2}{2\sigma^2}\right) \int \exp(-z^2/2) dz \\ &\propto \sigma^{-T} \exp\left(-\frac{vs^2}{2\sigma^2}\right) \\ &\propto \sigma^{-(v+1)} \exp\left(-\frac{vs^2}{2\sigma^2}\right) \end{aligned}$$

which corresponds to the inverted gamma distribution with v degrees of freedom and parameter s

- The explicit form (with constant of integration) of the inverted gamma is given by

$$P(\sigma|v, s) = \frac{2}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{vs^2}{2}\right)^{v/2} \sigma^{-(v+1)} \exp\left(-\frac{vs^2}{2\sigma^2}\right)$$

The Multiple Regression Model

- The regression model is given by

$$y = X\beta + u$$

where

y is a $T \times 1$ vector of the dependent variables

X is a $T \times M$ matrix with the first column being a $T \times 1$ vector of ones

β is an $M \times 1$ vector containing the intercept and $M-1$ slope coefficients

and u is a $T \times 1$ vector of residuals.

- We assume that $u_t \sim N(0, \sigma^2) \forall t = 1, \dots, T$ and IID through time
- The likelihood function is

$$\begin{aligned} P(y|X, \beta, \sigma) &\propto \sigma^{-T} \exp\left(-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta)\right) \\ &\propto \sigma^{-T} \exp\left\{-\frac{1}{2\sigma^2} \left[vs^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})\right]\right\} \end{aligned}$$

The Multiple Regression Model

where

$$v = T - M$$

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$s^2 = \frac{1}{v} (y - X\hat{\beta})' (y - X\hat{\beta})$$

► It follows since

$$\begin{aligned} (y - X\beta)'(y - X\beta) &= [y - X\hat{\beta} - X(\beta - \hat{\beta})]'[y - X\hat{\beta} - X(\beta - \hat{\beta})] \\ &= (y - X\hat{\beta})'(y - X\hat{\beta}) + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \end{aligned}$$

while the cross product is zero

Assuming Diffuse Prior

- The prior is modeled as

$$P(\beta, \sigma) \propto \frac{1}{\sigma}$$

- Then the joint posterior of β and σ is

$$P(\beta, \sigma | y, X) \propto \sigma^{-(T+1)} \exp \left\{ -\frac{1}{2\sigma^2} \left[v s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right] \right\}$$

- The conditional posterior of β is

$$P(\beta | \sigma, y, X) \propto \exp \left(-\frac{1}{2\sigma^2} (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right)$$

which obeys the multivariable normal distribution

$$N(\hat{\beta}, (X' X)^{-1} \sigma^2)$$

Assuming Diffuse Prior

- What about the marginal posterior for β ?

$$P(\beta|y, X) = \int P(\beta, \sigma|y, X) d\sigma$$
$$\propto \left[\nu s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right]^{-T/2}$$

which pertains to the multivariate student t with mean $\hat{\beta}$ and $T-M$ degrees of freedom

- What about the marginal posterior for σ ?

$$P(\sigma|y, X) = \int P(\beta, \sigma|y, X) d\beta$$
$$\propto \sigma^{-(\nu+1)} \exp\left(-\frac{\nu s^2}{2\sigma^2}\right)$$

which stands for the inverted gamma with $T-M$ degrees of freedom and parameter s

- You can simulate the distribution of β in two steps without solving analytically the integral, drawing first σ from its inverted gamma distribution and then drawing from the conditional of β given σ which is normal as shown earlier. This mechanism generates draws from the Student t distribution

Bayesian Updating/Learning

- Suppose the initial sample consists of T_1 observations of X_1 and y_1 .
- Suppose further that the posterior distribution of (β, σ) based on those observations is given by:

$$\begin{aligned} P(\beta, \sigma | y_1, X_1) &\propto \sigma^{-(T_1+1)} \exp \left[-\frac{1}{2\sigma^2} (y_1 - X_1\beta)'(y_1 - X_1\beta) \right] \\ &\propto \sigma^{-(T_1+1)} \exp \left\{ -\frac{1}{2\sigma^2} [v_1 s_1^2 + (\beta - \hat{\beta}_1)' X_1' X_1 (\beta - \hat{\beta}_1)] \right\} \end{aligned}$$

where

$$\begin{aligned} v_1 &= T_1 - M \\ \hat{\beta}_1 &= (X_1' X_1)^{-1} X_1' y_1 \\ v_1 s_1^2 &= (y_1 - X_1 \hat{\beta}_1)' (y_1 - X_1 \hat{\beta}_1) \end{aligned}$$

- You now observe one additional sample X_2 and y_2 of length T_2 observations
- The likelihood based on that sample is

$$P(y_2, X_2 | \beta, \sigma) \propto \sigma^{-T_2} \exp \left[-\frac{1}{2\sigma^2} (y_2 - X_2\beta)'(y_2 - X_2\beta) \right]$$

Bayesian Updating/Learning

- Combining the posterior based on the first sample (which becomes the prior for the second sample) and the likelihood based on the second sample yields:

$$\begin{aligned} P(\beta, \sigma | y_1, y_2, X_1, X_2) &\propto \sigma^{-(T_1+T_2+1)} \exp \left\{ -\frac{1}{2\sigma^2} [(y_1 - X_1\beta)'(y_1 - X_1\beta) + (y_2 - X_2\beta)'(y_2 - X_2\beta)] \right\} \\ &\propto \sigma^{-(T_1+T_2+1)} \exp \left\{ -\frac{1}{2\sigma^2} [v s^2 + (\beta - \tilde{\beta})' \mathfrak{M} (\beta - \tilde{\beta})] \right\} \end{aligned}$$

where

$$\begin{aligned} \mathfrak{M} &= X_1'X_1 + X_2'X_2 \\ \tilde{\beta} &= \mathfrak{M}^{-1}(X_1'y_1 + X_2'y_2) \\ v s^2 &= (y_1 - X_1\tilde{\beta})'(y_1 - X_1\tilde{\beta}) + (y_2 - X_2\tilde{\beta})'(y_2 - X_2\tilde{\beta}) \\ v &= T_1 + T_2 - M \end{aligned}$$

- Then the posterior distributions for β and σ follow using steps outlined earlier
- With more observations realized you follow the same updating procedure
- Notice that the same posterior would have been obtained starting with diffuse priors and then observing the two samples consecutively $Y=[y_1', y_2']'$ and $X = [X_1', X_2']'$.

Example: Predictive Regressions

- In finance and economics you often use predictive regressions of the form

$$y_{t+1} = a + b'z_t + u_{t+1}$$

$$u_{t+1} \sim N(0, \sigma^2) \forall t = 1, \dots, T - 1 \text{ and IID}$$

where y_{t+1} is an economic quantity of interest, be it stock or bond return, inflation, interest rate, exchange rate, and z_t is a collection of $M - 1$ predictive variables, e.g., the term spread

- At this stage, the initial observation of the predictors, z_0 , is assumed to be non stochastic.
- Stambaugh (1999) considers stochastic z_0 . Then, some complexities emerge as shown later.
- The predictive regression can be written more compactly as

$$y_{t+1} = x_t' \beta + u_{t+1}$$

where

$$x_t = [1, z_t]'$$

$$\beta = [a, b]'$$

- In a matrix form, comprising all time-series observations, the normal regression model obtains

$$y = X\beta + u$$

Predictive Distribution

- You are interested to uncover the predictive distribution of the unobserved y_{T+1}
- Let Φ denote the observed data and let θ denote the set of parameters β and σ^2
- The predictive distribution is:

$$P(y_{T+1}|\Phi) = \int_{\theta} P(y_{T+1}|\Phi, \theta) P(\theta|\Phi) d\theta$$

where

$P(y_{T+1}|\Phi, \theta)$ is the conditional or classical predictive distribution

$P(\theta|\Phi)$ is the joint posterior of β and σ^2

- Notice that the predictive distribution integrates out β and σ from the joint distribution

$$P(y_{T+1}, \beta, \sigma|\Phi)$$

since

$$P(y_{T+1}, \beta, \sigma|\Phi) = P(y_{T+1}|\Phi, \theta) P(\theta|\Phi)$$

Predictive Distribution

- The conditional distribution of the next period realization is

$$P(y_{T+1}|\theta, \Phi) \propto \sigma^{-1} \exp \left[-\frac{1}{2\sigma^2} (y_{T+1} - x_T' \beta)^2 \right]$$

- Thus $P(y_{T+1}, \beta, \sigma | \Phi)$ is proportional to $\sigma^{-(T+2)} \exp \left[-\frac{1}{2\sigma^2} [(y - X\beta)'(y - X\beta) + (y_{T+1} - x_T' \beta)^2] \right]$
- On integrating $P(y_{T+1}, \beta, \sigma | \Phi)$ with respect to σ we obtain

$$P(y_{T+1}, \beta | \Phi) \propto [(y - X\beta)'(y - X\beta) + (y_{T+1} - x_T' \beta)^2]^{-\frac{(T+1)}{2}}$$

- Now we have to integrate with respect to the M elements of β
- On completing the square on β we get

$$\begin{aligned} & (y - X\beta)'(y - X\beta) + (y_{T+1} - x_T' \beta)^2 \\ &= y'y + y_{T+1}^2 + \beta' \mathfrak{M} \beta - 2\beta'(X'y + x_T'y_{T+1}) \\ &= y'y + y_{T+1}^2 - (y'X + y_{T+1}'x_T) \mathfrak{M}^{-1} (X'y + x_T'y_{T+1}) \\ &+ [\beta - \mathfrak{M}^{-1} (X'y + x_T'y_{T+1})]' \mathfrak{M} (\beta - \mathfrak{M}^{-1} (X'y + x_T'y_{T+1})) \end{aligned}$$

where

$$\mathfrak{M} = X'X + x_T'x_T$$

Predictive Distribution

- Integrating with respect to β yields

$$P(y_{T+1}|\Phi) \propto [y'y + y_{T+1}^2 - (y'X + y_{T+1}'x_T)\mathfrak{M}^{-1}(X'y + x_T'y_{T+1})]^{-\frac{\nu+1}{2}}$$

where

$$\nu = T - M$$

- With some further algebra it can be shown that the predictive distribution is

$$P(y_{T+1}|\Phi) \propto [\nu + (y_{T+1} - x_T'\hat{\beta})H(y_{T+1} - x_T'\hat{\beta})]^{-\frac{\nu+1}{2}}$$

where

$$H = \frac{1}{s^2} (1 - x_T'\mathfrak{M}^{-1}x_T)$$

$$\nu s^2 = (y - X\hat{\beta})'(y - X\hat{\beta})$$

$$\hat{\beta} = (X'X)^{-1}X'y$$

Predictive Moments

- The first and second predictive moments, based on the t-distribution, are

$$\begin{aligned}\mu_y &= E(y_{T+1}|\Phi) = x_T' \hat{\beta} \\ E(y_{T+1} - \mu_y)^2 &= \frac{v}{v-2} H^{-1} \\ &= \frac{vs^2}{v-2} (1 - x_T' \mathbf{\Pi}^{-1} x_T)^{-1} \\ &= \frac{vs^2}{v-2} (1 + x_T' (X'X) x_T)\end{aligned}$$

- With diffuse prior, the predictive mean coincides with the classical (non Bayesian) mean.
- The predictive variance is slightly higher due to estimation risk.
- Kandel and Stambaugh (JF 1996) provide more economic intuition about the predictive density
- The estimation risk effect on the predictive variance is analytically derived by Avramov and Chordia (JFE 2006) in a multi-asset (asset pricing) context.
- Later, we will use the predictive distribution to recover asset allocation under estimation risk and even under model uncertainty considering informative priors.

Multivariate Regression Models

- Consider the multivariate form (N dependent variables) of the predictive regression

$$R = XB + U$$

where R and U are both a $T \times N$ matrix, X is a $T \times M$ matrix, B is an $M \times N$ matrix

$$\text{vec}(U) \sim N(0, \Sigma \otimes I_T)$$

and where vec denotes the vectorization operator and \otimes is the kronecker product

- The priors for B and Σ are assumed to be the normal inverted Wishart (conjugate priors)

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

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

Multivariate regression

where

$$b = \text{vec}(B)$$

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

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

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

where tr stands for the trace operator. This can be rewritten in a more convenient form as

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

where

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

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

Multivariate regression

- An equivalent representation for the likelihood function is given by

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

where

$$\hat{b} = \text{vec}(\hat{B})$$

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

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

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

Multivariate regression

where

$$\tilde{\Psi} = \Psi_0 + X'X$$

$$\tilde{b} = \text{vec}(\tilde{B})$$

$$\tilde{B} = \tilde{\Psi}^{-1}X'X\hat{B} + \tilde{\Psi}^{-1}\Psi_0B_0$$

$$\tilde{S} = \hat{S} + S_0 + \hat{B}'X'X\hat{B} + B_0'\Psi_0B_0 - \tilde{B}'\tilde{\Psi}\tilde{B}$$

$$\tilde{\nu} = \nu_0 + T$$

- So the posterior for B is normal and for Σ is inverted Wishart.
- Again,, that is the conjugate prior idea - the prior and posterior have the same distributions but with different parameters.
- Not surprisingly, \tilde{B} is a weighted average of B_0 and \hat{B} :

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

where $W = I - \tilde{\Psi}^{-1}X'X$. Notice, the weights are represented by matrices.

What if the posterior does not obey a well-known expression?

- Thus far, the posterior densities can readily be identified.
- However, what if the posterior does not obey a well known expression?
- Markov Chain Monte Carlo (MCMC) methods can be employed to simulate from the posterior.
- The basic intuition behind MCMC is straightforward.
- Suppose the distribution is $P(x)$ which is unrecognized.
- The MCMC idea is to define a Markov chain over possible values of x (x_0, x_1, x_2, \dots) such that as $n \rightarrow \infty$, we can guarantee that $x_n \sim P(x)$, that is, that we have a draw from the posterior.
- As the number of draws (each draw pertains to a distinct chain) gets larger you can simulate the posterior density.
- The simulation gets more precise with increasing number of draws.
- There are various ways to set up such Markov chains
- Here, we cover two MCMC methods: the Gibbs Sampling and the Metropolis Hastings.

Gibbs Sampling

- The Gibbs Sampling analysis is based on “Measuring the Pricing Error of the Arbitrage Pricing Theory” by Geweke and Zhou (RFS 1996).
- This paper advocates a Bayesian method in which to test the APT of Ross (1976).
- Both APT and ICAPM motivate multiple factors – extending the CAPM.
- While APT motivates statistical based factors, as shown below, the ICAPM motivates economic factors related to the marginal utility of the investor – such as consumption growth.
- The basic APT model assumes that returns on N risky portfolios are related to K pervasive unknown factors ($K < N$).
- The relation is described by the K factor model

$$r_t = \mu + \beta f_t + \epsilon_t$$

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

Gibbs Sampling

- Specifically,

$$E\{f_t\} = 0$$

$$E\{f_t f'_t\} = I_K$$

$$E\{\epsilon_t | f_t\} = 0$$

$$E\{\epsilon_t \epsilon'_t | f_t\} = \Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_N^2)$$

$$\beta = [\beta_1, \dots, \beta_K]$$

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

$$\mu = \lambda_0 + \beta_1 \lambda_1 + \dots + \beta_K \lambda_K$$

- Notice that λ_0 is the component of expected return unrelated to factor exposures.
- The original APT model is about an approximated relation.
- An exact version is derived by Huberman (1982) among others.
- The objective throughout is to explore a measure that summarizes the deviation from exact pricing.

Gibbs Sampling

- ▶ That measure is denoted by Q^2 and is given by

$$Q^2 = \frac{1}{N} \mu' [I_N - \beta^* (\beta'^* \beta^*)^{-1} \beta^{*'}] \mu$$

where

$$\beta^* = [1_N, \beta]$$

- ▶ Recovering the sampling distribution of Q^2 is hopeless.
- ▶ Notice that one cannot even recover an analytic expression for the posterior density of model parameters $P(\theta|R)$. $P(\theta|R, F)$ is something known – but this is not the posterior.
- ▶ However, using Gibbs sampling, we can simulate the posterior distribution of Q^2 as well as simulate the posterior density of all parameters and latent factors.
- ▶ In what follows, we assume that observed returns and latent factors are jointly normally distributed:

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

Gibbs Sampling

- Here are some additional notations:
 - Data: $R = [r_1', \dots, r_T']'$
 - Parameters: $\Theta = [\mu', \text{vec}(\beta)', \text{vech}(\Sigma)']'$ where *vech* denotes the distinct elements of the matrix
 - Latent factors: $f = [f_1', \dots, f_T']'$
- To evaluate the pricing error we need to simulate draws from the posterior distribution $P(\Theta|R)$.
- We draw from the joint posterior in a slightly different manner than that suggested in the Geweke-Zhou paper.
- First, we employ a multivariate regression setting. Moreover, the well-known identification (of the factors) problem is not accounted for to simplify the analysis.
- The prior on the diagonal covariance matrix is assumed to be non-informative

$$P_0(\Theta) \propto |\Sigma|^{-\frac{1}{2}} = (\sigma_1 \dots \sigma_N)^{-1}$$

Gibbs Sampling

- Re-expressing the arbitrage pricing equation, we obtain:

$$r'_t = F'_t B' + \epsilon'_t$$

where

$$F'_t = [1, f'_t]$$

$$B = [\mu, \beta]$$

- Rewriting the system in a matrix notation, we get

$$R = FB' + E$$

- Why do we need to use the Gibbs sampling technique?
- Because the likelihood function $P(R|\Theta)$ (and therefore the posterior density) cannot be expressed analytically.

