Limits to Arbitrage and Commodity Index Investment

Yiqun Mou

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

©2011

Yiqun Mou

All Rights Reserved

ABSTRACT

Limits to Arbitrage and Commodity Index Investment

Yiqun Mou

The dramatic growth of commodity index investment over the last decade has caused a heated debate regarding its impact on commodity prices among legislators, practitioners and academics. This paper focuses on the unique rolling activity of commodity index investors in the commodity futures markets and shows that the price impact due to this rolling activity is both statistically and economically significant. Two simple trading strategies, devised to exploit this market anomaly, yielded excess returns with positive skewness and annual Sharpe ratios as high as 4.4 in the period January 2000 to March 2010. The profitability of these trading strategies is decreasing in the amount of arbitrage capital employed in the futures markets and increasing in the size of index funds' investment relative to the total size of futures markets. Due to the price impact, index investors forwent on average 3.6% annual return, a 48% higher Sharpe ratio of the return, and billions of dollars over this period.

Table of Contents

I Limits to Arbitrage and Commodity Index Investment: Front-Running the Goldman Roll **1**

1 Abstract **2**

2 Introduction **3**

2.1 Related Literature 8

3 Commodity Index Investment **11**

3.1 The Goldman Roll 14

4 Empirical Analysis **17**

4.1 Preliminary Evidence of Price Impact 18

4.2 Front-Running the Goldman Roll 23

4.2.1 Performance of the Strategies 25

4.3 Limits to Arbitrage 34

4.4 Cost of the Price Impact 38

5 Conclusions **45**

II Learning about Consumption Dynamics **52**

6 Abstract **53**

7 Introduction **54**

8 The Environment **62**

8.1 Model 62

8.2 Information and learning 63

8.3 Initial beliefs 65

9 Time-series of subjective beliefs **67**

9.1 State and parameter learning 67

9.1.1 Beliefs about models and consumption dynamics 74

10 Does learning matter for asset prices? **82**

10.1 A new test for the importance of learning 82

10.2 Learning from additional macro variables 86

10.3 Additional asset pricing implications 92

10.3.1 The model 94

11 Conclusion **105**

12 Appendix **114**

12.1 Existing literature and alternative approaches for parameter, state, and model uncertainty. 114

12.2 Econometrics 116

12.3 Priors 118

12.4 Time-Averaging of Consumption Data and Model Probabilities 120

12.5 Model solution and pricing 122

List of Figures

2.1 Related Plots of Crude Oil (WTI) Example 5

4.1 Average Roll Yields of Index Commodities over the 15-day Rolling Window 20
4.2 Average Roll Yield of Index Commodities over an Alternative 15-day Window 21
4.3 Average Roll Yield of Out-of-Index Commodities over the Rolling Window 22
4.4 Average Monthly Excess Returns of the Four Sector Portfolios with Strategy 1 29
4.5 Average Number of Spread Position Taken by Speculators 35
4.6 Value of Two Indices with Different Rolling Dates 40
4.7 Estimated Size of Index Investment and Loss due to Price Impact 43

9.1 Evolution of Posterior Mean State Beliefs 69
9.2 Mean Parameter Beliefs of the Volatility and Transition Probabilities 71
9.3 Speed of Learning 72
9.4 Marginal Model Probabilities under Different Priors 73
9.5 Quarterly Expected Consumption Growth 75
9.6 Quarterly Predictive Consumption Growth Standard Deviation 77
9.7 Quarterly Predictive Consumption Growth Skewness 79
9.8 Quarterly Predictive Consumption Growth Kurtosis 81

10.1 Evolution of Mean State Beliefs with GDP 88
10.2 Uncertainty about state identification with/without GDP 89
10.3 Marginal Model Probabilities with GDP. 90
10.4 Conditional Expected Consumption Growth with GDP 91

12.1 Model Probabilities and Time-Averaging of Consumption Data 121

List of Tables

3.1 Commodity Futures, Their Weights in SP-GSCI and DJ-UBSCI and Rolling Scheme 13

4.1 Commodities Futures out of the SP-GSCI and their Rolling Scheme 18
4.2 Summary Statistics of Monthly Excess Returns with Two Trading Strategies 26
4.3 Summary Statistics of Annualized Excess Returns in 2000-2010 28
4.4 Summary Statistics of Monthly Excess Returns with Two Strategies using Out-of-Index Commodities 31
4.5 Regressions on the Trading Strategies' Excess Returns 1 33
4.6 Regressions on the Trading Strategys' Excess Returns 2 39
4.7 Summary Statistics of Two Indices with Different Rolling Periods 42

10.1 Updates in Beliefs versus Realized Stock Returns 84
10.2 Updates in Beliefs versus Realized Stock Returns with GDP 93
10.3 Asset Price Moments 97
10.4 Dividend Yield Regression 100
10.5 Real Risk-free Yield Volatilities 101
10.6 Return Forecasting Regressions 104

12.1 Priors Specification 126

Acknowledgments

It is difficult to overstate my appreciation to my advisors Michael Johannes and Lars Lochstoer. I am grateful I had the opportunity to meet and work with them. They have given me numerous supports and inspirations in my research. I really learned a lot from them.

I am thankful to my committee chair Pierre Collin-Dufresne for his help and encouragement. I am also grateful to Bjarni Torfason and seminar participants at Columbia University, University of Maryland, University of Wisconsin-Madison, Federal Reserve Board, University of Hong Kong, Nanyang Technological University, Norwegian School of Economics and Business Administration for helpful comments.