Gibbs Sampling

- However, $P(R|\Theta, F)$ does obey an analytical form:

$$P(R|\Theta, F) \propto |\Sigma|^{-\frac{T}{2}} \exp \left[-\frac{1}{2} \text{tr}\{[R - FB']'[R - FB']\Sigma^{-1}\} \right]$$

- Therefore, we can compute the full conditional posterior densities:

$$P(B|\Sigma, F, R)$$

$$P(\Sigma|B, F, R)$$

$$P(F|B, \Sigma, R)$$

- The Gibbs sampling chain is formed as follows:

1. Specify starting values $\Sigma^{(0)}$ and $F^{(0)}$ and set $i = 1$.

2. Draw from the full conditional distributions:

- Draw $B^{(i)}$ from $P(B|\Sigma^{(i-1)}, F^{(i-1)}, R)$

- Draw $\Sigma^{(i)}$ from $P(\Sigma|B^{(i)}, F^{(i-1)}, R)$

- Draw $F^{(i)}$ from $P(F|B^{(i)}, \Sigma^{(i)}, R)$

3. Set $i = i + 1$ and go to step 2.

Gibbs Sampling

- After m iterations the sample $B^{(m)}, \Sigma^{(m)}, F^{(m)}$ is obtained.
- Under mild regularity conditions (see, for example, Tierney, 1994), $(B^{(m)}, \Sigma^{(m)}, F^{(m)})$ converges in distribution to the relevant marginal and joint distributions:

$$\begin{aligned}P(B^{(m)}|R) &\rightarrow P(B|R) \\P(\Sigma^{(m)}|R) &\rightarrow P(\Sigma|R) \\P(F^{(m)}|R) &\rightarrow P(F|R) \\P(B^{(m)}, \Sigma^{(m)}, F^{(m)}|R) &\rightarrow P(B, \Sigma, F|R)\end{aligned}$$

- For m (burn-in draws) large enough, the G values
 $(B^{(g)}, \Sigma^{(g)}, F^{(g)})_{g=m+1}^{m+G}$

are a sample from the joint posterior.

- What are the full conditional posterior densities?
- Note:

$$P(B|\Sigma, F, R) \propto \exp \left\{ -\frac{1}{2} \text{tr}[R - FB']'[R - FB']\Sigma^{-1} \right\}$$

Gibbs Sampling

► Let $b = \text{vec}(B')$. Then

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

where

$$\hat{b} = \text{vec}[(F'F)^{-1}F'R]$$

► Therefore,

$$b|\Sigma, F, R \sim N(\hat{b}, \Sigma \otimes (F'F)^{-1})$$

► Also note:

$$P(\sigma_i|B, F, R) \propto \sigma_i^{-(T+1)} \exp \left(-\frac{TS_i^2}{2\sigma_i^2} \right)$$

where TS_i^2 is the i -th diagonal element of the $N \times N$ matrix

$$[R - FB']'[R - FB']$$

Gibbs Sampling

suggesting that:

$$\frac{TS_i^2}{\sigma_i^2} \sim \chi^2(T)$$

➤ Finally,

$$f_t | \mu, \beta, \Sigma, r_t \sim N(M_t, H_t)$$

where

$$M_t = \beta'(\beta\beta' + \Sigma)^{-1}(r_t - \mu)$$

$$H_t = I_K - \beta'(\beta\beta' + \Sigma)^{-1}\beta$$

- Here, we are basically done with the GS implementation
- Having all the essential draws from the joint posterior at hand, we can analyze the simulated distribution of the pricing errors and make a call about model's pricing abilities.

Metropolis Hastings (MH)

- Indeed, Gibbs Sampling is intuitive, easy to implement, and convergence to the true posterior density is accomplished relatively fast and with mild regularity conditions.
- Often, however, integrable expressions for the full conditional densities are infeasible.
- Then it is essential to resort to more complex methods in which draws from the posterior distributions could be highly correlated and convergence could be rather slow. Still, such methods are useful.
- One example is the Metropolis Hastings (MH) algorithm, a MCMC procedure introduced by Metropolis et al (1953) and later generalized by Hastings (1970).
- The basic idea in MH is to make draws from a candidate distribution which seems to be related to the target (unknown) distribution.
- The candidate draw from the posterior is accepted with some probability – the Metropolis rule. Otherwise, it is rejected and the previous draw is retained. As in the Gibbs Sampling, the Markov Chain starts with some initial value set by the researcher.
- The Gibbs Sampling is a special case of MH in which all draws are accepted with probability one.

Metropolis Hastings (MH)

- I will display two applications of the MH method in the context of financial economics.
- The first application goes to the seminal work of Jacquier, Polson, and Rossi (1994) on estimating a stochastic volatility (SVOL) model. Coming up on the next page.
- The second application is based on Stambaugh (1999) who analyzes predictive regressions when the first observation of the predictive variable is stochastic.
- Mostly, analyses of predictive regressions are conducted based on the assumption that the first observation is fixed non-stochastic.
- While analytically tractable this assumption does not seem to hold true.
- Relaxing that assumption entertains several complexities and the need to use MH to draw from the joint posterior distribution of the predictive regression parameters.
- I will discuss that application in the section on asset allocation when stock returns are predictable.

Stochastic Volatility (SVOL)

- The SVOL model is given by

$$\begin{aligned}y_t &= \sqrt{h_t} u_t \\ \ln(h_t) &= \alpha + \delta \ln(h_{t-1}) + \sigma_v v_t \\ \begin{pmatrix} u_t \\ v_t \end{pmatrix} &\sim N(0, I_2)\end{aligned}$$

- Notice that volatility varies through time rather than being constant. While in ARCH, GARCH, EGARCH models there is no stochastic innovation, here volatility is stochastic.
- Now let

$$\begin{aligned}h' &= [h_1 \dots h_T] \\ \beta' &= [\alpha, \delta] \\ w' &= [\alpha, \delta, \sigma_v] \\ y' &= [y_1 \dots y_T]\end{aligned}$$

- The posterior of w given values of h is available from the standard regression model described earlier: β has the multivariable normal and σ_v has the inverted gamma distribution.
- However, drawing from $h|w, y$ requires more efforts.

Stochastic Volatility (SVOL)

- We cannot really draw h at once, rather, we have to break down the joint posterior of the entire h vector by considering the series of uni-variate conditional densities.

$$P(h_t|h_{t-1}, h_{t+1}, w, y_t) \text{ for } t = 1, \dots, T$$

- If it were possible to draw directly these uni-variate densities, the algorithm would reduce to a Gibbs sampler in which we would draw successively from $P(w|h, y)$ and then each of the T univariate conditionals in turn to form one step in the Markov chain.
- The uni-variate conditional densities, however, exhibit an unusual form:

$$\begin{aligned} &P(h_t|h_{t-1}, h_{t+1}, w, y_t) \\ &\propto P(y_t|h_t)P(h_t|h_{t-1})P(h_{t+1}|h_t) \\ &\propto h_t^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{y_t^2}{h_t}\right) \frac{1}{h_t} \exp[-(\ln(h_t) - \mu_t)^2 / 2\sigma^2] \end{aligned}$$

where

$$\begin{aligned} \mu_t &= [\alpha(1 - \delta) + \delta(\ln(h_{t+1}) + \ln(h_{t-1}))] / (1 + \delta^2) \\ \sigma^2 &= \frac{\sigma_v^2}{1 + \delta^2} \end{aligned}$$

Stochastic Volatility (SVOL)

- The result follows by combining two likelihood normal terms and completing the square on $\ln(h_t)$.
- Notice, the density is not of a standard form. It is proportional to:

$$\frac{\exp\left(-\frac{y_t^2}{2h_t} - \frac{1}{2\sigma^2}(\ln(h_t) - \mu)^2\right)}{(h_t)^{3/2}} \\ = \frac{\exp\left\{-\frac{y_t^2}{2h_t} - \frac{1}{2\sigma^2}[\ln(h_t) - (\mu - \sigma^2/2)]^2\right\}}{h_t}$$

- A good proposal here can be the lognormal density given by

$$\frac{1}{\sqrt{2\pi\sigma h_t}} \exp\left(-\frac{1}{2\sigma^2}[\ln(h_t) - \mu + \sigma^2/2]^2\right)$$

Stochastic Volatility (SVOL)

- From here we compute the ratio of the target to the proposal as:

$$\frac{\text{target}}{\text{proposal}} \propto \frac{\exp\left(-\frac{y_t^2}{2h_t} - \frac{1}{2\sigma^2} [\ln(h_t) - (\mu - \sigma^2/2)]^2\right)/h_t}{\exp\left(-\frac{1}{2\sigma^2} [\ln(h_t) - (\mu - \sigma^2/2)]^2\right)/h_t} \\ \propto \exp\left(-\frac{y_t^2}{2h_t}\right)$$

- For the MH algorithm, the relevant ratio is

$$\exp\left(\frac{y_t^2}{2h_t} - \frac{y_t^2}{2h_t^*}\right)$$

where h_t^* is the new proposed draw and h_t is the current state (or previously accepted draw).

Bayesian Portfolio Analysis

- We next present topics on Bayesian portfolio analysis based upon a review paper of Avramov and Zhou (2010) that came up in the *Annual Review of Financial Economics*
- We first study asset allocation when stock returns are assumed to be IID
- We then incorporate potential return predictability based on macro economy variables.
- What are the benefits of using the Bayesian approach?
- There are at least three important benefits including (i) the ability to account for estimation risk and model uncertainty, (ii) the feasibility of powerful and tractable simulation methods, and (iii) the ability to elicit economically meaningful prior beliefs about the distribution of future returns.

Bayesian Asset Allocation

- ▶ We start with the mean variance framework
- ▶ Assume there are $N + 1$ assets, one of which is riskless and others are risky.
- ▶ Denote by r_{ft} and r_t the rates of returns on the riskless asset and the risky assets at time t , respectively.
- ▶ Then

$$R_t \equiv r_t - r_{ft} \mathbf{1}_N$$

are excess returns on the N risky assets, where $\mathbf{1}_N$ is an $N \times 1$ vector of ones.

- ▶ Assume that the joint distribution of R_t is IID over time, with mean μ and covariance matrix V .
- ▶ In the static mean-variance framework an investor at time T chooses his/her portfolio weights w , so as to maximize the quadratic objective function

$$U(w) = E[R_p] - \frac{\gamma}{2} \text{Var}[R_p] = w' \mu - \frac{\gamma}{2} w' V w$$

Bayesian Asset Allocation

where $R_p = w'R_{T+1}$ is the future uncertain portfolio return at time $T + 1$ and γ is the coefficient of relative risk aversion.

- When both μ and V are assumed to be known, the optimal portfolio weights are

$$w^* = \frac{1}{\gamma} V^{-1} \mu$$

and the maximized expected utility is

$$U(w^*) = \frac{1}{2\gamma} \mu' V^{-1} \mu = \frac{\theta^2}{2\gamma}$$

where $\theta^2 = \mu' V^{-1} \mu$ is the squared Sharpe ratio of the ex ante tangency portfolio of the risky assets.

- This is the well known mean-variance theory pioneered by Markowitz (1952).
- In practice, the problem is that w^* is not computable because μ and V are unknown. As a result, the above mean-variance theory is usually applied in two steps.
- In the first step, the mean and covariance matrix of the asset returns are estimated based on the observed data.

Bayesian Asset Allocation

- Given a sample size of T , the standard maximum likelihood estimators are

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T R_t$$
$$\hat{V} = \frac{1}{T} \sum_{t=1}^T (R_t - \hat{\mu})(R_t - \hat{\mu})'$$

- Then, in the second step, these sample estimates are treated as if they were the true parameters, and are simply plugged in to compute the estimated optimal portfolio weights,

$$\hat{w}^{\text{ML}} = \frac{1}{\gamma} \hat{V}^{-1} \hat{\mu}$$

- The two-step procedure gives rise to a parameter uncertainty problem because it is the estimated parameters, not the true ones, that are used to compute the optimal portfolio weights.
- Consequently, the utility associated with the plug-in portfolio weights can be substantially different from $U(w^*)$.

Bayesian Asset Allocation

- Denote by θ the vector of all the parameters (both μ and V).
- Mathematically, the two-step procedure maximizes the expected utility conditional on the estimated parameters, denoted by $\hat{\theta}$, being equal to the true ones,

$$\max_w [U(w) \mid \theta = \hat{\theta}]$$

and the uncertainty or estimation errors are ignored.

- To account for estimation risk, let us specify the posterior distribution of the parameters as

$$p(\mu, V \mid \Phi_T) = p(\mu \mid V, \Phi_T) \times p(V \mid \Phi_T)$$

with

$$p(\mu \mid V, \Phi_T) \propto |V|^{-1/2} \exp\left\{-\frac{1}{2} \text{tr}[T(\mu - \hat{\mu})(\mu - \hat{\mu})' V^{-1}]\right\}$$

$$P(V) \propto |V|^{-\frac{\nu}{2}} \exp\left\{-\frac{1}{2} \text{tr} V^{-1}(T\hat{V})\right\}$$

where $\nu = T + N$.

Bayesian Asset Allocation

- The predictive distribution is:

$$p(R_{T+1}|\Phi_T) \propto |\hat{V} + (R_{T+1} - \hat{\mu})(R_{T+1} - \hat{\mu})'/(T + 1)|^{-T/2}$$

which is a multivariate t -distribution with $T - N$ degrees of freedom.

- While the problem of estimation error is recognized by Markowitz (1952), it is only in the 70s that this problem receives serious attention.
- Winkler (1973) and Winkle and Barry (1975) are earlier examples of Bayesian studies on portfolio choice.
- Brown (1976, 1978) and Klein and Bawa (1976) lay out independently and clearly the Bayesian predictive density approach, especially Brown (1976) who explains thoroughly the estimation error problem and the associated Bayesian approach.

Bayesian Asset Allocation

➤ Later, Bawa, Brown, and Klein (1979) provide an excellent review of the early literature.

➤ Under the diffuse prior, it is known that the Bayesian optimal portfolio weights are

$$\hat{w}^{\text{Bayes}} = \frac{1}{\gamma} \left(\frac{T - N - 2}{T + 1} \right) \hat{V}^{-1} \hat{\mu}$$

➤ In contrast with the classical weights \hat{w}^{ML} , the Bayesian portfolio is proportion to $\hat{V}^{-1} \hat{\mu}$, but the proportional coefficient is $(T - N - 2)/(T + 1)$ instead of 1.

➤ The coefficient can be substantially smaller when N is large relative to T .

➤ Intuitively, the assets are riskier in the Bayesian framework since parameter uncertainty is an additional source of risk and this risk is now factored into the portfolio decision.

➤ As a result, the overall position in the risky assets are generally smaller than before.

Bayesian Asset Allocation

- However, in the classical framework, \hat{w}^{ML} is a biased estimator of the true weights since, under the normality assumption,

$$E\hat{w}^{\text{ML}} = \frac{T - N - 2}{T} w^* \neq w^*$$

- Let

$$\tilde{V}^{-1} = \frac{T - N - 2}{T} \hat{V}^{-1}$$

- Then \tilde{V}^{-1} is an unbiased estimator of V^{-1} .

- The unbiased estimator of w^* is

$$\bar{w} = \frac{1}{\gamma} \frac{T - N - 2}{T} \hat{V}^{-1} \hat{\mu}$$

which is a scalar adjustment of \hat{w}^{ML} .

Bayesian Asset Allocation

- The unbiased classical weights differ from their Bayesian counterparts by a scalar $T/(T + 1)$.
- The difference is independent of N , and is negligible for all practical sample sizes T .
- Hence, parameter uncertainty makes little difference between Bayesian and classical approaches if the diffuse prior is used.
- Therefore, to provide new insights, it is important for a Bayesian to use informative priors, which is a decisive advantage of the Bayesian approach that can incorporate useful information easily into portfolio analysis.
- In the following we show how factor models can be employed to form priors.

The Black-Litterman Model

- The Black-Litterman (BL) approach attempts to propose new estimates for expected returns.
- Indeed, the sample means are simply too noisy.
- Asset pricing models - even if misspecified - could potentially deliver a good guidance.
- To illustrate, you consider a K -factor model (factors are portfolio spreads) and run the time series regression

$$\underset{N \times 1}{r_t^e} = \underset{N \times 1}{\alpha} + \underset{N \times 1}{\beta_1} f_{1t} + \underset{N \times 1}{\beta_2} f_{2t} + \cdots + \underset{N \times 1}{\beta_K} f_{Kt} + \underset{N \times 1}{e_t}$$

- Then the estimated excess mean return is given by

$$\hat{\mu}^e = \hat{\beta}_1 \hat{\mu}_{f_1} + \hat{\beta}_2 \hat{\mu}_{f_2} + \cdots + \hat{\beta}_K \hat{\mu}_{f_K}$$

where $\hat{\beta}_1, \hat{\beta}_2 \dots \hat{\beta}_K$ are the sample estimates of the factor loadings, and $\hat{\mu}_{f_1}, \hat{\mu}_{f_2} \dots \hat{\mu}_{f_K}$ are the sample estimates of the factor mean returns.

The Black-Litterman Model

- The BL approach combines a model (CAPM) with some views, either relative or absolute, about expected returns.

- The BL vector of mean returns is given by

$$\mu_{BL} = \left[\begin{pmatrix} \tau & \Sigma \\ 1 \times 1 & N \times N \end{pmatrix}^{-1} + \begin{matrix} P' & \Omega^{-1} & P \\ N \times K & K \times K & K \times N \end{matrix} \right]^{-1} \left[\begin{pmatrix} \tau & \Sigma \\ 1 \times 1 & N \times N \end{pmatrix}^{-1} \begin{matrix} \mu^{eq} \\ N \times 1 \end{matrix} + \begin{matrix} P' & \Omega^{-1} \\ N \times K & K \times K \end{matrix} \begin{matrix} \mu^v \\ K \times 1 \end{matrix} \right]$$

- We need to understand the essence of the following parameters, which characterize the mean return vector: $\Sigma, \mu^{eq}, P, \tau, \Omega, \mu^v$
- Starting from the Σ matrix - you can choose any feasible specification either the sample covariance matrix, or the equal correlation, or an asset pricing based covariance.

The Black-Litterman Model

- The μ^{eq} , which is the equilibrium vector of expected return, is constructed as follows.
- Generate ω_{MKT} , the $N \times 1$ vector denoting the weights of any of the N securities in the market portfolio based on market capitalization.
- Of course, the sum of weights must be unity.
- Then, the price of risk is $\gamma = \frac{\mu_m - Rf}{\sigma_m^2}$ where μ_m and σ_m^2 are the expected return and variance of the market portfolio.
- Later, we will justify this choice for that price of risk.
- One could pick a range of values for γ and examine performance for each choice.
- If you work with monthly observations, then switching to the annual frequency does not change γ as both the numerator and denominator are multiplied by 12 under the IID assumption.
- It does change the Sharpe ratio, however, as the standard deviation grows with the horizon by the square root of the period, while the expected return grows linearly.

The Black-Litterman Model

- ▶ Having at hand both ω_{MKT} and γ , the equilibrium return vector is given by

$$\mu^{eq} = \gamma \Sigma \omega_{MKT}$$

- ▶ This vector is called neutral mean or equilibrium expected return.
- ▶ To understand why, notice that if you have a utility function that generates the tangency portfolio of the form

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

- ▶ Then using μ^{eq} as the vector of excess returns on the N assets would deliver ω_{MKT} as the tangency portfolio.

The Black-Litterman Model

- The question being – would you get the same vector of equilibrium mean return if you directly use the CAPM?
- Yes, if...
- Under the CAPM the vector of excess returns is given by

$$\mu^e = \beta \mu_m^e$$
$$\beta = \frac{\text{cov}(r^e, r_m^e)}{\sigma_m^2} = \frac{\text{cov}(r^e, (r^e)' w_{MKT})}{\sigma_m^2} = \frac{\sum w_{MKT}}{\sigma_m^2}$$
$$\text{CAPM: } \mu_{N \times 1}^e = \frac{\sum w_{MKT}}{\sigma_m^2} \mu_m^e = \gamma \sum w_{MKT}$$

The Black-Litterman Model

➤ Since

$$\mu_m^e = (\mu^e)' w_{MKT} \quad \text{and} \quad r_m^e = (r^e)' w_{MKT}$$

➤ Then

$$\mu^e = \frac{\mu_m^e}{\sigma_m^2} \sum w_{MKT} = \mu^{eq}$$

➤ So indeed, if you use (i) the sample covariance matrix, rather than any other specification, as well as (ii)

$$\gamma = \frac{\mu_m - Rf}{\sigma_m^2}$$

➤ Then the BL equilibrium expected returns and expected returns based on the CAPM are identical.

The Black-Litterman Model

- In the BL approach the investor/econometrician forms some views about expected returns as described below.
- P is defined as that matrix which identifies the assets involved in the views.
- To illustrate, consider two "absolute" views only.
- The first view says that stock 3 has an expected return of 5% while the second says that stock 5 will deliver 12%.
- In general the number of views is K .
- In our case $K=2$.
- Then P is a $2 \times N$ matrix.
- The first row is all zero except for the third entry which is one.
- Likewise, the second row is all zero except for the fifth entry which is one.

The Black-Litterman Model

- ▶ Let us consider now two "relative views".
- ▶ Here we could incorporate market anomalies into the BL paradigm.
- ▶ Anomalies are cross sectional patterns in stock returns unexplained by the CAPM.
- ▶ Example: price momentum, earnings momentum, value, size, accruals, credit risk, dispersion, and volatility.
- ▶ Let us focus on price momentum and the value effects.
- ▶ Assume that both momentum and value investing outperform.
- ▶ The first row of P corresponds to momentum investing.
- ▶ The second row corresponds to value investing.
- ▶ Both the first and second rows contain N elements.

The Black-Litterman Model

- Winner stocks are the top 10% performers during the past six months.
- Loser stocks are the bottom 10% performers during the past six months.
- Value stocks are 10% of the stocks having the highest book-to-market ratio.
- Growth stocks are 10% of the stocks having the lowest book-to-market ratios.
- The momentum payoff is a return spread – return on an equal weighted portfolio of winner stocks minus return on equal weighted portfolio of loser stocks.
- The value payoff is also a return spread – the return differential between equal weighted portfolios of value and growth stocks.

The Black-Litterman Model

- Suppose that the investment universe consists of 100 stocks
- The first row gets the value 0.1 if the corresponding stock is a winner (there are 10 winners in a universe of 100 stocks).
- It gets the value -0.1 if the corresponding stock is a loser (there are 10 losers).
- Otherwise, it gets the value zero.
- The same idea applies to value investing.
- Of course, since we have relative views here (e.g., return on winners minus return on losers) then the sums of the first row and the sum of the second row are both zero.
- The same applies to value versus growth stocks.

The Black-Litterman Model

- Rule: the sum of the row corresponding to absolute views is one, while the sum of the row corresponding to relative views is zero.
- μ^v is the $K \times 1$ vector of K views on expected returns.
- Using the absolute views above
$$\mu^v = [0.05, 0.12]'$$
- Using the relative views above, the first element is the payoff to momentum trading strategy (sample mean); the second element is the payoff to value investing (sample mean).
- Ω is a $K \times K$ covariance matrix expressing uncertainty about views.
- It is typically assumed to be diagonal.
- In the absolute views case described above $\Omega(1,1)$ denotes uncertainty about the first view while $\Omega(2,2)$ denotes uncertainty about the second view – both are at the discretion of the econometrician/investor.

The Black-Litterman Model

- In the relative views described above: $\Omega(1,1)$ denotes uncertainty about momentum. This could be the sample variance of the momentum payoff.
- $\Omega(2,2)$ denotes uncertainty about the value payoff. This could be the sample variance of the value payoff.
- There are many debates among professionals about the right value of τ .
- From a conceptual perspective it should be $1/T$ where T denotes the sample size.
- You can pick $\tau = 0.1$
- You can also use other values and examine how they perform in real-time investment decisions.

Extending BL to Incorporate Sample Moments

- ▶ Consider a sample of size T , e.g., $T=60$ monthly observations.
- ▶ Let us estimate the mean and covariance of our N assets based on the sample.
- ▶ Then the vector of expected return that serves as an input for asset allocation is given by

$$\mu = [\Delta^{-1} + (V_{sample}/T)^{-1}]^{-1} \cdot [\Delta^{-1}\mu_{BL} + (V_{sample}/T)^{-1}\mu_{sample}]$$

where

$$\Delta = [(\tau\Sigma)^{-1} + P'\Omega^{-1}P]^{-1}$$

Bayesian Asset Allocation: Asset Pricing Prior

- ▶ Pástor (2000) and Pástor and Stambaugh (1999) introduce interesting priors that reflect an investor's degree of belief in an asset pricing model.
- ▶ To see how this class of priors is formed, assume $R_t = (y_t, x_t)$, where y_t contains the excess returns of m non-benchmark positions and x_t contains the excess returns of $K (= N - m)$ benchmark positions.
- ▶ Consider a factor model multivariate regression

$$y_t = \alpha + Bx_t + u_t$$

where u_t is an $m \times 1$ vector of residuals with zero means and a non-singular covariance matrix $\Sigma = V_{11} - BV_{22}B'$, and α and B are related to μ and V through

$$\alpha = \mu_1 - B\mu_2, \quad B = V_{12}V_{22}^{-1}$$

where μ_i and V_{ij} ($i, j = 1, 2$) are the corresponding partitions of μ and V ,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

Bayesian Asset Allocation: Asset Pricing Prior

- For a factor-based asset pricing model, such as the three-factor model of Fama and French (1993), the restriction is $\alpha = 0$.
- To allow for mispricing uncertainty, Pástor (2000), and Pástor and Stambaugh (2000) specify the prior distribution of α as a normal distribution conditional on Σ ,

$$\alpha|\Sigma \sim N \left[0, \sigma_{\alpha}^2 \left(\frac{1}{s_{\Sigma}^2} \Sigma \right) \right]$$

where s_{Σ}^2 is a suitable prior estimate for the average diagonal elements of Σ . The above alpha-Sigma link is also explored by MacKinlay and Pástor (2000) in the classical framework.

- The magnitude of σ_{α} represents an investor's level of uncertainty about the pricing ability of a given model.
- When $\sigma_{\alpha} = 0$, the investor believes dogmatically in the model and there is no mispricing uncertainty.
- On the other hand, when $\sigma_{\alpha} = \infty$, the investor believes that the pricing model is entirely useless.

Bayesian Asset Allocation: Investing in Mutual Funds

- Baks, Metrick, and Wachter (JF 2001) (henceforth BMW) and Pastor and Stambaugh (JFE 2002) have explored the role of prior information about fund performance in making investment decisions.
- BMW consider a mean variance optimizing investor who is largely skeptical about the ability of a fund manager to pick stocks and time the market.
- They find that even with a high degree of skepticism about fund performance the investor would allocate considerable amounts to actively managed funds.
- Pastor and Stambaugh nicely extend the BMW methodology to the case where prior uncertainty is not only about managerial skills but also about model pricing abilities.
- In particular, starting from Jensen (1965), mutual fund performance is typically defined as the intercept in the regression of the fund's excess returns on excess return of one or more benchmark assets.
- However, the intercept in such time series regressions could reflect a mix of fund performance as well as model mispricing.

Bayesian Asset Allocation: Mutual Funds

- In particular, consider the case wherein benchmark assets used to define fund performance are unable to explain the cross section dispersion of passive assets, that is, the sample alpha in the regression of nonbenchmark passive assets on benchmarks assets is nonzero.
- Then model mispricing emerges in the performance regression.
- Thus, Pastor and Stambaugh formulate prior beliefs on both performance and mispricing.
- Geczy, Stambaugh, and Levin (2005) apply the Pastor Stambaugh methodology to study the cost of investing in socially responsible mutual funds.
- Comparing portfolios of these funds to those constructed from the broader fund universe reveals the cost of imposing the socially responsible investment (SRI) constraint on investors seeking the highest Sharpe ratio.
- This SRI cost depends crucially on the investor's views about asset pricing models and stock-picking skill by fund managers.

Mutual Funds: Prior dependence across funds

- ▶ BMW and Pastor and Stambaugh assume that the prior on alpha is independent across funds.
- ▶ Jones and Shanken (JFE 2002) show that under such an independence assumption, the maximum posterior mean alpha increases without bound as the number of funds increases and "extremely large" estimates could randomly be generated.
- ▶ This is true even when fund managers display no skill.
- ▶ Thus they propose incorporating prior dependence across funds.
- ▶ Then, investors aggregate information across funds to form a general belief about the potential for abnormal performance.
- ▶ Each fund's alpha estimate is shrunk towards the aggregate estimate, mitigating extreme views

Bayesian Asset Allocation: With Predictive Variables

- Consider a one-period optimizing investor who must allocate at time T funds between the value-weighted NYSE index and one-month Treasury bills.
- The investor makes portfolio decisions based on estimating the predictive system

$$\begin{aligned}r_t &= a + b'z_{t-1} + u_t \\z_t &= \theta + \rho z_{t-1} + v_t\end{aligned}$$

where r_t is the continuously compounded NYSE return in month t in excess of the continuously compounded T-bill rate for that month, z_{t-1} is a vector of M predictive variables observed at the end of month $t - 1$, b is a vector of slope coefficients, and u_t is the regression disturbance in month t .

- The evolution of the predictive variables is essentially stochastic, as shown earlier. Here, the evolution is crucial for understanding expected return and risk over long horizons.
- The regression residuals are assumed to obey the normal distribution and are IID.

Bayesian Asset Allocation

- In particular, let $\eta_t = [u_t, v_t']'$ then $\eta_t \sim N(0, \Sigma)$

where

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{vu} & \Sigma_v \end{bmatrix}$$

- The distribution of r_{T+1} , e.g., the time $T + 1$ NYSE excess return, conditional on data and model parameters is $N(a + b'z_T, \sigma_u^2)$.
- Assuming the inverted Wishart prior distribution for Σ and multivariate normal prior for the intercept and slope coefficients in the predictive system, the Bayesian predictive distribution $P(r_{T+1}|\Phi_T)$ obeys the Student t density.
- Then, considering a power utility investor with parameter of relative risk aversion denoted by γ the optimization formulation is

$$\omega^* = \arg \max_{\omega} \int_{r_{T+1}} \frac{[(1 - \omega) \exp(r_f) + \omega \exp(r_f + r_{T+1})]^{1-\gamma}}{1 - \gamma} P(r_{T+1}|\Phi_T) dr_{T+1}$$

subject to ω being nonnegative.

Bayesian Asset Allocation

- It is infeasible to have analytic solution for the optimal portfolio.
- Then, considering a power utility investor with parameter of relative risk aversion denoted by γ the optimization formulation is

$$\omega^* = \arg \max_{\omega} \int_{r_{T+1}} \frac{[(1 - \omega) \exp(r_f) + \omega \exp(r_f + r_{T+1})]^{1-\gamma}}{1 - \gamma} P(r_{T+1} | \Phi_T) dr_{T+1}$$

subject to ω being nonnegative.

- Then, given G independent draws for R_{T+1} from the predictive distribution, the optimal portfolio is found by implementing a constrained optimization code to maximize the quantity

$$\frac{1}{G} \sum_{g=1}^G \frac{\{(1 - \omega) \exp(r_f) + \omega \exp(r_f + R_{T+1}^{(g)})\}^{1-\gamma}}{1 - \gamma}$$

subject to ω being nonnegative.

Predictive Regressions with Stochastic Initial Observation

- Let us revisit the predictive regression with one predictor only

$$\begin{aligned}r_t &= a + bz_{t-1} + u_t \\z_t &= \theta + pz_{t-1} + v_t \\ \begin{bmatrix} u_t \\ v_t \end{bmatrix} &\sim \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}\end{aligned}$$

- The system of two equations can be re-expressed in the normal multivariate form

$$R = X\beta + u$$

where R is a $T \times 2$ matrix with the first (second) column consisting of excess stock return (current values of the predictors) and X is a $T \times 2$ matrix with the first (second) column consisting of a $T \times 1$ vector of ones (lagged values of the predictors).

- Previously we analyzed such multivariable regressions, assuming the initial observation is non stochastic.
- In particular, the posterior is given by (recall from the section on multivariate regression)

$$|\Sigma|^{-\frac{T+N+1}{2}} \exp \left(-\frac{1}{2} (\beta - \tilde{\beta}) [\Sigma^{-1} \otimes \tilde{\Psi}] (\beta - \tilde{\beta}) + \text{tr}[\tilde{S}\Sigma^{-1}] \right)$$

Stochastic Initial Observation

- In the case of stochastic initial observation we multiply this posterior by $P(z_0|b, \Sigma)$ which is

$$\left(\frac{1-p^2}{2\pi\sigma_v^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1-p^2}{2\sigma_v^2}\left(z_0 - \frac{\theta}{1-p}\right)^2\right\}$$

- Notice that this is the normal distribution with unconditional mean and variance of the predictive variables.
- Integrating the posterior analytically to obtain the marginal posterior density of the parameters does not appear to be feasible
- Instead, the posterior density can be obtained using the MH algorithm.
- The candidate distribution is normal/inverted Wishart
- See Stambaugh (1999) for a detailed discussion.

Bayesian Asset Allocation

- ▶ Kandel and Stambaugh (1996) show that even when the statistical evidence on predictability, as reflected through the R^2 of the predictive regression, is weak, the current values of the predictive variables, z_T , can exert a substantial influence on the optimal portfolio.
- ▶ Whereas Kandel and Stambaugh (1996) study asset allocation in a single-period framework, Barberis (2000) analyzes multi-period investment decisions, considering both a buy-and-hold investor as well as an investor who dynamically rebalances the optimal stock-bond allocation.
- ▶ Implementing long horizon asset allocation in a buy-and-hold setup is quite straightforward.
- ▶ In particular, let K denote the investment horizon, and let $R_{T+K} = \sum_{k=1}^K r_{T+k}$ be the cumulative (continuously compounded) return over the investment horizon.

Bayesian Asset Allocation

- Avramov (2002) shows that the distribution for R_{T+K} conditional on the data (denoted Φ_T) and set of parameters (denoted Θ) is given by

$$R_{T+K} | \Theta, \Phi_T \sim N(\lambda, \Upsilon)$$

where

$$\lambda = Ka + b'[(\rho^K - I_M)(\rho - I_M)^{-1}]z_T + b'[\rho(\rho^{K-1} - I_M)(\rho - I_M)^{-1} - (K-1)I_M](\rho - I_M)^{-1}\theta$$

$$\Upsilon = K\sigma_u^2 + \sum_{k=1}^K \delta(k)\Sigma_v\delta(k)' + \sum_{k=1}^K \sigma_{uv}\delta(k)' + \sum_{k=1}^K \delta(k)\sigma_{vu}$$

$$\delta(k) = b'[(\rho^{k-1} - I_M)(\rho - I_M)^{-1}]$$

- Drawing from the Bayesian predictive distribution is done in two steps.
- First, draw the model parameters Θ from their posterior distribution with either fixed or stochastic first observation.
- Second, conditional on model parameters, draw R_{T+K} from the normal distribution.
- The optimal portfolio is accomplished by numerically maximizing

$$\frac{1}{G} \sum_{g=1}^G \frac{\left\{ (1-\omega) \exp(r_f) + \omega \exp(r_f + R_{T+k}^{(g)}) \right\}^{1-\gamma}}{1-\gamma}$$

Investment Opportunities: Risk for the Long Run

- The mean and variance in an IID world increase linearly with the investment horizon.
- There is no horizon effect when (i) returns are IID and (ii) estimation risk is not accounted for, as indeed shown by Samuelson (1969) and Merton (1969) in an equilibrium framework.
- Incorporating estimation risk in an IID setup, Barberis (2000) shows that stocks appear riskier in longer horizons.
- Incorporating return predictability and estimation risk, Barberis (2000) shows that investors allocate considerably more heavily to equity the longer their horizon.
- This is not clear ex ante - there is a tradeoff between mean reversion and estimation risk. The mean reversion effect appears to be stronger.