Special thanks be to my wife Qiqi Deng for her selflessly support. I could not get through all the difficult times without your support and understanding. Also I wish to thank my parents, Jianhua Mou and Aijuan Tan. Your unconditional love is the best thing I have ever got in my life.

To my parents Jianhua and Aijuan

my wife Qiqi and daughter Cindy

Part I

Limits to Arbitrage and Commodity Index Investment: Front-Running the Goldman Roll

Chapter 1

Abstract

The dramatic growth of commodity index investment over the last decade has caused a heated debate regarding its impact on commodity prices among legislators, practitioners and academics. This paper focuses on the unique rolling activity of commodity index investors in the commodity futures markets and shows that the price impact due to this rolling activity is both statistically and economically significant. Two simple trading strategies, devised to exploit this market anomaly, yielded excess returns with positive skewness and annual Sharpe ratios as high as 4.4 in the period January 2000 to March 2010. The profitability of these trading strategies is decreasing in the amount of arbitrage capital employed in the futures markets and increasing in the size of index funds' investment relative to the total size of futures markets. Due to the price impact, index investors forwent on average 3.6% annual return, a 48% higher Sharpe ratio of the return, and billions of dollars over this period.

Chapter 2

Introduction

Arbitrage is the basis of the efficient market hypothesis, as in theory, rational arbitrageurs can engage in risk-less arbitrage to quickly eliminate any market anomalies. In reality, arbitrage opportunities are often limited, because arbitrageurs are typically capital-constrained and any arbitrage carries some risks. Recent literature on limits to arbitrage (Shleifer and Vishny, 1997) shows that a market anomaly can persist for a long period due to slow-moving arbitrage capital and the resulting delayed arbitrage, as summarized by Duffie (2010). While previous empirical evidence of limits to arbitrage was often found in equity and bond markets, in this paper I find a significant and persistent market anomaly in the commodity futures markets, which are attracting more and more attention from legislators, investors and economists. The market anomaly arises due to the increasing size of commodity index investment and its mechanical rolling forward of futures contracts.

Commodity index investment experienced dramatic growth over the last decade and now constitutes a significant fraction of investment in commodity futures markets. When commodity prices reached dizzying heights in mid-2008, the value of total long positions held by index investors reached $256 billion, up from about $6 billion in 1999. At the same time, the average estimated ratio of these long positions relative to total open interest increased from 6.7% in 1999 to 44% in mid-2008 across 19 largest commodity markets that this paper studies. After the commodity prices collapsed in the fall of 2008, commodity index investment dropped, but it quickly recovered. The value of index investors' long positions increased from $112 billion at the end of 2008 to $211 billion at the end of 2009,

and the average estimated market ratio also increased from 39% to 52%. While there has been a heated debate on the impact of this surge in index investment on commodity price levels[1], little attention has been devoted to the impact on a separate, but quantitatively at least as important, component of index funds' returns called the "roll yield", which depends on the slope of commodity futures curves[2].

This paper documents that the mechanical rolling forward of futures contracts explicit in index funds' investment strategies exerts large and time-varying price pressure on the futures curve in the largest commodity markets. The estimated losses incurred by index investors as a group, due to this price-pressure and arbitrageurs' front-running of their trades, amounted to $26 billion over the period 2000 to 2009, compared to the estimated total management fees of about $5 billion. Commodity index investors also forwent on average 3.6% annual return and a 48% higher Sharpe ratio of returns over this period.

The Standard and Poor's-Goldman Sachs Commodity Index (SP-GSCI) was the first commercially available commodity index and is also most popular. The SP-GSCI rolls futures forward from the fifth business day to the ninth business day of each month, and its rolling activity is usually called the *Goldman roll* by practitioners. To help understand the Goldman roll and its impact, I use crude oil (WTI) as an example and look at a 15-business-day window ending on February 13, 2001. The SP-GSCI rolled the futures of crude oil (WTI) forward from February 7 to February 13 by shorting the March contracts and longing the April contracts. Panel A of 2.1 shows the term structure of crude oil (WTI) futures on February 7, 2001. As we can see, the slope was negative, which means contracts with shorter maturities were trading at premiums. This kind of term structure is called in *backwardation* by the literature. Because the March contract was more expensive, by shorting the March contract at $31.27 and longing the April contract at $30.98, the SP-GSCI

[1]See Singleton (2010), Master and White (2008), Buyuksahin et al. (2008), Buyuksahin and Harris (2009), Hamilton (2009), Kilian (2009), Stoll and Whaley (2010).

[2]Unlike equity index funds which invest directly in the underlying assets, commodity index funds obtain commodity price exposure by entering long positions in commodity futures contracts. In order to maintain the long exposure, the funds need to unwind the maturing contracts before they expire and initiate new long positions in contracts that have later maturity dates. The roll yield refers to the difference between log price of the maturing contract they roll from and the deferred contract they roll into.

got a positive roll yield $ln(31.27/30.98) = 0.93\%$.