Risk for the Long Run – Mean Reversion

- Let r_t be the cc return in time t and r_{t+1} be the cc return in time $t + 1$.
- Assume that $\text{var}(r_t) = \text{var}(r_{t+1})$.
- Then the cumulative two period return is $R(t, t + 1) = r_t + r_{t+1}$.
- The question of interest: is $\text{var}(R(t, t + 1))$ greater than equal to or smaller than $2\text{var}(r_t)$.
- Of course if stock returns are iid the variance grows linearly with the investment horizon, as long as estimation risk is overlooked.
- However, let us assume that stock returns can be predictable by the dividend yield:

$$\begin{aligned}r_t &= \alpha + \beta \text{div}_{t-1} + \epsilon_t \\ \text{div}_t &= \phi + \delta \text{div}_{t-1} + \eta_t\end{aligned}$$

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

Risk for the Long Run: Mean Reversion

- It follows that

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

- Thus, if $\beta^2\sigma_2^2 + 2\beta\sigma_{12} < 0$ the conditional variance of two period return is less than the twice conditional variance of one period return, which is indeed the case based on the empirical evidence.
- This is the mean-reversion property.
- While mean reversion makes stocks appear less risky with the investment horizon, there are other effects (estimation risk, uncertainty about current and future mean return) which make stocks appear riskier.
- In the next slide, we decompose the predictive variance of long horizon return to all these effects.

Risk for the Long Run: The Predictive Variance of Long Horizon Cumulative Return

- In the traditional predictive regression setup, the predictive variance consists of four parts:

$$\text{Var}(r_{T,T+k} | D_T) =$$

a. $E(k\sigma_r^2 | D_T)$

b. $+E\left[\sum_{i=1}^{k-1} 2b_r(I - B_x)^{-1}(I - B_x^i)\Sigma_{xr} | D_T\right]$

c. $+E\left[\sum_{i=1}^{k-1} \left(b_r(I - B_x)^{-1}(I - B_x^i)\right)\Sigma_x \left(b_r(I - B_x)^{-1}(I - B_x^i)\right)' | D_T\right]$

d. $+Var\left[ka_r + b_r(I - B_x)^{-1}\left(\left(kI - (I - B_x)^{-1}(I - B_x^k)\right)a_x + (I - B_x^k)x_T\right) | D_T\right]$

a. iid uncertainty *b.* mean reversion *c.* future expected return uncertainty *d.* estimation risk

The Predictive Variance of Long Horizon Cumulative Return

- The first is the IID component
- The second is the mean-reversion component
- The third reflects the uncertainty about future mean return
- The fourth is the estimation risk component
- In the predictive system of Pastor and Stambaugh (2012) there is a fifth component describing uncertainty about current mean return.
- Accounting for model uncertainty induces one more component of the predictive variance

Investment Opportunities: Risk for the Long Run

- Pastor and Stambaugh (2012) implement a predictive system to show that stocks may be more risky over long horizons from an investment perspective.
- They exhibit one additional source of uncertainty about current mean return, which increases with the horizon.
- Avramov, Cederbug, and Kvasnakova (2017) suggest model based priors on the return dynamics and show that per-period variance can be either higher or lower with the horizon.
- In particular, prospect theory and habit formation (Long Run Risk) investors perceive less (more) risky equities due to strong (weak) mean reversion.
- Bottom line: Economic theory could give important guidance about investments for the long run as the sample is not informative enough.
- The next slide gives more intuition about mean reversion.

Model Uncertainty

- Financial economists have identified variables that predict future stock returns, as noted earlier.
- However, the “correct” predictive regression specification has remained an open issue for several reasons.
- For one, existing equilibrium pricing theories are not explicit about which variables should enter the predictive regression.
- This aspect is undesirable, as it renders the empirical evidence subject to data over-fitting concerns.
- Indeed, Bossaerts and Hillion (1999) confirm in-sample return predictability, but fail to demonstrate out-of-sample predictability.

Model Uncertainty

- Moreover, the multiplicity of potential predictors also makes the empirical evidence difficult to interpret.
- For example, one may find an economic variable statistically significant based on a particular collection of explanatory variables, but often not based on a competing specification.
- Given that the true set of predictive variables is virtually unknown, the Bayesian methodology of model averaging is attractive, as it explicitly incorporates model uncertainty.
- The idea is to compute posterior probability for each candidate return forecasting model – then predicted return is the weighted average of return forecasting models with weights being the posterior model probabilities.
- Avramov (2002) derives analytic expressions for model posterior probability. Often numerical methods are proposed.
- Assuming equal prior model probability, the posterior probability is a normalized version of the marginal likelihood, which in turn is a mix of two components standing for model complexity and goodness-of-fit.

Bayesian Asset Allocation under Model Uncertainty

- In the context of asset allocation, the Bayesian weighted predictive distribution of cumulative excess continuously compounded returns averages over the model space, and integrates over the posterior distribution that summarizes the within-model uncertainty about Θ_j where j is the model identifier. It is given by

$$P(R_{T+K}|\Phi_T) = \sum_{j=1}^{2^M} P(\mathcal{M}_j|\Phi_T) \int_{\Theta_j} P(\Theta_j|\mathcal{M}_j, \Phi_T) P(R_{T+K}|\mathcal{M}_j, \Theta_j, \Phi_T) d\Theta_j$$

where $P(\mathcal{M}_j|\Phi_T)$ is the posterior probability that model \mathcal{M}_j is the correct one.

- Drawing from the weighted predictive distribution is done in three steps.
- First draw the correct model from the distribution of models.
- Then conditional upon the model implement the two steps, noted above, of drawing future stock returns from the model specific Bayesian predictive distribution.

Bayesian Asset Allocation - Stock return predictability and asset pricing models - informative priors

- The classical approach has examined whether return predictability is explained by rational pricing or whether it is due to asset pricing misspecification [see, e.g., Campbell (1987), Ferson and Korajczyk (1995), and Kirby (1998)].
- Studies such as these approach finance theory by focusing on two polar viewpoints: rejecting or not rejecting a pricing model based on hypothesis tests.
- The Bayesian approach incorporates pricing restrictions on predictive regression parameters as a reference point for a hypothetical investor's prior belief.
- The investor uses the sample evidence about the extent of predictability to update various degrees of belief in a pricing model and then allocates funds across cash and stocks.
- Pricing models are expected to exert stronger influence on asset allocation when the prior confidence in their validity is stronger and when they explain much of the sample evidence on predictability.

Bayesian Asset Allocation

- In particular, Avramov (2004) models excess returns on N investable assets as

$$r_t = \alpha(z_{t-1}) + \beta f_t + u_{rt},$$

$$\alpha(z_{t-1}) = \alpha_0 + \alpha_1 z_{t-1}$$

$$f_t = \lambda(z_{t-1}) + u_{ft}$$

$$\lambda(z_{t-1}) = \lambda_0 + \lambda_1 z_{t-1}$$

where f_t is a set of K monthly excess returns on portfolio based factors, α_0 stands for an N -vector of the fixed component of asset mispricing, α_1 is an $N \times M$ matrix of the time varying component, and β is an $N \times K$ matrix of factor loadings.

- Now, a conditional version of an asset pricing model (with fixed beta) implies the relation

$$\mathbb{E}(r_t \mid z_{t-1}) = \beta \lambda(z_{t-1})$$

for all t , where \mathbb{E} stands for the expected value operator.

- The model imposes restrictions on parameters and goodness of fit in the multivariate predictive regression

$$r_t = \mu_0 + \mu_1 z_{t-1} + v_t$$

where μ_0 is an N -vector and μ_1 is an $N \times M$ matrix of slope coefficients.

Bayesian Asset Allocation

$$r_t = (\mu_0 - \beta\lambda_0) + (\mu_1 - \beta\lambda_1)z_{t-1} + \beta f_t + u_{rt}$$

$$\mu_0 = \alpha_0 + \beta\lambda_0$$

$$\mu_1 = \alpha_1 + \beta\lambda_1.$$

- That is, under pricing model restrictions where $\alpha_0 = \alpha_1 = 0$ it follows that:

$$\mu_0 = \beta\lambda_0$$

$$\mu_1 = \beta\lambda_1$$

- This means that stock returns are predictable ($\mu_1 \neq 0$) iff factors are predictable.
- Makes sense; after all, under pricing restrictions the systematic component of returns on N stocks is captured by K common factors.
- Of course, if we relax the fixed beta assumption - time varying beta could also be a source of predictability. More later!
- Is return predictability explained by asset pricing models? Probably not!
- Kirby (1998) shows that returns are too predictable to be explained by asset pricing models.
- Ferson and Harvey (1999) show that $\alpha_1 \neq 0$.

Bayesian Asset Allocation

- Avramov and Chordia (2006) show that strategies that invest in individual stocks conditioning on time varying alpha perform extremely well. More later!
- So, should we disregard asset pricing restrictions? Not necessarily!
- The notion of rejecting or not rejecting pricing restrictions on predictability reflects extreme polar views.
- What if you are a Bayesian investor who believes pricing models could be useful albeit not perfect?
- As discussed earlier, such an approach has been formalized by Black and Litterman (1992) and Pastor (2000) in the context of IID returns and by Avramov (2004) who accounts for predictability.
- The idea is to mix the model and data.
- This is shrinkage approach to asset allocation.
- Let μ_d and Σ_d (μ_m and Σ_m) be the expected return vector and variance covariance matrix based on the data (model).

Bayesian Asset Allocation

- Simplistically speaking, moments used for asset allocation are

$$\begin{aligned}\mu &= \omega\mu_d + (1 - \omega)\mu_m \\ \Sigma &= \omega\Sigma_d + (1 - \omega)\Sigma_m\end{aligned}$$

where ω is the shrinkage factor

- In particular,
 - If you completely believe in the model you set $\omega = 0$.
 - If you completely disregard the model you set $\omega = 1$.
 - Going with the shrinkage approach means that $0 < \omega < 1$.
- The shrinkage of Σ is quite meaningless in this context.
- There are other quite useful shrinkage methods of Σ - see, for example, Jagannathan and Ma (2005).

Bayesian Asset Allocation

- ▶ Avramov (2004) derives asset allocation under the pricing restrictions alone, the data alone, and pricing restrictions and data combined.
- ▶ He shows that
 - Optimal portfolios based on the pricing restrictions deliver the lowest Sharpe ratios.
 - Completely disregarding pricing restrictions results in the second lowest Sharpe ratios.
 - Much higher Sharpe ratios are obtained when asset allocation is based on the shrinkage approach.

Could one exploit predictability to design outperforming trading strategies?

- Avramov and Chordia (2006), Avramov and Wermers (2006), and Avramov, Kosowski, Naik, and Teo (2009) are good references here.
- Let us start with Avramov and Chordia (2006).
- They study predictability through the out of sample performance of trading strategies that invest in individual stocks conditioning on macro variables.
- They focus on the largest NYSE-AMEX firms by excluding the smallest quartile of firms from the sample.
- They capture 3123 such firms during the July 1972 through November 2003 investment period.
- The investment universe contains 973 stocks, on average, per month.

Bayesian Asset Allocation - The Evolution of Stock Returns

- The underlying statistical models for excess stock returns, the market premium, and macro variables are

$$\begin{aligned}r_t &= \alpha(z_{t-1}) + \beta(z_{t-1})\text{mkt}_t + v_t \\ \alpha(z_{t-1}) &= \alpha_0 + \alpha_1 z_{t-1} \\ \beta(z_{t-1}) &= \beta_0 + \beta_1 z_{t-1} \\ \text{mkt}_t &= a + b'z_{t-1} + \eta_t \\ z_t &= c + dz_{t-1} + e_t\end{aligned}$$

- Stock level predictability could come up from:
 1. Model mispricing that varies with changing economic conditions ($\alpha_1 \neq 0$);
 2. Factor sensitivities are predictable ($\beta_1 \neq 0$);
 3. The equity premium is predictable ($b \neq 0$).

Bayesian Asset Allocation

- In the end, time varying model alpha is the major source of predictability and investment profitability focusing on individual stocks, portfolios, mutual funds, and hedge funds.
- In the mutual fund and hedge fund context alpha reflects skill (but can also entails mispricing).
- Indeed, alpha reflects skill only if the benchmarks used to measure performance are able to price all passive payoffs.

Bayesian Asset Allocation - The Proposed Strategy

- ▶ Avramov and Chordia (2006) form optimal portfolios from the universe of AMEX-NYSE stocks over the period 1972 through 2003 with monthly rebalancing on the basis of various models for stock returns.
- ▶ For instance, when predictability in alpha, beta, and the equity premium is permissible, the mean and variance used to form optimal portfolios are

$$\mu_{t-1} = \hat{\alpha}_0 + \hat{\alpha}_1 z_{t-1} + \hat{\beta}(z_{t-1})[\hat{a} + \hat{b}z_{t-1}]$$

$$\Sigma_{t-1} = \hat{\beta}(z_{t-1})\hat{\beta}(z_{t-1})'\hat{\sigma}_{mkt}^2 + \hat{\Psi}$$

$$+ \underbrace{\delta_1 \hat{\beta}(z_{t-1})\hat{\beta}(z_{t-1})'\hat{\sigma}_{mkt}^2 + \delta_2 \hat{\Psi}}_{\text{Estimation risk}}.$$

- ▶ The trading strategy is obtained by maximizing

$$w_t = \arg \max_{w_t} \left\{ w_t' \mu_t - \frac{1}{2(1/\gamma_t - r_{ft})} w_t' [\Sigma_{t-1} + \mu_{t-1} \mu_{t-1}'] w_t \right\}$$

where γ_t is the risk aversion level.

- ▶ We do not permit short selling of stocks but we do allow buying on margin.

Bayesian Asset Allocation - Performance evaluation

- We implement a recursive scheme:
 - The first optimal portfolio is based on the first 120 months of data on excess returns, market premium, and predictors. (That is, the first estimation window is July 1962 through June 1972.)
 - The second optimal portfolio is based on the first 121 months of data.
- Altogether, we form 377 optimal portfolios on a monthly basis for each model under consideration.
- We record the realized excess return on any strategy
$$r_{p,t+1} = \omega_t' r_{t+1}.$$
- We evaluate the ex-post out-of-sample performance of the trading strategies based on the realized returns.
- Ultimately, we are able to assess the (quite large) economic value of predictability as well as show that our strategies successfully rotate across the size, value, and momentum styles during changing business conditions.

Bayesian Asset Allocation - So far, we have shown

- Over the 1972-2003 investment period, portfolio strategies that condition on macro variables outperform the market by about 2% per month.
- Such strategies generate positive performance even when adjusted by the size, value, and momentum factors as well as by the size, book-to-market, and past return characteristics.
- In the period prior to the discovery of the macro variables, investment profitability is primarily attributable to the predictability in the equity premium.
- In the post-discovery period, the relation between the macro variables and the equity premium is attenuated considerably.
- Nevertheless, incorporating macro variables is beneficial because such variables drive stock-level alpha and beta variations.
- Predictability based strategies hold small, growth, and momentum stocks and load less (more) heavily on momentum (small) stocks during recessions.
- Such style rotation has turned out to be successful ex post.

Bayesian Asset Allocation - Exploiting Predictability in Mutual Fund Returns

- Can we use our methodology to generate positive performance based on the universe of actively managed no-load equity mutual funds?
- What do we know about equity mutual funds?
 - In 2015 about \$6 trillion is currently invested in U.S. equity mutual funds, making them a fundamental part of the portfolio of a domestic investor.
 - Active fund management underperforms, on average, passive benchmarks.
 - Strategies that attempt to identify subsets of funds using information variables such as past returns or new money inflows (“hot hands” or “smart money” strategies) underperform when investment payoffs are adjusted the Fama-French and momentum benchmarks.
- Avramov and Wermers (2006) show that strategies that invest in no-load equity funds conditioning on macro variables generate substantial positive performance.

Understanding the SDF approach and the Hansen-Jagannathan Distance Measure

The SDF Approach

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

$$E[\xi_{t+1}R_{i,t+1}|\mathcal{I}_t] = 1,$$

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

$$\xi_{t+1} = a_t + b'_t f_{t+1}.$$

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

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

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

Is the pricing Kernel linear or nonlinear in the factors?

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

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

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

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

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

- This first order approximation results in the traditional CAPM.

- The second order approximation is given by

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

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

Are pricing kernel parameters fixed or time-varying?

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

$$\begin{aligned}\xi_{t+1} &= 1 - \gamma \Delta c_{t+1} + o(c_t), \\ &= a + b \Delta c_{t+1}.\end{aligned}$$

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

- The utility function under habit formation

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

where H_t is the consumption habit.

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

$$\xi_{t+1} = a(z_t) + b(z_t)f_{t+1},$$

$$a_t = a_0 + a_1 z_t,$$

$$b_t = b_0 + b_1 z_t.$$

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

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

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

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

The Hansen Jagannathan — HJ — (1997)

Distance Measure

- The HJ measure is used for comparing and testing asset pricing models.
- Suppose you want to compare the performance of competing, not necessarily nested, asset pricing models.
- If there is only one asset, then you can compare the pricing error, i.e., the difference between the market price of an asset and the hypothetical price implied by a particular SDF.
- However, when there are many assets, it is rather difficult to compare the pricing errors across the different candidate SDFs unless pricing errors of one SDF are always smaller across all assets.
- One simple idea would be to examine the pricing error on the portfolio (there are infinitely many such portfolios) that is most mispriced by a given model.
- Then, the superior model is the one with the smallest pricing error.
- However, there is a practical problem in implementing this simple idea.
- Suppose there are at least two assets which do not have the same pricing error for a given candidate SDF

- ▶ Let R_{1t} and R_{2t} denote the corresponding gross returns.
- ▶ Suppose that (i) the date $t - 1$ market prices of these payoffs are both unity, and (ii) the model assigns prices of $1 + \psi_i$, i.e., the pricing errors are ψ_1 and ψ_2 .
- ▶ Consider now forming a zero-investment portfolio by going long one dollar in security 1 and short one dollar in security 2.
- ▶ The pricing error of this zero-cost position is $\psi_1 - \psi_2$.
- ▶ That is, as long as the difference is not zero the pricing error of any portfolio of the two assets can be arbitrarily large by adding a scale multiple of this zero-investment portfolio.

The HJ idea

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

- Suppose that the SDF is modeled as

$$\xi_t(\Theta) = \Theta_0 + \Theta_{vw}R_t^{vw} + \Theta_{prem}R_{t-1}^{prem} + \Theta_{labor}R_t^{labor} = \Theta'Y_t$$

where

$$\Theta = [\Theta_0, \Theta_{vw}, \Theta_{prem}, \Theta_{labor}]',$$

$$Y_t = [1, R_t^{vw}, R_{t-1}^{prem}, R_t^{labor}]'$$

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

$$f_t(\Theta) = R_t\xi_t(\Theta) - \iota_N = R_tY_t'\Theta - \iota_N.$$

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

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

- This is the HJ distance measure - is not the HJ bound.

- Since the vector Θ is unknown, a natural way to estimate the system is to choose those values that minimize the function.
- We can then assess the specification error of a given stochastic discount factor by examining the maximum pricing error δ .
- Next, compute some sample moments

$$D_T = \frac{1}{T} \sum_{t=1}^T \frac{\partial f_t(\Theta)}{\partial \Theta} = \frac{1}{T} \sum_{t=1}^T R_t Y_t' = \frac{1}{T} R' Y,$$

$$g_T(\Theta) = \frac{1}{T} \sum_{t=1}^T f_t(\Theta) = D_T \Theta - \iota_N,$$

$$G_T = \frac{1}{T} \sum_{t=1}^T R_t R_t' = \frac{1}{T} R' R,$$

where

$$R = [R_1, R_2, \dots, R_T]',$$

$$Y = [Y_1, Y_2, \dots, Y_T].$$

- The sample analog of the HJ distance is thus

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

- The first order condition of the minimization problem

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

is given by

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

which gives an analytic expression for the sample minimizer

$$\begin{aligned}\hat{\Theta} &= (D_T' G_T^{-1} D_T)^{-1} D_T' G_T^{-1} \iota_N, \\ &= T(Y' R(R' R)^{-1} R' Y)^{-1} Y' R(R' R)^{-1} \iota_N.\end{aligned}$$

➤ It follows that

$$g_T(\hat{\Theta}) = R' Y(Y' R(R' R)^{-1} R' Y)^{-1} Y' R(R' R)^{-1} \iota_N - \iota_N$$

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

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

➤ If the data is serially uncorrelated, the estimate of the variance matrix of pricing errors is given by

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

- That is, the estimator $\hat{\Theta}$ is equivalent to a GMM estimator defined by Hansen (1982) with the moment condition $E[f(\Theta)] = 0$ and the weighting matrix G^{-1} .
- If the weighing matrix is optimal in the sense of Hansen (1982), then $T\delta_T^2$ is asymptotically a random variable of χ^2 distribution with $N - m$ dof, where m is the dimension of Θ .
- Moreover, the optimal variance of $\hat{\Theta}$ becomes

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

- However, G is generally not optimal, and thus the distribution of $T\delta_T^2$ is not χ_{N-m}^2 .
- Instead, the limiting distribution of this statistic is given by

$$u = \sum_{j=1}^{N-m} \lambda_j v_j,$$

where v_1, v_2, \dots, v_{N-m} are independent $\chi^2(1)$ random variables, and $\lambda_1, \lambda_2, \dots, \lambda_{N-m}$ are $N - m$ nonzero eigenvalues of the matrix A given by

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

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

- As long as we have a consistent estimate S_T of the matrix S , we can estimate the matrix A by replacing S and G by S_T and G_T , respectively.
- One can generate a large number of draws from the nonstandard distribution to determine the p -value of the HJ distance measure, or whether or not it is equal to zero.
- You can follow the below-described algorithm to compute the empirical p -value:
 1. Compute $T\delta_T^2 = Tg_T(\hat{\Theta})G_T^{-1}g_T(\hat{\Theta})$.
 2. Obtain the $N - m$ largest eigenvalues of \hat{A} , a consistent estimate of A .
 3. Generate $N - m$ independent draws from $\chi^2(1)$. For example, using the Matlab command $g = \text{chi2rnd}(v, 1000, 1)$ generates 1000 independent draws from $\chi^2(v)$
 4. Based on these independent draws, compute the statistic u_i
 5. If $u_i > T\delta_T^2$ set $I_i = 1$. Otherwise set $I_i = 0$
 6. Repeat steps 3-5 100,000 times
 7. The empirical p -value is given by $\frac{1}{100,000} \sum_{i=1}^{100,000} I_i$.

Considering conditional pricing models

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

$$\xi_{t+1} = (\Theta_0' \mathbf{X}_t) + (\Theta_1' \mathbf{X}_t)f_{t+1}^1 + \cdots + (\Theta_K' \mathbf{X}_t)f_{t+1}^K,$$

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

- As noted earlier the first order condition implies that

$$E[R_{t+1}((\Theta_0' \mathbf{X}_t) + (\Theta_1' \mathbf{X}_t)f_{t+1}^1 + \cdots + (\Theta_K' \mathbf{X}_t)f_{t+1}^K) | \mathbf{Z}_t] = \iota_N.$$

- We collect the vector of errors,

$$\begin{aligned} f_{t+1} &= R_{t+1}((\Theta_0' \mathbf{X}_t) + (\Theta_1' \mathbf{X}_t)f_{t+1}^1 + \cdots + (\Theta_K' \mathbf{X}_t)f_{t+1}^K) - \iota_N, \\ &= R_{t+1} Y_{t+1}' \Theta - \iota_N, \end{aligned}$$

where

$$\begin{aligned} \Theta' &= [\Theta_0', \Theta_1', \dots, \Theta_K'], \\ Y_{t+1}' &= [\mathbf{X}_t', \mathbf{X}_t' f_{t+1}^1, \dots, \mathbf{X}_t' f_{t+1}^K]. \end{aligned}$$

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

$$E[\mathbf{f}_{t+1}|\mathbf{Z}_t] = 0,$$

and therefore

$$E[(\mathbf{f}_{t+1} \otimes \mathbf{Z}_t)|\mathbf{Z}_t] = 0.$$

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

$$g_T(\Theta) = \frac{1}{T} \sum_{t=0}^{T-1} [\mathbf{f}_{t+1} \otimes \mathbf{X}_t].$$

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

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

with D and G being estimated by

$$D_T = \frac{1}{T} \sum_{t=0}^{T-1} [(R_{t+1} \otimes \mathbf{X}_t) Y_{t+1}'],$$

$$G_T = \frac{1}{T} \sum_{t=0}^{T-1} [(\mathbf{R}_{t+1} \otimes \mathbf{X}_t)(\mathbf{R}_{t+1} \otimes \mathbf{X}_t)']$$

➤ Finally, we get

$$g_T(\hat{\Theta}) = D_T \hat{\Theta} - (\iota_N \otimes \bar{X}),$$

where

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

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

HJ distance measure vs. the standard GMM

- Both the GMM and HJ distance measure are cross sectional tests of asset pricing models.
 1. Since the distance measure is formed using a weighting matrix that is invariant across competing SDF candidates it can be used to compare the performance of nested and non-nested asset pricing models.
 2. In the standard GMM the optimal weighting matrix S^{-1} varies across competing specifications.
 3. Therefore, the standard GMM cannot be used for comparing misspecification across competing models.

4. The HJ distance measure avoids the pitfall embedded in the standard GMM of favoring pricing models that produce volatile pricing errors.
5. In the HJ distance measure the weighting matrix is not a function of the parameters, which may result in a more stable estimation procedure.
6. On the other hand, the optimal GMM provides the most efficient estimate among estimates that use linear combinations of pricing errors as moments, in the sense that the estimated parameters have the smallest asy. covariance
7. While the idea looks neat be careful using the HJ distance measure.
8. For one, the more factors you throw in (as meaningless as they can get) the smaller the distance to the extent that artificial factors are not rejected by the test.
9. Moreover, if one model displays smaller distance it is considered better. That is imprecise as the gap of distance measures is a random quantity.
10. Indeed, the HJ statistic applies to large samples while small sample performance is troublesome.
11. In my opinion, Bayesian methods are more plausible to test and compare models. See in particular, Avramov and Chao (2006).
12. Bayesian econometrics is coming up next.

Spectral Analysis in Asset Pricing

- For a comprehensive coverage of the econometrics of spectral analysis it is suggested to consult the textbook of John H. Cochrane
http://faculty.chicagobooth.edu/john.cochrane/research/papers/time_series_book.pdf.
- Surprisingly, there are only a few papers in asset pricing implementing spectral analysis
- One of the first studies goes back to Daniel and Marshall (1998) – henceforth DM.
- There is a nice follow up work by Yu (2012).
- Here is the motivation of both studies.
- The sample correlation between the market excess return and consumption growth (on a quarterly frequency) is only 0.15.
- Such low correlation makes it difficult for consumption based models to match the data.
- Low correlation is often attributed to short term frictions including transaction and adjustment costs.
- Such factors could be meaningful in high frequencies (short horizons) but they should not interrupt the model over low frequencies (long horizons)
- DM and Yu perform coherence analysis of the consumption growth and excess market return
- Essentially, what the coherence analysis does is to split each of the two series into a set of Fourier (periodic) components at different frequencies, and then to determine the correlation of a set of Fourier components for the two series around each frequency.

- Such correlation is not a single number but rather varies at each frequency
- Since coherence is always positive, the sign of the correlation at different frequencies cannot be determined from the coherence spectrum.
- To identify the sign of the correlation, the co-spectrum can be examined.
- The co-spectrum at frequency ω can be interpreted as the portion of the covariance between consumption growth and asset returns that is attributable to cycles with frequency ω .
- Since the covariance can be positive or negative, the co-spectrum can also be positive or negative.
- Spectral analysis also yields the phase relation between the two series, which is a measure of how far the series must be shifted to maximize the correlation of the sets of Fourier components.
- The slope of the phase spectrum at any frequency ω is the group delay at frequency ω and precisely measures the number of leads or lags between consumption growth and asset returns.
- When this slope is positive, consumption leads the market return and vice versa.
- Therefore, the coherence, co-spectrum, and phase spectrum provide a convenient tool for analyzing the lead-lag relation and the correlations at different frequencies between time series.
- DM show that while there is a complete lack of correlation between asset returns and consumption growth at high frequencies, the coherence/correlation between the two series at lower frequencies is above 60%.

- Yu shows how the presence of a persistent habit process leads to an attenuated correlation between consumption and returns at low frequencies.
- Specifically, he shows that as long as the external habit model produces a countercyclical risk premium or a pro-cyclical price-dividend ratio, the model implies that the covariation between consumption and returns is greater in high-frequency components, whereas in the data, the opposite occurs.
- Parker and Julliard (2006) build on the long run correlation to revive the CCAPM in the cross section of average return.
- Rather than measuring risk by the contemporaneous covariance of an asset's return and consumption growth, they measure risk by the covariance of an asset's return and consumption growth cumulated over many quarters.
- They find that while contemporaneous consumption risk explains little of the variation in average returns across the 25 Fama-French portfolios, their measure of ultimate consumption risk at a horizon of three years explains a large fraction of this variation
- While they don't use spectral analysis their paper is motivated by the findings of DM
- Otrok, Ravikumarb, and Whiteman (2002) is another paper that implements spectral analysis to habit-formation preferences.

- They show that habit agents are much more averse to high-frequency fluctuations than to low-frequency fluctuations, and further, the relatively high equity premium in the habit model is determined by a relatively insignificant amount of high-frequency volatility in U.S. consumption.
- Otrok, Ravikumar, and Whiteman (2002) and Yu (2012) basically ask how the price of a consumption claim depends on the spectral density of consumption and its relation with returns.
- As noted by Dew-Becker and Giglio (2016) since the price of the asset reflects a combination of preferences and dynamics, it is impossible to evaluate the relative importance of the two.
- Instead, Dew-Becker and Giglio quantify preferences over the dynamics of shocks by deriving frequency-specific risk prices that capture the price of risk of consumption fluctuations at each frequency
- The frequency-specific risk prices are derived analytically for leading models.
- The decomposition helps measure the importance of economic fluctuations at different frequencies
- They precisely quantify the meaning of long-run in the context of Epstein-Zin preferences and measure the exact relevance of business-cycle fluctuations.
- Last, they estimate frequency-specific risk prices and show that cycles longer than the business cycle long-run risks are significantly priced in the equity market

The cospectrum, phase, and long run risk (based on Jianfeng, Yu 2012)

- Recall, following the notation of Bansal and Yaron (2004) the consumption and dividend dynamics are

$$\begin{aligned}x_{t+1} &= \rho x_t + \varphi_e \sigma e_{t+1} \\g_{c,t+1} &= \mu + x_t + \sigma \eta_{t+1} \\g_{d,t+1} &= \mu_d + \phi x_t + \varphi_d \sigma u_{t+1}\end{aligned}$$

- Then we can show that when Epstein-Zin (1989) recursive preferences excess return on the dividend claim can be written as:

$$r_{t+1}^e \approx \bar{r} + \frac{\phi - \frac{1}{\psi}}{1 - k_{1m}\rho} k_{1m} \varphi_e \sigma e_{t+1} + \varphi_d \sigma u_{t+1}$$

where ψ is the IES, k_{im} is the constant in the CS log linearization and $\rho = \text{corr}(\eta_{t+1}, u_{t+1})$

- The spectral representations of $\{g_{c,t}\}, \{r_t^{ex}\}, \{x_t\}, \{e_t\}, \{u_t\}$, and $\{\eta_t\}$ are

$$dZ_{g_c} = e^{-i\lambda} dZ_x + \sigma dZ_\eta$$

$$dZ_x = \rho e^{-i\lambda} dZ_x + \varphi_e \sigma dZ_e$$

$$dZ_r = \frac{\phi - \frac{1}{\psi}}{1 - k_{1m}\rho} k_{1m} \varphi_e \sigma dZ_e + \varphi_d \sigma dZ_u$$

- Define $A_{1m} = \frac{\phi - \frac{1}{\psi}}{1 - k_{1m}\rho}$ rearrange, and solve for dZ_{g_c} and dZ_r to obtain:

$$dZ_{g_c} = \frac{\varphi_e \sigma e^{-i\omega}}{1 - \rho e^{-i\omega}} dZ_e + \sigma dZ_\eta$$

$$dZ_r = k_{1m} A_{1m} \varphi_e \sigma dZ_e + \varphi_d \sigma dZ_u$$

- Thus, the multivariate spectrum is given by

$$f_{rr} = (k_{1m}A_{1m}\varphi_e\sigma)^2 + (\varphi_d\sigma)^2$$

$$f_{gg} = \left| e^{-i\lambda} \frac{\varphi_e\sigma}{1-\rho e^{-iw}} \right|^2 + \sigma^2$$

$$\begin{aligned} f_{gr} &= \left(\frac{e^{-iw}}{1-\rho e^{-iw}} \right) k_{1m}A_{1m}\varphi_e^2\sigma^2 + \varphi_d\sigma^2\rho_{\eta u} \\ &= \left(\frac{e^{-iw}-\rho}{1+\rho^2-2\rho\cos(w)} \right) k_{1m}A_{1m}\varphi_e^2\sigma^2 + \varphi_d\sigma^2\rho_{\eta u} \end{aligned}$$

- Solving for the cospectrum $C_{sp}(w)$, the real part of the cross spectrum $f_{12}(w)$, yields:

$$C_{sp}(w) = \left(\frac{\cos(w) - \rho}{1 + \rho^2 - 2\rho\cos(w)} \right) k_{1m}A_{1m}\varphi_e^2\sigma^2 + \varphi_d\sigma^2\rho_{\eta u}$$

- Taking the derivative and rearranging the equation yield

$$C'_{sp}(w) = \frac{\sin(w)(\rho^2 - 1)}{(1 + \rho^2 - 2\rho\cos(w))^2} \cdot \frac{\phi - \frac{1}{\psi}}{1 - k_{1m}\rho} k_{1m}\phi_e^2\sigma^2$$