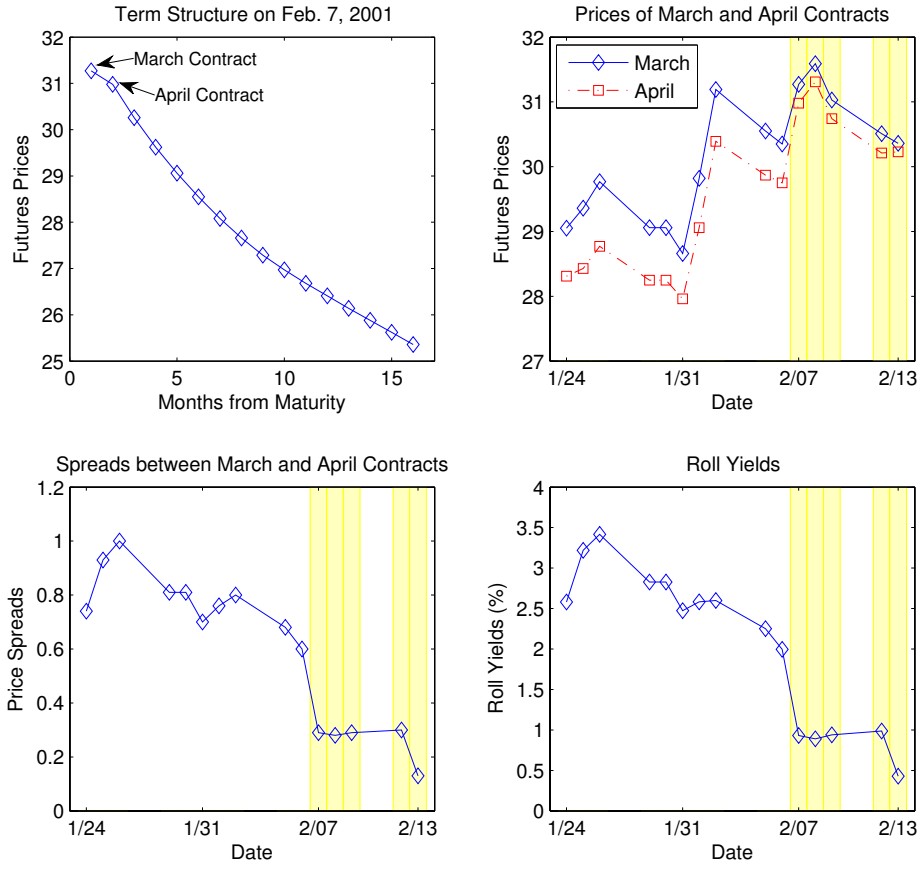

Figure 2.1: Related Plots of Crude Oil (WTI) Example

Panel B of 2.1 shows how the prices of the March and April contracts moved during the 15-day window. Although the two contracts shared the same general price pattern, their prices were much closer during the rolling period. The difference between the prices of two contracts is called the *spread.* As shown in Panel C and Panel D, the spreads and roll yields were much lower in the rolling period. More importantly, we can clearly observe a large \$0.31 drop in spread and a 1.1% drop in roll yield when entering the rolling period. This suggests that due to the large size of index investment, the shorting demand exerted by the Goldman roll caused the March contract to be temporarily underpriced, and the longing demand caused the April contract to be temporarily overpriced. The resulting price impact

also caused the roll yield to drop.

The plots also indicate how this mispricing due to the price impact could be easily exploited by long-short strategies similar to those used in the equity market. For example, on January 24, we can short the March contract at \$29.05, anticipating that it would be relatively underpriced after 10 business days. At the same time, we long the April contract at \$28.31, expecting it to be relatively overpriced when the Goldman roll happens. In this way, we create a calendar spread position with net value equal to the spread \$0.74, and our long-short spread position is not exposed to the change in absolute price level of crude oil. After the mispricing happens on February 7 due to the Goldman roll, we unwind the positions by longing the March contract and shorting the April contract to exactly offset the positions of the SP-GSCI, paying the spread \$0.29. This front-running strategy profits from the drop in the spread \$0.74−\$0.29=\$0.45, and if we post full collateral for the spread position: \$28.68 ($=\frac{\$29.05+\$28.31}{2}$), the strategy yields an unleveled excess return of 1.57% in 10 business days. In the real world, initiating such a spread position only requires 2-3% margin of the nominal value, so the strategy can be easily implemented with very high leverage. As indicated by the plots, this front-running strategy can still yield high excess returns even if we initiate our positions just a few days before the Goldman roll.

I focus on 19 commodities in the SP-GSCI that are traded on US exchanges. These commodities are very representative, because they have the largest and also the most liquid commodity futures markets, with a total weight of 93.22% in the SP-GSCI in 2010. The sample period is from January 1980 to March 2010. The year 2000 is set as a cut-off point, because index investment was nonexistent or very small (less than \$6 billion) before 2000. Two simple trading strategies, like the one above, are designed to exploit the price impact. The only difference is that Strategy 1 front-runs the Goldman roll by 10 business days, and Strategy 2 front-runs it by just 5 business days. In the example above, Strategy 1 would initiate spread positions from January 24 to January 30, and Strategy 2 would initiate positions from January 31 to February 6. Both strategies unwind positions from February 7 to February 13, when the SP-GSCI rolls futures forward.

The 19 commodities are grouped in sectors to form 4 equally weighted sector portfolios (agriculture, livestock, energy and metals) and one total portfolio. In the period 1980-

1999, the portfolios' Sharpe ratios were typically low or negative. However, in the period 2000-2010, both strategies yielded very high abnormal returns. Under the assumption that capital was invested in risk-free assets when it was not utilized for the strategies, the annualized Sharpe ratios ranged from 1.09 to 2.75 with Strategy 1, and ranged from 0.46 to 1.78 with Strategy 2. More importantly, the excess returns were positively skewed for most portfolios, with a maximum skewness of 2.23 with Strategy 1 and 2.45 with Strategy 2. Energy sector is overall the best performing sector. With Strategy 1, the energy portfolio has unleveled annual excess return of 4.43%, with annual Sharpe ratio of 2.2, skewness of 0.88 and maximum drawdown of 0.94%. From the perspective of a money manager who has multiple trading opportunity and who only cares about performance in the trading periods, the annualized Sharpe ratios ranged from 2.0 to 3.99 with Strategy 1, and ranged from 1.16 to 4.39 with Strategy 2. Besides the metals portfolio, the mean of *unlevered* annual excess returns ranged from 7.8% to 10.5% with Strategy 1, and ranged from 5.2% to 10.8% with Strategy 2. A closer examination of the strategies' performance reveals that the exact choice of cut-off year is not important. For the energy and livestock portfolios, the strategies' excess returns were mostly positive as early as 1992, right after the launch of the SP-GSCI in November 1991.