- From the expression of f_{gr} - under the assumption of $\rho_{\eta u} = 0$, we can solve for the phase spectrum $\phi_{12}(w)$:

$$\tan(\phi_{12}(w)) = \frac{-\sin(w) - \rho}{\cos(w) - \rho}$$

- Thus, taking the derivative, it follows that

$$\begin{aligned}\phi'_{12}(w) &\propto \frac{\partial \frac{-\sin(w) - \rho}{\cos(w) - \rho}}{\partial w} \\ &\propto -\cos(w)^2 + \rho \cos(w) - \sin(w)^2 - \rho \sin(w) \\ &= \rho(\cos(w) - \sin(w)) - 1 < 0\end{aligned}$$

where ' \propto ' denotes that both sides of ' \propto ' have the same sign. Thus, the phase spectrum is always decreasing

The cospectrum, phase, and Habit formation (based on Jianfeng, Yu 2012)

- The growth of log consumption is modeled as

$$g_{c,t} = \mu_c + \varepsilon_{c,t}$$

- The approximated excess return on dividend claim is

$$r_{t+1}^e \approx \alpha - \beta_S \sum_{j=1}^{\infty} \phi_S^{j-1} g_{c,t+1-j} + \beta_c \varepsilon_{c,t+1} + \beta_\delta \varepsilon_{\delta,t+1}$$

- Now let us understand the cospectrum and phase of the joint consumption return process.
- We first replace the consumption and return dynamics using inverse FT

$$dZ_{g_c}(w) = dZ_{\varepsilon_c}(w)$$

$$dZ_r(w) = -\beta_S \sum_{j=1}^{\infty} \phi_S^{j-1} \exp(-ijw) dZ_{g_c}(w) + \beta_c dZ_{\varepsilon_c}(w) + \beta_\delta dZ_{\varepsilon_\delta}(w)$$

► Notice that

$$\sum_{j=1}^{\infty} \phi_s^{j-1} \exp(-ijw) = \frac{\exp(-iw)}{1 - \phi_s \exp(-iw)}$$

► So:

$$dZ_r(w) = \left(\frac{-\beta_s \exp(-iw)}{1 - \phi_s \exp(-iw)} + \beta_c \right) dZ_{\varepsilon_c}(w) + \beta_\delta dZ_{\varepsilon_\delta}(w)$$

► Then, the multivariate spectrum is given by

$$2\pi f_{11}(w) = \sigma_c^2$$

$$2\pi f_{22}(w) = \left| \frac{-\beta_s \exp(-iw)}{1 - \phi_s \exp(-iw)} + \beta_c \right|^2 \sigma_c^2 + \beta_\delta^2 \sigma_\delta^2$$

$$+ 2\operatorname{Re} \left(\frac{-\beta_s \exp(-iw)}{1 - \phi_s \exp(-iw)} + \beta_c \right) \beta_\delta \sigma_{c\delta}$$

$$\begin{aligned}
2\pi f_{12}(w) &= \left(\frac{-\beta_s \exp(-iw)}{1 - \phi_s \exp(-iw)} + \beta_c \right)' \sigma_c^2 + \beta_\delta \sigma_{c\delta} \\
&= \left(\frac{\beta_s (\phi_s - \exp(iw))}{1 + \phi_s^2 - 2\phi_s \cos(w)} + \beta_c \right) \sigma_c^2 + \beta_\delta \sigma_{c\delta}
\end{aligned}$$

- For the cospectrum $C_{sp}(w)$ the real part of the cross spectrum $f_{12}(w)$

$$2\pi C_{sp}(w) = \left(\frac{\beta_s (\phi_s - \cos(w))}{1 + \phi_s^2 - 2\phi_s \cos(w)} + \beta_c \right) \sigma_c^2 + \beta_\delta \sigma_{c\delta}$$

- Therefore, the derivative of the cospectrum is

$$C'_{sp}(w) = \beta_s \sigma_c^2 \frac{\sin(w)(1 + \phi_s^2 - 2\phi_s \cos(w)) + 2\phi_s \sin(w)(\cos(w) - \phi_s)}{(1 + \phi_s^2 - 2\phi_s \cos(w))^2}$$

$$= \frac{\beta_s \sigma_c^2 \sin(w)}{(1 + \phi_s^2 - 2\phi_s \cos(w))^2} (1 - \phi_s^2)^2 \geq 0$$

and the portion of covariance contributed by components at frequency w is increasing in the frequency w .

- By definition, the coherence and the phase are, respectively

$$h(w) = \frac{|f_{12}|}{\sqrt{f_{11}f_{22}}}$$

$$\tan(\phi_{12}(w)) = \frac{\frac{\beta_s \sin(w)}{1 + \phi_s^2 - 2\phi_s \cos(w)} \sigma_c^2}{\left(\frac{\beta_s (\phi_s - \cos(w))}{1 + \phi_s^2 - 2\phi_s \cos(w)} + \beta_c \right) \sigma_c^2 + \beta_\delta \sigma_{c\delta}}$$

- At the frequency $w=0$, the cospectrum is

$$\begin{aligned} C_{sp}(0) &= \left(-\beta_s \frac{1-\phi_s}{1+\phi_s^2-2\phi_s} + \beta_c \right) \sigma_c^2 + \beta_\delta \sigma_{c\delta} \\ &= \left(1 - a_1 k_1 - \frac{a_1(1-k_1)}{\bar{S}(1-\phi_s)} \right) \sigma_c^2 + \beta_\delta \sigma_{c\delta} \end{aligned}$$

- Therefore, the low-frequency correlation between consumption growth and asset returns is negative if and only if

$$\left(1 - a_1 k_1 - \frac{a_1(1 - k_1)}{\bar{S}(1 - \phi_s)}\right) + \beta_\delta \frac{\sigma_{c\delta}}{\sigma_c^2} < 0$$

- By differentiating the following equation

$$\tan(\phi_{12}(w)) = \frac{\frac{\beta_s \sin(w)}{1 + \phi_s^2 - 2\phi_s \cos(w)} \sigma_\varepsilon^2}{\left(\frac{\beta_s(\phi_s - \cos(w))}{1 + \phi_s^2 - 2\phi_s \cos(w)} + \beta_c\right) \sigma_c^2 + \beta_\delta \sigma_{c\delta}}$$

the sign of the slope of the phase spectrum can be examined

- To see this, start with

$$\phi'_{12}(w)$$

$$\propto -[\beta_s(\phi_s - \cos(w))\sigma_c^2 + (\beta_c\sigma_c^2 + \beta_\delta\sigma_{c\delta})(1 + \phi_s^2 - 2\phi_s\cos(w))] \cdot \beta_s\cos(w) \\ + \beta_s\sin(w)[\beta_s\sin(w)\sigma_c^2 + 2\phi_s(\beta_c\sigma_c^2 + \beta_\delta\sigma_{c\delta})\sin(w)]$$

where \propto denotes that both sides of \propto have the same sign.

- Rearrange and simplify to obtain

$$\phi'_{12}(w)$$

$$\propto -\frac{a_1(k_1\phi_s)-1}{\bar{S}} + 2\phi_s \left(1 + a_1k_1 \frac{1-\bar{S}}{\bar{S}}\right) + 2\phi_s\beta_\delta \frac{\sigma_{c\delta}}{\sigma_c^2} \\ - \left\{1 + a_1k_1 \frac{1-\bar{S}}{\bar{S}} + \beta_\delta \frac{\sigma_{c\delta}}{\sigma_c^2} + \phi_s^2 + \frac{-a_1k_1\bar{S}\phi_s^2}{\bar{S}} + \frac{a_1\phi_s}{\bar{S}} + \beta_\delta\phi_s^2 \frac{\sigma_{c\delta}}{\sigma_c^2}\right\} \cos(w) \\ \geq -\left[1 - a_1k_1 - a_1 \frac{1-k_1}{\bar{S}(1-\phi_s)} + \beta_\delta \frac{\sigma_{c\delta}}{\sigma_c^2}\right] (1 - \phi_s)^2$$

- The inequality above requires the assumption

$$(1 - a_1 k_1)(1 + \phi_s^2) + \frac{a_1 k_1}{\bar{S}} + \frac{a_1 \phi_s}{\bar{S}} + \beta_\delta (1 + \phi_s^2) \frac{\sigma_{c\delta}}{\sigma_c^2} > 0$$

which is true if the correlation between the innovations of return and consumption is positive.

- Thus, the phase spectrum is increasing as long as

$$\left[1 - a_1 k_1 - \frac{a_1 (1 - k_1)}{\bar{S} (1 - \phi_s)} + \beta_\delta \frac{\sigma_{c\delta}}{\sigma_c^2} \right] + \beta_\delta \frac{\sigma_{c\delta}}{\sigma_c^2} < 0$$

and the correlation between the innovations of return and consumption is positive

Appendix

Quality Investing using Return on Invested Capital

► **The Little Book That Beats the Market - by Joel Greenblatt:** The idea here is to combine only two financial ratios – earnings yield ($\text{EBIT} / \text{enterprise value}$) and return on capital ($\text{EBIT} / \text{net fixed assets plus working capital}$). Greenblatt suggests the “magic formula”: purchasing 30 cheap stocks with a high earnings yield and a high return on capital. The receipt is below.

1. Decide on minimum market capitalization (usually greater than \$50 million).
2. Exclude utility and financial stocks.
3. Exclude foreign companies (American Depositary Receipts).
4. Compute company's earnings yield = $\text{EBIT} / \text{enterprise value}$.
5. Compute company's return on capital = $\text{EBIT} / (\text{net fixed assets} + \text{working capital})$.
6. Rank all companies above the threshold market capitalization by highest earnings yield and highest return on capital.
7. Invest in 20–30 highest ranked companies.
8. Re-balance portfolio once per year, selling losers one week before the year-mark and winners one week after the year mark.

Quality investing based on the G Score

- **G Score:** The G-Score is due to Mohanram (2005). It combines traditional fundamentals, such as earnings and cash flows, with measures tailored for growth firms, such as earnings stability, growth stability and intensity of R&D, capital expenditure and advertising. A long–short strategy based on GSCORE earns significant excess returns, though most of the returns come from the short side. Thus, to form an attractive trading strategy one could take long positions based on the F-Score or the F-score combined with the book-to-market ratio and short position based on the G-score.
- The formation of the G-score based on 8 binary variables as follow:
 - G1 is equal 1 if a firm's ROA is greater than the contemporaneous median ROA for all low BM firms in the same industry and 0 otherwise. ROA, defined as the ratio of net income before extraordinary items scaled by average total assets.
 - G2=1 if a firm's cash flow ROA exceeds the contemporaneous median for all low BM firms in the same industry and 0 otherwise. Cash flows ROA is similar to the above-defined ROA except that operating cash flows replace net income.
 - G3=1 if a firm's cash flow from operations exceeds net income and 0 otherwise.

Quality investing based on the G-Score

- $G4 = 1$ if a firm's earnings variability is less than the contemporaneous median for all low BM firms in the same industry and 0 otherwise.
- $G5 = 1$ if a firm's sales growth variability is less than the contemporaneous median for all low BM firms in the same industry and 0 otherwise.
- $G6$, $G7$ and $G8$ are defined to equal 1 if a firm's R&D, capital expenditure and advertising intensity respectively, are greater than the contemporaneous medians of the corresponding variables for all low BM firms in the same industry and 0 otherwise. The intensity of R&D, capital expenditure and advertising are measured by deflating these variables by beginning assets.
- The signals relating to profitability and cash flows ($G1:G3$) as well as those related to conservatism ($G6:G8$) are created using the annualized financials. The two signals earnings variability and sales growth variability ($G4, G5$) are generated from quarterly financials of the past 4 years, with the constraint that at least six quarters information be available. While quarterly information might induce variability owing to seasonality, the industry adjustment should mitigate this.

External Financing

- Richardson and Sloan (2003) nicely summarize all external financing transactions in one measure.
- They show that their comprehensive measure of external financing has a stronger relation with future returns relative to measures based on individual transactions.
- The external financing measure, denoted by $\Delta XFIN$ is the total cash received from issuance of new debt and equity offerings minus cash used for retirement of existing debt and equity. All components are normalized by the average value of total assets.
- This measure considers all sorts of equity offerings including common and preferred stocks as well as all sorts of debt offerings including straight bonds, convertible bonds, bank loans, notes, etc. Interest payments on debt as well as dividend payments on preferred stocks are not considered as retiring debt or equity. However, dividend payments on common stocks are considered as retiring equity. In essence, dividends on common stocks are treated as stock repurchases.
- The $\Delta XFIN$ measure can be decomposed as

$$\Delta XFIN = \Delta CEquity + \Delta PEquity + \Delta Debt$$

where:

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

Downside Risk

- Downside risk is the financial risk associated with losses.
- There are various down side risk measures which quantify the risk of losses, the expected loss given the realization of a loss, or even the worst case scenario characterizing a particular investment.
- All down side risk measures exclusively focus on the left tail of the return distribution, whereas volatility measures are both about the upside and downside outcomes.
- Typical downside risk measures include the Value at Risk (VaR), expected shortfall, semi-variance, maximum drawdown, downside beta, and shortfall probability. To establish the trading strategy one can focus on VaR, which is a very well-used measure in risk management.
- Down side risk measures are often positively and often negatively associated with average returns.

Tail risk as a common factor and equity premium predictor

- Kelly and Jiang (2014) propose the tail exponent as both a common factor and a predictor of the equity premium.
- The tail exponent is an aggregate variable constructed based on daily returns.
- In particular, consider daily returns of all stocks within a particular month and identify the *5th* percentile of the cross sectional distribution or the return threshold.
- Then only for those daily returns which fall below the return threshold take the simple average of the natural log of return divided by the return threshold

$$\lambda_t = \frac{1}{k_t} \sum_{k=1}^{k_t} \ln \frac{r_{kt}}{u_t}$$

where k_t is the number of exceedances, u_t is the return threshold, r_{kt} is the daily return that falls below the threshold.

- Perhaps an open question would be how tail risk is associated with other anomalous patterns in the cross section of average returns.

Proof of result on the maximum Sharpe ratio

- If the SDF is conditionally log-normally distributed, hence we can apply the general formula:

$$SR_t = \sqrt{\exp(\text{Var}_r(\log M_{t+1}) - 1)} \approx \sigma_t(\log M_{t+1})$$

Proof: use the log-normal formula

$$E(\exp(X)) = \exp\left(E(X) + \frac{1}{2}\text{Var}(X)\right)$$

and compute

$$E_t(M_{t+1}) = \exp\left(E_t(\log M_{t+1}) + \frac{1}{2}\text{Var}_t(\log M_{t+1})\right)$$

$$\begin{aligned}\text{Var}_t(M_{t+1}) &= E_t(M_{t+1}^2) - [E_t(M_{t+1})]^2 \\ &= \exp(2E_t(\log M_{t+1}) + 2\text{Var}_t(\log M_{t+1})) \\ &\quad - \exp(2E_t(\log M_{t+1}) + \text{Var}_t(\log M_{t+1}))\end{aligned}$$

The result follows.

References

- ▶ Abel, Andrew B. 1990. *Asset Prices Under Habit Formation and Catching Up with the Joneses*.
- ▶ Amihud, Yakov. 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5 (1): 31-56.
- ▶ Amihud, Yakov, and Clifford M. Hurvich. 2004. Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis* 39 (4): 813-41.
- ▶ Amihud, Yakov, Clifford M. Hurvich, and Yi Wang. 2008. Multiple-predictor regressions: Hypothesis testing. *The Review of Financial Studies* 22 (1): 413-34.
- ▶ Andrews, Donald WK. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*: 817-58.
- ▶ Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2009. High idiosyncratic volatility and low returns: International and further US evidence. *Journal of Financial Economics* 91 (1): 1-23.
- ▶ Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2006. The cross-section of volatility and expected returns. *The Journal of Finance* 61 (1): 259-99.

- Antoniou, Constantinos, John A. Doukas, and Avanidhar Subrahmanyam. 2010. Investor sentiment and price momentum. *Work.Pap., Cent.Empir.Res.Finance, Durham Univ./Old Dominion Univ./Judge Bus.Sch., Cambridge Univ.*
- Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. Value and momentum everywhere. *The Journal of Finance* 68 (3): 929-85.
- Avramov, Doron. 2003. Stock return predictability and asset pricing models. *The Review of Financial Studies* 17 (3): 699-738.
- Avramov, Doron. 2002. Stock return predictability and model uncertainty. *Journal of Financial Economics* 64 (3): 423-58.
- Avramov, Doron, and Yan Xu, 2010. Predicting currency returns: New evidence on the forward premium puzzle and the dollar-trade strategy.
- Avramov, Doron, Scott Cederburg, and Katarina Kvasnakova, 2017. On the Riskiness of Stocks for the Long Run: Taking Cues from Economic Theory. *The Review of Financial Studies* Forthcoming
- Avramov, Doron, Scott Cederburg, and Katarina Lucivjanska. 2016. Are stocks riskier over the long run? taking cues from economic theory.
- Avramov, Doron, and John C. Chao. 2006. An exact bayes test of asset pricing models with application to international markets. *The Journal of Business* 79 (1): 293-324.

- Avramov, Doron, Si Cheng, and Allaudeen Hameed. 2016. Mutual funds and mispriced stocks.
- Avramov, Doron, Si Cheng, and Allaudeen Hameed. 2016. Time-varying liquidity and momentum profits. *Journal of Financial and Quantitative Analysis* 51 (6): 1897-923.
- Avramov, Doron, and Tarun Chordia. 2006. Asset pricing models and financial market anomalies. *The Review of Financial Studies* 19 (3): 1001-40.
- Avramov, Doron, and Tarun Chordia. 2006. Predicting stock returns. *Journal of Financial Economics* 82 (2): 387-415.
- Avramov, Doron, Tarun Chordia, and Amit Goyal. 2006a. Liquidity and autocorrelations in individual stock returns. *The Journal of Finance* 61 (5): 2365-94.
- Avramov, Doron, Tarun Chordia, and Amit Goyal. 2006b. The impact of trade on daily volatility. *The Review of Financial Studies* 19, 1241-1277.
- Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov. 2017. Bonds, stocks, and sources of mispricing.
- Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov. 2013. Anomalies and financial distress. *Journal of Financial Economics* 108 (1): 139-59.
- Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov. 2009. Credit ratings and the cross-section of stock returns. *Journal of Financial Markets* 12 (3): 469-99.

- Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov. 2009. Dispersion in analysts' earnings forecasts and credit rating. *Journal of Financial Economics* 91 (1): 83-101.
- Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov. 2007. Momentum and credit rating. *The Journal of Finance* 62 (5): 2503-20.
- Avramov, Doron, and Satadru Hore. 2017. Cross-sectional factor dynamics and momentum returns. *Journal of Financial Markets* 32 : 69-96.
- Avramov, Doron, Guy Kaplanski, and Avanidhar Subrahmanyam, 2019, The Predictability of Equity Returns from Past Returns: A New Moving Average-Based Perspective. Working Paper.
- Avramov, Doron, Robert Kosowski, Narayan Y. Naik, and Melvyn Teo. 2011. Hedge funds, managerial skill, and macroeconomic variables. *Journal of Financial Economics* 99 (3): 672-92.
- Avramov, Doron, and Russ Wermers. 2006. Investing in mutual funds when returns are predictable. *Journal of Financial Economics* 81 (2): 339-77.
- Avramov, Doron, and Guofu Zhou. 2010. Bayesian portfolio analysis. *Annu.Rev.Financ.Econ.* 2 (1): 25-47.
- Bailey, Warren, Alok Kumar, and David Ng. 2011. Behavioral biases of mutual fund investors. *Journal of Financial Economics* 102 (1): 1-27.

- Baks, Klaas P., Andrew Metrick, and Jessica Wachter. 2001. Should investors avoid all actively managed mutual funds? A study in bayesian performance evaluation. *The Journal of Finance* 56 (1): 45-85.
- Bakshi, Gurdip, and Dilip Madan. 2000. Spanning and derivative-security valuation,, *Journal of Financial Economics* 55:205-238.
- Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan. 2003. Stock return characteristics, skew laws and the differential pricing of individual equity options. *Review of Financial Studies* 16: 101-143.
- Ball, Ray, and Philip Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research*: 159-78.
- Bansal, Ravi. 2007. *Long-Run Risks and Financial Markets*.
- Bansal, Ravi, Robert Dittmar, and Dana Kiku. 2007. Cointegration and consumption risks in asset returns. *The Review of Financial Studies* 22 (3): 1343-75.
- Bansal, Ravi, Dana Kiku, and Amir Yaron. 2009. *An Empirical Evaluation of the Long-Run Risks Model for Asset Prices*.
- Bansal, Ravi, and Amir Yaron. 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance* 59 (4): 1481-509.
- Banz, Rolf W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9 (1): 3-18.
- Barberis, Nicholas. 2000. Investing for the long run when returns are predictable. *The Journal of Finance* 55 (1): 225-64.

- Barberis, Nicholas, and Ming Huang. 2008. Stocks as lotteries: The implications of probability weighting for security prices. *The American Economic Review* 98 (5): 2066-100.
- Barberis, Nicholas, Ming Huang, and Tano Santos. 2001. Prospect theory and asset prices. *The Quarterly Journal of Economics* 116 (1): 1-53.
- Barberis, Nicholas, Abhiroop Mukherjee, and Baolian Wang. 2016. Prospect theory and stock returns: An empirical test. *The Review of Financial Studies* 29 (11): 3068-107.
- Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. A model of investor sentiment. *Journal of Financial Economics* 49 (3): 307-43.
- Barro, Robert J. 2005. *Rare Events and the Equity Premium*.
- Basu, Sanjoy. 1983. The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. *Journal of Financial Economics* 12 (1): 129-56.
- Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein. 1979. Estimation risk and optimal portfolio choice. *NORTH-HOLLAND PUBL.CO., N.Y., 190 Pp.*
- Beeler, Jason, and John Y. Campbell. 2009. *The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment*.
- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, 2014, Inference on treatment effects after selection among high-dimensional controls, *The Review of Economic Studies* 81, 608-650.

- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, 2014, High-dimensional methods and inference on structural and treatment effects, *Journal of Economic Perspectives* 28, 29-50.
- Berk, Jonathan B., Richard C. Green, and Vasant Naik. 1999. Optimal investment, growth options, and security returns. *The Journal of Finance* 54 (5): 1553-607.
- Binsbergen, Jules van and Ralph SJ Koijen. 2010. Predictive regressions: A present-value approach. *The Journal of Finance* 65 (4): 1439-71.
- Boehmer, Ekkehart, Zsuzsa R. Huszar, and Bradford D. Jordan. 2010. The good news in short interest. *Journal of Financial Economics* 96 (1): 80-97.
- Bossaerts, Peter, and Pierre Hillion. 1999. Implementing statistical criteria to select return forecasting models: What do we learn? *The Review of Financial Studies* 12 (2): 405-28.
- Boudoukh, Jacob, Matthew Richardson, and Robert F. Whitelaw. 2006. The myth of long-horizon predictability. *The Review of Financial Studies* 21 (4): 1577-605.
- Breeden, Douglas T. 1984. Futures markets and commodity options: Hedging and optimality in incomplete markets. *Journal of Economic Theory* 32 (2): 275-300.
- Breeden, Douglas T. 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics* 7 (3): 265-96.
- Breeden, Douglas T., Michael R. Gibbons, and Robert H. Litzenberger. 1989. Empirical tests of the consumption-oriented CAPM. *The Journal of Finance* 44 (2): 231-62.

- Breeden, Douglas T., and Robert H. Litzenberger. 1978. Prices of state-contingent claims implicit in option prices. *Journal of Business*: 621-51.
- Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam. 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics* 49 (3): 345-73.
- Brown, Stephen J. 1978. The portfolio choice problem: Comparison of certainty equivalence and optimal bayes portfolios. *Communications in Statistics-Simulation and Computation* 7 (4): 321-34.
- Brown, Stephen J. 1976. Optimal portfolio choice under uncertainty: A bayesian approach. *Ph.D.Dissertation, Graduate School of Business, University of Chicago*.
- Campbell, John Y. 1992. *Intertemporal Asset Pricing without Consumption Data*.
- Campbell, John Y. 1987. Stock returns and the term structure. *Journal of Financial Economics* 18 (2): 373-99.
- Campbell, John Y., and John Ammer. 1993. What moves the stock and bond markets? A variance decomposition for long-term asset returns. *The Journal of Finance* 48 (1): 3-37.

- Campbell, John Y., and John H. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107 (2): 205-51.
- Campbell, John Y., Jens Hilscher, and Jan Szilagyi. 2008. In search of distress risk. *The Journal of Finance* 63 (6): 2899-939.
- Campbell, John Y., and Robert J. Shiller. 1988. The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies* 1 (3): 195-228.
- Campbell, John Y., and Samuel B. Thompson. 2007. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies* 21 (4): 1509-31.
- Campbell, John Y., and Tuomo Vuolteenaho. 2004. Bad beta, good beta. *The American Economic Review* 94 (5): 1249-75.
- Carhart, Mark M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52 (1): 57-82.
- Carr, P., and D. Madan. 2001. Optimal position in derivative securities. *Quantitative Finance* 1: 19-37.
- Chan, Yeung Lewis, and Leonid Kogan. 2002. Catching up with the joneses: Heterogeneous preferences and the dynamics of asset prices. *Journal of Political Economy* 110 (6): 1255-85.
- Chamberlain, Gary, and Michael Rothschild, 1982, Arbitrage, factor structure, and mean-variance analysis on large asset markets, NBER Working Paper.

- Chen, Hong-Yi, Sheng-Syan Chen, Chin-Wen Hsin, and Cheng-Few Lee. 2010. Price, earnings, and revenue momentum strategies. *Electronic Copy Available at: [Http://ssrn.com/abstract](http://ssrn.com/abstract) 1571883* .
- Chen, Long, and Xinlei Zhao. 2009. Return decomposition. *The Review of Financial Studies* 22 (12): 5213-49.
- Chen, Nai-fu. 1991. Financial investment opportunities and the macroeconomy. *The Journal of Finance* 46 (2): 529-54.
- Chen, Nai-Fu, Richard Roll, and Stephen A. Ross. 1986. Economic forces and the stock market. *Journal of Business*: 383-403.
- Chordia, Tarun, and Lakshmanan Shivakumar. 2002. Momentum, business cycle, and time-varying expected returns. *The Journal of Finance* 57 (2): 985-1019.
- Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong. 2014. Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics* 58 (1): 41-58.
- Chui, Andy CW, Sheridan Titman, and KC John Wei. 2010. Individualism and momentum around the world. *The Journal of Finance* 65 (1): 361-92.

- Cochrane, John H. 2007. The dog that did not bark: A defense of return predictability. *The Review of Financial Studies* 21 (4): 1533-75.
- Cochrane, John H. 2001. Asset pricing, 2001.
- Cochrane, John H. 2001. Long-Term debt and optimal policy in the fiscal theory of the price level. *Econometrica* 69 (1): 69-116.
- Cohen, Lauren, and Andrea Frazzini. 2008. Economic links and predictable returns. *The Journal of Finance* 63 (4): 1977-2011.
- Coibion, Olivier, and Yuriy Gorodnichenko. 2012. What Can Survey Forecasts Tell Us about Information Rigidities? *Journal of Political Economy* 120 (1): 116-159.
- Coibion, Olivier, and Yuriy Gorodnichenko. 2015a. Is The Phillips Curve Alive and Well After All? Inflation Expectations and the Missing Disinflation. *American Economic Journal: Macroeconomics* 7 (1), 197-232.
- Coibion, Olivier, and Yuriy Gorodnichenko. 2015b. Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts. *American Economic Review* 105 (8), 2644–2678.
- Connor, Gregory, and Robert A. Korajczyk. 1988. Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics* 21 (2): 255-89.
- Conrad, Jennifer, Nishad Kapadia, and Yuhang Xing. 2014. Death and jackpot: Why do individual investors hold overpriced stocks? *Journal of Financial Economics* 113 (3): 455-75.
- Constantinides, George M. 1990. Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy* 98 (3): 519-43.
- Cooper, Michael J., Huseyin Gulen, and Michael J. Schill. 2008. Asset growth and the cross-section of stock returns. *The Journal of Finance* 63 (4): 1609-51.

- Cooper, Michael J., Roberto C. Gutierrez, and Allaudeen Hameed. 2004. Market states and momentum. *The Journal of Finance* 59 (3): 1345-65.
- Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam. 1998. Investor psychology and security market under-and overreactions. *The Journal of Finance* 53 (6): 1839-85.
- Daniel, Kent, and David Marshall. 1997. Equity-premium and risk-free-rate puzzles at long horizons. *Macroeconomic Dynamics* 1 (2): 452-84.
- Daniel, Kent, and Tobias J. Moskowitz. 2016. Momentum crashes. *Journal of Financial Economics* 122 (2): 221-47.
- De Bondt, Werner FM, and Richard H. Thaler. 1987. Further evidence on investor overreaction and stock market seasonality. *Journal of Finance*: 557-81.
- De Bondt, Werner FM, and Richard Thaler. 1985. Does the stock market overreact? *The Journal of Finance* 40 (3): 793-805.

- De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance* 45 (2): 379-95.
- Detemple, Jerome, and Shashidhar Murthy. 1994. Intertemporal asset pricing with heterogeneous beliefs. *Journal of Economic Theory* 62 (2): 294-320.
- Dew-Becker, Ian, and Stefano Giglio. 2016. Asset pricing in the frequency domain: Theory and empirics. *The Review of Financial Studies* 29 (8): 2029-68.
- Diamond, Douglas W., and Robert E. Verrecchia. 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics* 18 (2): 277-311.
- Dichev, Ilia D. 1998. Is the risk of bankruptcy a systematic risk? *The Journal of Finance* 53 (3): 1131-47.
- Dichev, Ilia D., and Joseph D. Piotroski. 2001. The Long-Run stock returns following bond ratings changes. *The Journal of Finance* 56 (1): 173-203.
- Diether, Karl B., Christopher J. Malloy, and Anna Scherbina. 2002. Differences of opinion and the cross section of stock returns. *The Journal of Finance* 57 (5): 2113-41.
- Duffie, Darrell, and Larry G. Epstein. 1992. Stochastic differential utility. *Econometrica: Journal of the Econometric Society*: 353-94.

- Epstein, Larry G., and Stanley E. Zin. 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica: Journal of the Econometric Society*: 937-69.
- Fama, Eugene F. 1998. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics* 49 (3): 283-306.
- Fama, Eugene F. 1991. Efficient capital markets: II. *The Journal of Finance* 46 (5): 1575-617.
- Fama, Eugene F., and Kenneth R. French. 2016. Dissecting anomalies with a five-factor model. *The Review of Financial Studies* 29 (1): 69-103.
- Fama, Eugene F., and Kenneth R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116 (1): 1-22.
- Fama, Eugene F., and Kenneth R. French. 2008. Dissecting anomalies. *The Journal of Finance* 63 (4): 1653-78.
- Fama, Eugene F., and Kenneth R. French. 2006. Profitability, investment and average returns. *Journal of Financial Economics* 82 (3): 491-518.

- Fama, Eugene F., and Kenneth R. French. 1996. Multifactor explanations of asset pricing anomalies. *The Journal of Finance* 51 (1): 55-84.
- Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3-56.
- Fama, Eugene F., and Kenneth R. French. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47 (2): 427-65.
- Fama, Eugene F., and Kenneth R. French. 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25 (1): 23-49.
- Fama, Eugene F., and James D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81 (3): 607-36.
- Fan, J. and Li, R. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* 96, 1348–1360.
- Fan, J. and Peng, H. 2004. Nonconcave penalized likelihood with a diverging number of parameters, *Ann. Statist.* 32, 928–961.
- Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2017, Taming the factor zoo, Working Paper.
- Ferson, Wayne E., and Campbell R. Harvey. 1999. Conditioning variables and the cross section of stock returns. *The Journal of Finance* 54 (4): 1325-60.
- Ferson, Wayne E., and Campbell R. Harvey. 1991. The variation of economic risk premiums. *Journal of Political Economy* 99 (2): 385-415.
- Ferson, Wayne E., and Robert A. Korajczyk. 1995. Do arbitrage pricing models explain the predictability of stock returns? *Journal of Business*: 309-49.

- Ferson, Wayne, Suresh Nallareddy, and Biqin Xie. 2013. The “out-of-sample” performance of long run risk models. *Journal of Financial Economics* 107 (3): 537-56.
- Friend, Irwin, and Marshall Blume, 1970, Measurement of portfolio performance under uncertainty, *The American Economic Review* 60, 561-575.
- Foster, F. Douglas, Tom Smith, and Robert E. Whaley. 1997. Assessing goodness-of-fit of asset pricing models: The distribution of the maximal R². *The Journal of Finance* 52 (2): 591-607.
- Frazzini, Andrea, and Lasse H. Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1-25.
- Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2017, Dissecting characteristics nonparametrically, Working Paper.
- Friewald, Nils, Christian Wagner, and Josef Zechner. 2014. The Cross-Section of credit risk premia and equity returns. *The Journal of Finance* 69 (6): 2419-69.
- Fu, Wenjiang J., 1998, Penalized regressions: the bridge versus the lasso, *Journal of computational and graphical statistics* 7, 397-416.
- Fu, Fangjian. 2009. Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics* 91 (1): 24-37.
- Gabaix, Xavier. 2009. Power laws in economics and finance. *Annu.Rev.Econ.* 1 (1): 255-94.
- Garlappi, Lorenzo, Tao Shu, and Hong Yan. 2006. Default risk, shareholder advantage, and stock returns. *The Review of Financial Studies* 21 (6): 2743-78.
- Garlappi, Lorenzo, and Hong Yan. 2011. Financial distress and the cross-section of equity returns. *The Journal of Finance* 66 (3): 789-822.

- Gebhardt, William R., Soeren Hvidkjaer, and Bhaskaran Swaminathan. 2005. Stock and bond market interaction: Does momentum spill over? *Journal of Financial Economics* 75 (3): 651-90.
- Geczy, Christopher, and Mikhail Samonov. 2013. 212 years of price momentum (the World's longest backtest: 1801–2012). *Available at SSRN*.
- Geczy, Christopher, Robert F. Stambaugh, and David Levin. 2005. Investing in socially responsible mutual funds.
- Geweke, John, and Guofu Zhou. 1996. Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies* 9 (2): 557-87.
- Gibbons, Michael R., Stephen A. Ross, and Jay Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*: 1121-52.
- Gomes, Joao, Leonid Kogan, and Lu Zhang. 2003. Equilibrium cross section of returns. *Journal of Political Economy* 111 (4): 693-732.
- Greenblatt, Joel. 2006. *The little book that beats the market* John Wiley & Sons.
- Greenig, Douglas. 1986. Non-separable preferences, stochastic returns and intertemporal substitution in consumption.