When the same strategies are applied to 18 commodities not included in the SP-GSCI, there were no abnormal returns earned in either period. The annualized Sharpe ratios of similar portfolios were either negative or very small, with a maximum of 0.31. Results from panel regressions show that the average excess returns with both strategies were not significantly different from 0 for either commodities out of the SP-GSCI over the full sample period, or commodities in the SP-GSCI before the launch of the index (or the commodities' inclusion into the SP-GSCI). After the commodities were included in the SP-GSCI, the average excess return was 0.35% with Strategy 1 in 10 days and 0.24% with Strategy 2 in 5 days. Both are statistically significant at the 1% level.

All information about the Goldman roll is publicly available. What is more, compared to the equity and bond markets, there are fewer barriers to arbitrage in commodity futures markets. There is no short-sell constraint. Anyone can enter into both long and short positions freely. High leverage can be easily obtained by the low margin requirements. The

commodities in the SP-GSCI have very liquid futures markets, and the contracts involved in the Goldman roll are also the most liquid contracts in each commodity market. If the market was well arbitraged, we would not observe this market anomaly, because arbitrageurs would quickly eliminate the price impact. However, the performance of the strategies suggests that this market anomaly has persisted for a long period and arbitrage capital can be slow-moving.

CFTC's Commitment of Traders (COT) reports publish the number of positions held by different traders in commodity futures market from 1986. I find little increase in the number of spread positions held by speculators before 2004 in the 17 commodities' futures markets that have data available, which indicates that very few arbitrageurs were exploiting the market anomaly before 2004. It could be due to the inattention of arbitrageurs to commodity markets and thus their unawareness of this market anomaly. However, the number of spread positions held by speculators has experienced a dramatic jump since 2004 in all 17 commodity markets, most more than 5-fold. It suggests that as commodity markets and commodity index investment gained more attention from the investment community, arbitrageurs were getting aware of the market anomaly, and more arbitrage capital was utilized to exploit the price impact. Consistent with the limits to arbitrage theory, the paper shows that the performances of front-running strategies are significantly related to the net forces of the size of index investment and size of arbitrage capital utilized to take advantage of the market anomaly. The arbitrage profit is lower when there is a reduction in index investment or an increase in arbitrage capital.

The remainder of the introduction relates the paper to the literature. Section 2 describes some facts about commodity index investment and the Goldman roll. Section 3 presents the empirical analysis. Section 4 concludes.

2.1 Related Literature

There is a large literature on limits to arbitrage, as summarized by Shleifer (2000), Barberis and Thaler (2001) and Duffie (2010). In theory, arbitrageurs often have to bear three kinds of risks: fundamental risk (Shleifer and Vishny, 1997), noise trader risk (Delong et al.,

1990) and synchronization risk (Abreu and Brunnermeier, 2002). These risks can prevent arbitrageurs from eliminating a market anomaly quickly and thus cause delayed arbitrage. Duffie (2010) proposed that arbitrageurs' inattention can also cause slow-moving arbitrage capital and delayed arbitrage. In this paper, I contribute an empirical example of limits to arbitrage in commodity futures markets. Here, there are two possible explanations for the persistence of the market anomaly. One is the limited knowledge of the existence of the market anomaly, which is consistent with the theory of inattentive arbitrageurs. The anomaly can also persist due to the fundamental risk involved in the arbitrage. Although the mispriced futures contracts have the same underlying commodity, they are still not perfect substitutes for each other because their maturities are different. The fundamental value of this partially hedged portfolio might change due to exogenous demand shocks or a supply crunch, which could lead to a loss for arbitrageurs. The concern of this fundamental risk may delay the action of arbitrageurs, especially when the price impact of commodity index investment was not large enough.

Many empirical studies on limits to arbitrage focus on the effects of index investment in the equity market. First is the inclusion effect. Petajisto (2010) shows that in the period 1990-2005, prices increased an average 8.8% around the event for stocks added to the S&P 500, and dropped -15.1% if the stocks are deleted from the index[3]. The effect generally grew with the size of index fund assets. Second, Morck and Yang (2001) and Cremers, Petajisto and Zitzewitz (2010) find significantly large price premiums attached to index membership. Third, Kaul, Mehrotra and Morck (2000) show that when the index increased the weights of stocks, prices experienced significant increases during the event week with no reversal afterwards, even when the adjustment was previously announced. In this paper, I extend the research into commodity markets, and find that commodity index investors get significantly lower roll yields due to the price impact of their mechanical rolling activity.