- Grossman, Sanford J., Angelo Melino, and Robert J. Shiller. 1987. Estimating the continuous-time consumption-based asset-pricing model. *Journal of Business & Economic Statistics* 5 (3): 315-27.
- Grundy, Bruce D., and J. Spencer Martin Martin. 2001. Understanding the nature of the risks and the source of the rewards to momentum investing. *The Review of Financial Studies* 14 (1): 29-78.
- Gu, Shihao, Bryan T. Kelly, and Dacheng Xiu, 2018, Empirical Asset Pricing Via Machine Learning, Working Paper.
- Hansen, Lars Peter. 1982. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*: 1029-54.
- Hansen, Lars P., John C. Heaton, and Nan Li, 2008, Consumption strikes back? Measuring long-run risk, *Journal of Political economy* 116, 260-302.
- Hansen, Lars Peter, and Ravi Jagannathan. 1997. Assessing specification errors in stochastic discount factor models. *The Journal of Finance* 52 (2): 557-90.
- Hansen, Lars Peter, and Scott F. Richard. 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica: Journal of the Econometric Society*: 587-613.
- Hansen, Lars P., and Thomas J. Sargent, 2005, Robust estimation and control under commitment, *Journal of Economic Theory* 124, 258-301.
- Hansen, Lars Peter, and Kenneth J. Singleton. 1983. Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy* 91 (2): 249-65.

- Hansen, Lars Peter, and Kenneth J. Singleton. 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society*: 1269-86.
- Harrison, J. Michael, and David M. Kreps. 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20 (3): 381-408.
- Harrison, J. Michael, and David M. Kreps. 1978. Speculative investor behavior in a stock market with heterogeneous expectations. *The Quarterly Journal of Economics* 92 (2): 323-36.
- Harvey, Campbell R. 1991. The world price of covariance risk. *The Journal of Finance* 46 (1): 111-57.
- Harvey, Campbell R. 1989. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics* 24 (2): 289-317.
- Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2016. ... And the cross-section of expected returns. *The Review of Financial Studies* 29 (1): 5-68.
- Harvey, Campbell R., and Akhtar Siddique. 2000. Conditional skewness in asset pricing tests. *The Journal of Finance* 55 (3): 1263-95.
- Hastings, W. Keith. 1970. Monte carlo sampling methods using markov chains and their applications. *Biometrika* 57 (1): 97-109.

- Haugen, Robert A., and Nardin L. Baker. 1996. Commonality in the determinants of expected stock returns. *Journal of Financial Economics* 41 (3): 401-39.
- Heaton, John, and Deborah Lucas. 1992. The effects of incomplete insurance markets and trading costs in a consumption-based asset pricing model. *Journal of Economic Dynamics and Control* 16 (3-4): 601-20.
- Heaton, John, and Deborah J. Lucas. 1996. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy* 104 (3): 443-87.
- Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser. 1993. Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *The Journal of Finance* 48 (1): 93-130.
- Hoerl, Arthur E., and Robert W. Kennard, 1970, Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics* 12, 55-67.
- Hoerl, Arthur E., and Robert W. Kennard, 1970, Ridge regression: applications to nonorthogonal problems, *Technometrics* 12, 69-82.
- Hong, Harrison, Terence Lim, and Jeremy C. Stein. 2000. Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance* 55 (1): 265-95.
- Hong, Harrison, and David A. Sraer, 2016, Speculative betas, *The Journal of Finance* 71, 2095-2144.
- Hou, Kewei, Chen Xue, and Lu Zhang. 2015. Digesting anomalies: An investment approach. *The Review of Financial Studies* 28 (3): 650-705.

- Huang, J., J. L. Horowitz, and F. Wei, 2010, Variable Selection in Nonparametric Additive Models, *Annals of statistics* 38, 2282-2313.
- Huberman, Gur. 1982. A simple approach to arbitrage pricing theory. *Journal of Economic Theory* 28 (1): 183-91.
- Ian, Martin,. 2017. What is the expected return on the market? *Quarterly Journal of Economics* 132 (1), 367-433: 0033-5533
- Ian, Martin,. and Christian Wagner. 2018. What is the Expected Return on a Stock? *Journal of finance*. 74 (4): 1887-1929
- Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi. 2002. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics* 20 (1): 69-87.
- Jean-Philippe, Bouchaud., Philipp, Kruger., Augustin, Landier., and David Thesmar. 2019. *Journal of Finance* 74 (2): 639-674.
- Jegadeesh, Narasimhan. 1990. Evidence of predictable behavior of security returns. *The Journal of Finance* 45 (3): 881-98.
- Jegadeesh, Narasimhan, and Sheridan Titman. 2002. Cross-sectional and time-series determinants of momentum returns. *The Review of Financial Studies* 15 (1): 143-57.

- Jegadeesh, Narasimhan, and Sheridan Titman. 2001. Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance* 56 (2): 699-720.
- Jegadeesh, Narasimhan, and Sheridan Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* 48 (1): 65-91.
- Jensen, Michael C. 1968. The performance of mutual funds in the period 1945–1964. *The Journal of Finance* 23 (2): 389-416.
- Jensen, Michael C., Fischer Black, and Myron S. Scholes. 1972. The capital asset pricing model: Some empirical tests.
- John H. Cochrane. 2011. Presidential Address: Discount Rates. *The Journal of Finance* 57 (4): 1047-1108
- Johnson, Timothy C. 2002. Rational momentum effects. *The Journal of Finance* 57 (2): 585-608.
- Jones, Christopher S., and Jay Shanken. 2005. Mutual fund performance with learning across funds. *Journal of Financial Economics* 78 (3): 507-52.
- Jostova, Gergana, Stanislava Nikolova, Alexander Philipov, and Christof W. Stahel. 2013. Momentum in corporate bond returns. *The Review of Financial Studies* 26 (7): 1649-93.
- Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*: 263-91.
- Kandel, Shmuel, and Robert F. Stambaugh. 1996. On the predictability of stock returns: An asset-allocation perspective. *The Journal of Finance* 51 (2): 385-424.

- Keim, Donald B., and Robert F. Stambaugh. 1986. Predicting returns in the stock and bond markets. *Journal of Financial Economics* 17 (2): 357-90.
- Kelly, Bryan, and Hao Jiang. 2014. Tail risk and asset prices. *The Review of Financial Studies* 27 (10): 2841-71.
- Kendall, Maurice G. 1954. Note on bias in the estimation of autocorrelation. *Biometrika* 41 (3-4): 403-4.
- Kirby, Chris. 1998. The restrictions on predictability implied by rational asset pricing models. *The Review of Financial Studies* 11 (2): 343-82.
- Kirby, Chris. 1997. Measuring the predictable variation in stock and bond returns. *The Review of Financial Studies* 10 (3): 579-630.
- Klein, Roger W., and Vijay S. Bawa. 1976. The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics* 3 (3): 215-31.
- Korajczyk, Robert A., and Ronnie Sadka. 2004. Are momentum profits robust to trading costs? *The Journal of Finance* 59 (3): 1039-82.
- Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, Interpreting factor models, *The Journal of Finance* 73, 1183-1223.
- Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, Shrinking the cross section, NBER Working Paper.
- Kumar, Alok. 2009. Who gambles in the stock market? *The Journal of Finance* 64 (4): 1889-933.
- Lee, Charles, and Bhaskaran Swaminathan. 2000. Price momentum and trading volume. *The Journal of Finance* 55 (5): 2017-69.
- Lehmann, Bruce N. 1990. Fads, martingales, and market efficiency. *The Quarterly Journal of Economics* 105 (1): 1-28.
- Lehmann, Bruce N., and David M. Modest, 1988, The empirical foundations of the arbitrage pricing theory, *Journal of Financial Economics* 21, 213-254.

- Lettau, Martin, and Sydney Ludvigson. 2001. Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance* 56 (3): 815-49.
- Lettau, Martin, and Sydney Ludvigson. 2001. Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109 (6): 1238-87.
- Lewellen, Jonathan, and Stefan Nagel. 2006. The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics* 82 (2): 289-314.
- Laibson, David, 1997, Golden eggs and hyperbolic discounting, *The Quarterly Journal of Economics* 112, 443-478.
- Liew, Jimmy, and Maria Vassalou. 2000. Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics* 57 (2): 221-45.
- Lintner, John. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*: 13-37.
- Lo, Andrew W., and A. Craig MacKinlay. 1990. When are contrarian profits due to stock market overreaction? *The Review of Financial Studies* 3 (2): 175-205.
- Lucas Jr, Robert E. 1978. Asset prices in an exchange economy. *Econometrica: Journal of the Econometric Society*: 1429-45.
- Lustig, Hanno N., and Stijn G. Van Nieuwerburgh. 2005. Housing collateral, consumption insurance, and risk premia: An empirical perspective. *The Journal of Finance* 60 (3): 1167-219.
- Luyang, Chen., Markus, Pelger†., and Jason Zhu. 2019. Deep Learning in Asset Pricing. Working paper.

- MacKinlay, A., and Ľuboš Pástor. 2000. Asset pricing models: Implications for expected returns and portfolio selection. *The Review of Financial Studies* 13 (4): 883-916.
- Mankiw, N. Gregory, and Stephen P. Zeldes. 1991. The consumption of stockholders and nonstockholders. *Journal of Financial Economics* 29 (1): 97-112.
- Mankiw, N. Gregory, and Ricardo Reis. 2002. Sticky information versus sticky prices. *Quarterly Journal of Economics* 117(4): 1295-1328
- Markowitz, Harry. 1952. Portfolio selection. *The Journal of Finance* 7 (1): 77-91.
- McLean, R. David, and Jeffrey Pontiff. 2016. Does academic research destroy stock return predictability? *The Journal of Finance* 71 (1): 5-32.
- Mehra, Rajnish, and Edward C. Prescott. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15 (2): 145-61.
- Menzly, Lior, Tano Santos, and Pietro Veronesi. 2004. Understanding predictability. *Journal of Political Economy* 112 (1): 1-47.
- Merton, Robert C. 1987. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance* 42 (3): 483-510.
- Merton, Robert C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance* 29 (2): 449-70.

- Merton, Robert C. 1973. Theory of rational option pricing. *The Bell Journal of Economics and Management Science*: 141-83.
- Merton, Robert C. 1969. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*: 247-57.
- Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21 (6): 1087-92.
- Miller, Edward M. 1977. Risk, uncertainty, and divergence of opinion. *The Journal of Finance* 32 (4): 1151-68.
- Mohanram, Partha S. 2005. Separating winners from losers among lowbook-to-market stocks using financial statement analysis. *Review of Accounting Studies* 10 (2): 133-70.
- Moskowitz, Tobias J., and Mark Grinblatt. 1999. Do industries explain momentum? *The Journal of Finance* 54 (4): 1249-90.
- Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen. 2012. Time series momentum. *Journal of Financial Economics* 104 (2): 228-50.

- Mossin, Jan. 1966. Equilibrium in a capital asset market. *Econometrica: Journal of the Econometric Society*: 768-83.
- Newey, Whitney K., and Kenneth D. West. 1987. Hypothesis testing with efficient method of moments estimation. *International Economic Review*: 777-87.
- Novy-Marx, Robert. 2013. The other side of value: The gross profitability premium. *Journal of Financial Economics* 108 (1): 1-28.
- Novy-Marx, Robert. 2010. Operating leverage. *Review of Finance* 15 (1): 103-34.
- Ohlson, James A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*: 109-31.
- Otrok, Christopher, B. Ravikumar, and Charles H. Whiteman. 2002. Habit formation: A resolution of the equity premium puzzle? *Journal of Monetary Economics* 49 (6): 1261-88.
- Parker, Jonathan A., and Christian Julliard. 2005. Consumption risk and the cross section of expected returns. *Journal of Political Economy* 113 (1): 185-222.
- Pástor, Ľuboš. 2000. Portfolio selection and asset pricing models. *The Journal of Finance* 55 (1): 179-223.

- Pástor, Ľuboš, Meenakshi Sinha, and Bhaskaran Swaminathan. 2008. Estimating the intertemporal risk–return tradeoff using the implied cost of capital. *The Journal of Finance* 63 (6): 2859-97.
- Pástor, Ľuboš, and Robert F. Stambaugh. 2012. Are stocks really less volatile in the long run? *The Journal of Finance* 67 (2): 431-78.
- Pástor, Ľuboš, and Robert F. Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111 (3): 642-85.
- Pástor, Ľuboš, and Robert F. Stambaugh. 2002. Investing in equity mutual funds. *Journal of Financial Economics* 63 (3): 351-80.
- Pástor, Ľuboš, and Robert F. Stambaugh. 2002. Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics* 63 (3): 315-49.
- Pástor, Ľuboš, and Robert F. Stambaugh. 2000. Comparing asset pricing models: An investment perspective. *Journal of Financial Economics* 56 (3): 335-81.
- Pástor, Ľuboš, and Robert F. Stambaugh. 1999. Costs of equity capital and model mispricing. *The Journal of Finance* 54 (1): 67-121.
- Pastor, Lubos ; Robert F. Stambaugh, and Lucian A. Taylor, 2017, Do funds make more when they trade more, *Journal of Finance* 72, 1483-1528.

- Piotroski, Joseph D. 2000. Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*: 1-41.
- Pye, Gordon. 1973. Lifetime portfolio selection in continuous time for a multiplicative class of utility functions. *The American Economic Review* 63 (5): 1013-6.
- Pye, Gordon. 1972. Preferential tax treatment of capital gains, optimal dividend policy, and capital budgeting. *The Quarterly Journal of Economics* 86 (2): 226-42.
- Rapach, David E., Jack K. Strauss, and Guofu Zhou. 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies* 23 (2): 821-62.
- Richardson, Scott A., and Richard G. Sloan. 2003. External financing and future stock returns.
- Rietz, Thomas A. 1988. The equity risk premium a solution. *Journal of Monetary Economics* 22 (1): 117-31.
- Roll, Richard, and Stephen A. Ross. 1980. An empirical investigation of the arbitrage pricing theory. *The Journal of Finance* 35 (5): 1073-103.

- Ross, Stephen A. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13 (3): 341-60.
- Rouwenhorst, K. Geert. 1998. International momentum strategies. *The Journal of Finance* 53 (1): 267-84.
- Rubinstein, Mark. 1976. The valuation of uncertain income streams and the pricing of options. *The Bell Journal of Economics*: 407-25.
- Samuelson, Paul A. 1969. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*: 239-46.
- Schwert, G. William. 2003. Anomalies and market efficiency. *Handbook of the Economics of Finance* 1 : 939-74.
- Shanken, Jay. 1992. On the estimation of beta-pricing models. *The Review of Financial Studies* 5 (1): 1-33.
- Shanken, Jay. 1990. Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics* 45 (1-2): 99-120.
- Sharpe, William F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19 (3): 425-42.
- Sloan, Allan. 1996. Create an account or log in. *Accounting Review* 71 (3): 289-315.

- Stambaugh, Robert F. 1999. Predictive regressions. *Journal of Financial Economics* 54 (3): 375-421.
- Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan. 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance* 70 (5): 1903-48.
- Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104 (2): 288-302.
- Stambaugh, Robert F., and Yu Yuan. 2016. Mispricing factors. *The Review of Financial Studies* 30 (4): 1270-315.
- Sundaresan, Suresh M. 1989. Intertemporally dependent preferences and the volatility of consumption and wealth. *Review of Financial Studies* 2 (1): 73-89.
- Tibshirani, Robert, 1996, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B (Methodological)* , 267-288.
- Tierney, Luke. 1994. Markov chains for exploring posterior distributions. *The Annals of Statistics*: 1701-28.
- Timmermann, Allan G. 1993. How learning in financial markets generates excess volatility and predictability in stock prices. *The Quarterly Journal of Economics* 108 (4): 1135-45.
- Titman, Sheridan, KC John Wei, and Feixue Xie. 2004. Capital investments and stock returns. *Journal of Financial and Quantitative Analysis* 39 (4): 677-700.

- Tversky, Amos, and Daniel Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5 (4): 297-323.
- Vassalou, Maria, and Yuhang Xing. 2004. Default risk in equity returns. *The Journal of Finance* 59 (2): 831-68.
- Vissing-Jørgensen, Annette. 2002. Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy* 110 (4): 825-53.
- Weil, Philippe. 1989. The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics* 24 (3): 401-21.
- Weitzman, Martin. 2007. *Structural Uncertainty and the Value of Statistical Life in the Economics of Catastrophic Climate Change*.
- Welch, Ivo, and Amit Goyal. 2007. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* 21 (4): 1455-508.
- Wheatley, Simon. 1988. Some tests of international equity integration. *Journal of Financial Economics* 21 (2): 177-212.

- White, Halbert. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*: 817-38.
- Winkler, Robert L. 1973. Bayesian models for forecasting future security prices. *Journal of Financial and Quantitative Analysis* 8 (3): 387-405.
- Winkler, Robert L., and Christopher B. Barry. 1975. A bayesian model for portfolio selection and revision. *The Journal of Finance* 30 (1): 179-92.
- Yu, Jianfeng. 2012. Using long-run consumption-return correlations to test asset pricing models. *Review of Economic Dynamics* 15 (3): 317-35.
- Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics*.
- Zhang, Cun H. 2010. Nearly unbiased variable selection under minimax concave penalty, *Ann. Statist.* 38, 894–942.
- Zhang, X. 2006. Information uncertainty and stock returns. *The Journal of Finance* 61 (1): 105-37.
- Zheng, Lu. 1999. Is money smart? A study of mutual fund investors' fund selection ability. *The Journal of Finance* 54 (3): 901-33.
- Zhou, Guofu, and Yingzi Zhu. 2013. *The Long-Run Risks Model: What Differences can an Extra Volatility Factor make?*.
- Zou, Hui, and Trevor Hastie, 2005, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 301-320.