The paper is also related to a classic theory called the *Theory of Normal Backwardation* (Keynes (1930), Hicks (1939) and Cootner (1967)) in commodity markets. The theory emphasizes the interaction between hedgers and speculators. In the theory, the commod-

[3]Other studies of this effect include Harris and Gurel (1986), Shleifer (1986), Lynch and Mendenhall (1997), Chen, Noronha and Singal (2004), and many others.

ity producers are typically the hedgers and short futures contracts due to risk aversion. Speculators earn a risk premium by taking long positions to meet the hedging demand of producers. Empirical evidence[4] shows that the risk premium is higher when the producers' hedging demand is higher. Commodity indices were originally designed to capture this risk premium, so index investors are often called index speculators. However, in this paper, index investors are actually the hedgers. Because the commodity index funds and banks selling swaps have to follow the exact rolling rules of the indices they track in order to fully hedge themselves, they have great hedging demand when they roll futures contracts forward. By meeting this hedging demand, speculators could earn very high excess returns. Hirshleifer's (1988, 1990) theoretical models indicate that in equilibrium a friction to investing in commodity futures must exist for the hedging demand to affect prices. Bessembinder and Lemmon (2002) model this friction as the absence of storage in electricity markets, while Acharya, Lochstoer and Ramadorai (2010) model the friction as the limit on the risk-taking capacity of speculators. Here, the friction arises from the restriction of index investors to follow fixed rolling rules, which are publicly known.

[4]See Carter, Rausser and Schmitz (1983), Chang (1985), Bessembinder (1992), de Roon, Nijman and Veld (2000) and Acharya, Lochstoer and Ramadorai (2010).

Chapter 3

Commodity Index Investment

Commodity index investment has become increasingly popular among institutional and individual investors in recent years. The first commercially available commodity index was launched at the end of 1991, and now there are hundreds of different indices. Institutional investors, such as pension funds and endowment funds, usually enter into over-the-counter (OTC) commodity index swaps with big banks. In a typical commodity index swap, the institutional investor pays the 3-month Treasury-bill rate plus a management fee to a Wall Street bank, and the bank pays the total return on a particular commodity index. The management fee ranges from 0.5% to 1% per year depending on the index and nominal amount. Institutional investors can also put their funds under the management of a commodity index fund, which tracks a particular index. For individual investors, the main investment channel is to buy exchanged-traded funds (ETFs) and notes (ETNs) which are tied to particular indices. The management fees associated with ETFs or ETNs are typically higher than the fees of swaps. Like other index investors, commodity index investors are usually long-term investors and mostly passive in the sense that there is no attempt to time the market or identify under-priced commodities. Most of the indices only take long positions in futures contracts[1], and all the positions are fully collateralized, with the collateral invested in 3-month Treasury bills.

[1]Starting from 2006, some new commodity indices take both long and short positions depending on the term structures and other factors, like the Morningstar long and short commodity index. However, the majority of commodity indices still only take long positions.

The Standard and Poor's-Goldman Sachs Commodity Index (SP-GSCI) and the Dow Jones-UBS Commodity Index (DJ-UBSCI) are the two most popular commodity indices and used as industry benchmarks. According to Masters and White (2008), the estimated market share was approximately 63% for the SP-GSCI and 32% for the DJ-UBSCI in 2008. The SP-GSCI was the first commercially available commodity index and was launched in November 1991. It includes 24 commodities now, and the composition has remained the same since 2002. The DJ-UBSCI was launched in July 1998 and includes 19 commodities, 18 of which it shares with the SP-GSCI. The weighting schemes of the two indices are different. The weights in the SP-GSCI are primarily based on the delayed five-year rolling averages of world production quantities, while the DJ-UBSCI chooses weights based on liquidity and world production values, where liquidity is the dominant factor[2]. Since the SP-GSCI is the most popular index and includes almost all commodities in the DJ-UBSCI and other indices, I will focus on the 19 commodities in the SP-GSCI that are traded on US exchanges[3]. These commodities also have the largest futures markets, and will be referred to as *index commodities*. 3.1 lists these commodities and their weights in the two indices in 2010[4]. The aggregate weights of the 19 commodities are 93.44% in the SP-GSCI and 78.21% in the DJ-USBCI in 2010, so they are very representative. As shown in Table 1, the SP-GSCI is heavily weighted on the energy sector, with a total weight of 69.25% and a crude oil weight of 50.05%. The weights in the DJ-UBSCI are more evenly dispersed, and the total energy weight is only 33%.

[2]The DJ-UBSCI also impose lower bound of 2% for individual weight and upper bound of 33% for sector weight.

[3]I exclude six industrial metals that are traded on London Metal Exchange (LME), because the maturity structure of the futures contracts listed on LME is very different from that in US. The maturities of these futures contracts range from one day to 3 months consecutively. It is not clear which contracts these indices choose and how they roll the contracts forward.

[4]The weights are taken in 2010. The index committee may revise the weights depending on various factors each year, so the weights in previous years can be different from the current weights, but the differences are not very big.

Trading Facility	Commodity (Contracts)	SP-GSCI Weights	DJ-UBSCI Weights	Futures Since	Maturity of contracts at Month Begin											
					1	2	3	4	5	6	7	8	9	10	11	12
Agriculture (8 Commodities)																
ICE	Cocoa	0.36%	0.0%	1964.12	H	H	K	K	N	N	U	U	Z	Z	Z	H
ICE	Coffee "C"	0.78%	2.56%	1972.08	H	H	K	K	N	N	U	U	Z	Z	Z	H
CBOT	Corn	3.99%	7.09%	1964.12	H	H	K	K	N	N	U	U	Z	Z	Z	H
ICE	Cotton #2	0.96%	2.00%	1964.12	H	H	K	K	N	N	Z	Z	Z	Z	Z	H
CBOT	Soybean	2.77%	7.91%	1964.12	H	H	K	K	N	N	X	X	X	X	F	F
ICE	Sugar #11	1.92%	2.89%	1964.12	H	H	K	K	N	N	V	V	V	H	H	H
KBOT	Wheat (Kansas)	0.86%	0.0%	1970.01	H	H	K	K	N	N	U	U	Z	Z	Z	H
CBOT	Wheat	4.05%	4.70%	1964.12	H	H	K	K	N	N	U	U	Z	Z	Z	H
Livestock (3 Commodities)																
CME	Feeder Cattle	0.56%	0.0%	1972.03	H	H	J	K	Q	Q	Q	U	V	X	F	F
CME	Lean Hogs	1.54%	2.10%	1966.02	G	J	J	M	M	N	Q	V	V	Z	Z	G
CME	Live Cattle	3.01%	3.55%	1964.12	G	J	J	M	M	Q	Q	V	V	Z	Z	G
Energy (6 Commodities)																
ICE	Crude Oil (Brent)	13.14%	0.0%	1989.07	H	J	K	M	N	Q	U	V	X	Z	F	G
NYMEX	Crude Oil (WTI)	36.91%	14.34%	1983.03	G	H	J	K	M	N	Q	U	V	X	Z	F
ICE	Gasoil	4.78%	0.0%	1986.06	G	H	J	K	M	N	Q	U	V	X	Z	F
NYMEX	Gasoline (RBOB)	4.56%	3.53%	1984.12	G	H	J	K	M	N	Q	U	V	X	Z	F
NYMEX	Heating Oil #2	4.54%	3.58%	1978.11	G	H	J	K	M	N	Q	U	V	X	Z	F
NYMEX	Natural Gas	5.32%	11.55%	1990.04	G	H	J	K	M	N	Q	U	V	X	Z	F
Metals (2 Commodities)																
NYMEX	Gold	2.86%	9.12%	1974.12	G	J	J	M	M	Q	Q	Z	Z	Z	Z	G
NYMEX	Silver	0.31%	3.29%	1964.12	H	H	K	K	N	N	U	U	Z	Z	Z	H

F: January G: February H: March J: April K: May M: June N: July Q: August U: September V: October X: November Z: December

Table 3.1: Commodity Futures, Their Weights in SP-GSCI and DJ-UBSCI and Rolling Scheme

Commodity index investments give investors exposure to commodity prices. There is both academic and industry research that suggests that even when a commodity index may be a poor stand-alone investment, it is still desirable because of the hedging against inflation and the diversification benefit added to the investors' total portfolio. Gorton and Rouwenhorst (2006) find that over the period between July 1959 and March 2004, the returns of investing in commodity futures were negatively correlated with equity and bond returns, but positively correlated with inflation. Based on the examination of asset class data from 1970 to 2004, Idzorek (2006) shows that by adding commodity indices to the portfolio, the average improvement in historical return at each risk level (standard deviation range of approximately 2.4% to 19.8%) was approximately 1.33%, with a maximum of 1.88%. However, a recent study by Tang and Xiong (2010) find that with the boom of commodity index investments, commodity prices have been increasingly exposed to market-wide shocks, and shocks to other commodities, such as oil. Therefore, it is unknown whether or not the diversification benefit of commodity index investment is sustainable in the future.

3.1 The Goldman Roll

Since futures contracts have expiration dates, to maintain the long exposure to commodity prices, commodity indices need to roll the positions forward, i.e., by closing the long positions in the maturing contracts and initiating new long positions in contracts that have later maturity dates. 3.1 shows the rolling scheme of the SP-GSCI by listing the maturities of the futures contracts held by the index on the first business day of each calendar month. If the index holds different contracts at the beginnings of two consecutive months, it means that the index rolls futures forward in the first month. For example, the SP-GSCI holds the March and May wheat contracts at the beginning of February and March respectively, so the index rolls the wheat futures forward in February by closing the March contracts and initiating the May contacts. Since the liquidity of contracts drops very quickly as the maturity increases, commodity indices usually hold contracts with short maturities. Different commodities have different rolling frequencies. Agricultural commodities are typically rolled forward 4 or 5 times a year. The livestock commodities are rolled forward a bit more

frequently, 6 to 8 times a year. The SP-GSCI rolls the energy commodities every month. Gold and silver are rolled forward 5 times a year. The rolling scheme of the DJ-UBSCI is the same for most commodities except energy commodities, which the DJ-UBSCI rolls every two months.

In the rolling month, both the SP-GSCI and DJ-UBSCI have a rolling period of 5 business days. The SP-GSCI starts on the fifth business day of the month, and ends on the ninth business day, while the DJ-UBSCI rolls from the sixth business day to the tenth business day, so the rolling periods of the two indices greatly overlap. Many other indices and ETFs also roll in this period, like the former Lehman Brothers Commodity Index and the largest crude oil ETF: United States Oil Fund (USO). On each day in the rolling period, both indices roll forward 20% of the positions for commodities that need to be rolled. Since the DJ-UBSCI's rolling rules are mostly the same as the SP-GSCI and the SP-GSCI is much more popular, in the following empirical analysis, I will focus on the rolling activity of the SP-GSCI, which is called the Goldman roll by practitioners.

The total excess return of investing in futures consists of spot return and roll yield. Spot return captures the price change of the futures contracts that investor holds. Roll yield (also called roll return) captures the slope of futures curve when investors roll futures forward. From now on, the contracts held by the SP-GSCI will be referred to as the maturing contracts, and the contracts that the SP-GSCI rolls into will be referred to as the deferred contracts. Suppose the price of the maturing contract is F_{t,T_1} at time t with maturity T_1, and F_{t,T_2} is the price of the deferred contract with maturity T_2, where $T_2 > T_1$. The roll yield is defined as

$$Roll\ Yield = ln(F_{t,T_1}) - ln(F_{t,T_2}) \tag{1}$$

When the maturing contract is more expensive $F_{t,T_1} > F_{t,T_2}$, the term structure is usually called in *backwardation* and the roll yield is positive. When the maturing contract is at a discount $F_{t,T_1} < F_{t,T_2}$, the term structure is called in *contango* and the roll yield is negative.

Historically, the roll yield is an important component of the total excess return. Anson (1998) shows that the roll yield provided most of commodity investments' total excess

return in the period between 1985 and 1997, and in the case of the SP-GSCI, the average annual roll yield was 6.11% while the average spot return was -0.08%. Nash (2001) and Feldman and Till (2006) find that from 1983 to 2004, whether a commodity was in structural backwardation or not largely determined its returns, and roll yield has been the dominant driver of commodity futures returns.

Chapter 4

Empirical Analysis

The daily prices for individual commodity's futures contracts are obtained from the Commodity Research Bureau (CRB) and the full sample period is from January 2, 1980 to March 31, 2010[1]. In the following analysis, the year 2000 is often set as a cutoff point, since commodity index investment was nonexistent or very small (less than $6 billion) before 2000[2]. To facilitate the analysis, I form a control group using 18 commodities not included in the SP-GSCI with futures trading on US exchanges since 2005 or earlier. These commodities will be referred to as *out-of-index commodities*. I apply a similar rolling scheme as the SP-GSCI by matching the sector and maturity structures of futures markets. The rolling periods of these commodities are exactly the same as the SP-GSCI. 4.1 lists the commodities in this control group[3] and the rolling scheme. Many commodities in the control

[1]I exclude the sample before 1980 due to the following considerations. First, there could be some potential structural changes in commodity futures markets, so the data further back may not be so relevant. Second, the SP-GSCI is heavily weighted on energy sector, and the first energy commodity futures (heating oil) started trading at the end of 1979. Third, I check the empirical analysis using all available data and the results are very similar. The results using whole history are available upon request.

[2]The exact choice of the cutoff point is not important, and would not change the results.

[3]The soybean oil is actually included in the DJ-UBSCI and some smaller indices, but the weight is very low. The orange juice is also included in some smaller indices. The copper here is traded on NYMEX, so it is not the same contract which the SP-GSCI and DJ-USBCI hold. I put milk and butter in the livestock sector because they are produced by livestock and I can have more than one commodity in livestock sector when I form sector portfolios later.

Trading Facility	Commodity (Contracts)	Futures Data Period	Maturity of contracts at Month Begin											
			1	2	3	4	5	6	7	8	9	10	11	12
Agriculture (8 Commodities)														
CME	Lumber	1969.10–2010.3	H	H	K	K	N	N	U	U	X	X	F	F
CBOT	Oats	1959.7–2010.3	H	H	K	K	N	N	U	U	Z	Z	Z	H
ICE	Orange Juice	1967.2–2010.3	H	H	K	K	N	N	U	U	X	X	F	F
CBOT	Rough Rice	1986.8–2010.3	H	H	K	K	N	N	U	U	X	X	F	F
CBOT	Soybean Meal	1959.7–2010.3	H	H	K	K	N	N	V	V	V	F	F	F
CBOT	Soybean Oil	1959.7–2010.3	H	H	K	K	N	N	V	V	V	F	F	F
ICE	Sugar #14	1985.7–2008.2	H	H	K	K	N	N	U	U	X	X	F	F
MGEX	Wheat, Spring	1970.1–2010.3	H	H	K	K	N	N	U	U	Z	Z	Z	H
Livestock (3 Commodities)														
CME	Butter	1996.9–2010.3	G	H	K	K	N	N	U	U	V	Z	Z	G
CME	Milk, Class III	1996.1–2010.3	G	J	J	M	M	N	U	U	Z	Z	Z	G
CME	Pork Bellies	1966.2–2010.3	G	H	K	K	N	N	Q	G	G	G	G	G
Energy (4 Commodities)														
NYMEX	Coal	2001.7–2010.3	G	H	J	K	M	N	Q	U	V	X	Z	F
NYMEX	Electricity, PJM	2003.4–2010.3	G	H	J	K	M	N	Q	U	V	X	Z	F
CBOT	Ethanol	2005.3–2010.3	G	H	J	K	M	N	Q	U	V	X	Z	F
NYMEX	Propane	1987.8–2009.9	G	H	J	K	M	N	Q	U	V	X	Z	F
Metals (3 Commodities)														
NYMEX	Copper	1959.7–2010.3	G	H	J	K	M	N	Q	U	V	X	Z	F
NYMEX	Palladium	1977.1–2010.3	H	H	M	M	M	U	U	U	Z	Z	Z	H
NYMEX	Platinum	1968.3–2010.3	J	J	J	N	N	N	V	V	V	F	F	F

Table 4.1: Commodities Futures out of the SP-GSCI and their Rolling Scheme

group are closely related to some index commodities.

4.1 Preliminary Evidence of Price Impact

Given the massive size of investment tied to the SP-GSCI, when it rolls futures forward, the large shorting demand of the maturing contract (being rolled from) could potentially push its price down, while the large longing demand of the deferred contract (being rolled into) could push its price up[4]. Together, the resulting price impact would cause the roll yield to drop in the rolling period. In the following analysis, I will provide some preliminary and

[4]Some market participants state that they tend to avoid trading in the SP-GSCI rolling periods if they want to do similar trading as the SP-GSCI does.

visual evidence based on this intuition to show the existence of the price impact.

First, a 15-business-day window is constructed to examine the change of roll yields, with the last five days being the rolling dates of the SP-GSCI. This window is labeled "rolling window". Days after SP-GSCI's rolling period are not included here, because for energy commodities, after the SP-GSCI unwinds the maturing contracts, these contracts typically have less than a week before the last trading days. Previous empirical studies usually exclude such contracts with just a few days to expire, because these contracts have great liquidity concerns. The full sample is divided into two sub-samples: 1980-1999 and 2000-2010. 4.1 shows the average roll yields (in percentage) of four representative index commodities (crude oil WTI, heating oil, gasoline RBOB and live cattle) over the rolling window in the two periods.

The plots in 4.1 reveal some interesting facts. First, before 2000, the average roll yields were positive on every day for all 4 commodities. It is consistent with the findings of Litzenberger and Rabinowitz (1995) and Casassus and Collin-Dufresne (2005) that these commodities were often in backwardation. In the period 2000-2010, the average roll yields dropped, especially in the SP-GSCI's rolling period. Second and more interestingly, before 2000, the roll yields showed no clear trend in the window, and the average roll yields in the SP-GSCI's rolling period were not significantly lower than the average roll yields in the first 5 days of the window. The roll yields were also very smooth across the days. However, in the period 2000-2010, we can observe very clear drops of roll yields when entering SP-GSCI's rolling period, especially for 3 energy commodities. There are decreasing trends for all commodities, and the average roll yields in the SP-GSCI's rolling period are much lower than the average roll yields in the first 5 days, with statistical significance at the 1% level for three energy commodities and at the 5% level for live cattle.

There are also some drops of roll yields from day 6 to day 10, which could be due to the price impact of some other commodity indices that roll futures forward a little earlier than the SP-GSCI. For example, the Reuters/Jefferies-CRB Index (CRB) rolls futures forward between the 1st and 4th business days of the rolling month (day 7 to day 10), and the Deutsche Bank Liquid Commodity Index (DBLCI) has a rolling period which is between the 2nd and 6th business day (day 8 to day 11).

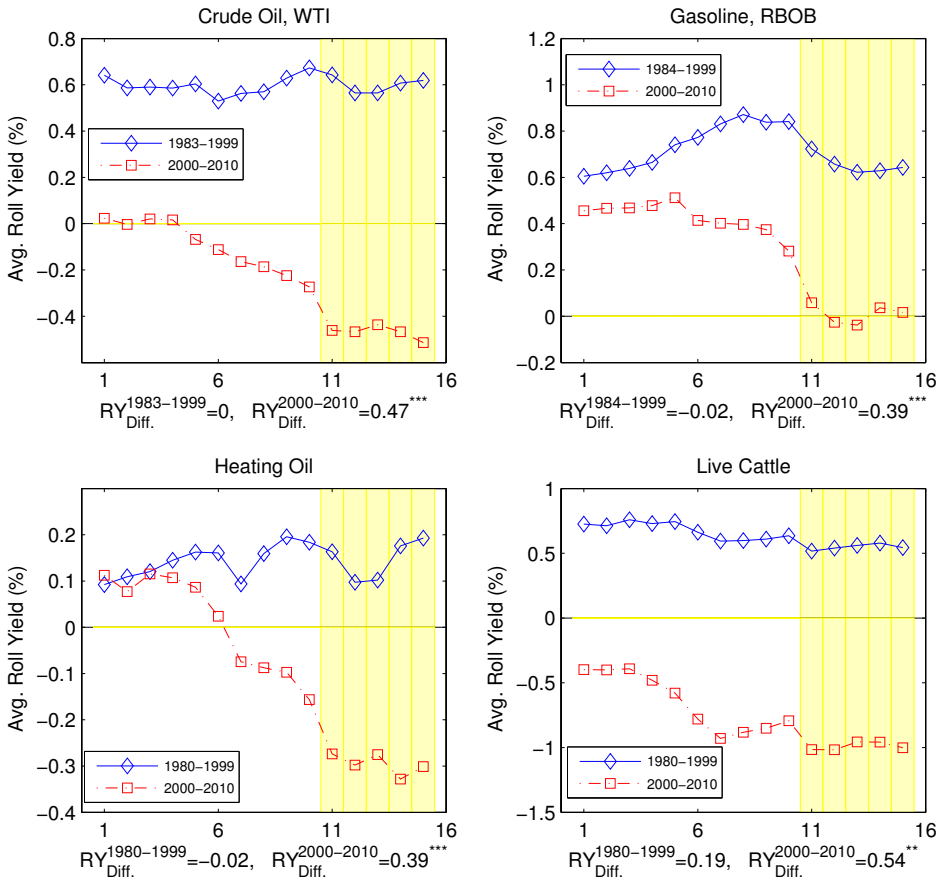

Figure 4.1: Average Roll Yields of Index Commodities over the 15-day Rolling Window

Second, I examine an alternative 15-business-day window, with the last day being one day earlier than the first day of the rolling window, so the two windows are consecutive. As shown in 4.2, there were no clear trends over the window and drops on any particular day for all commodities in both time periods. The average roll yields in the last 5 days of the window were not significantly lower than the average roll yields in the first 5 days. In the case of gasoline and heating oil, the average roll yields in the two periods were very close to each on each day.

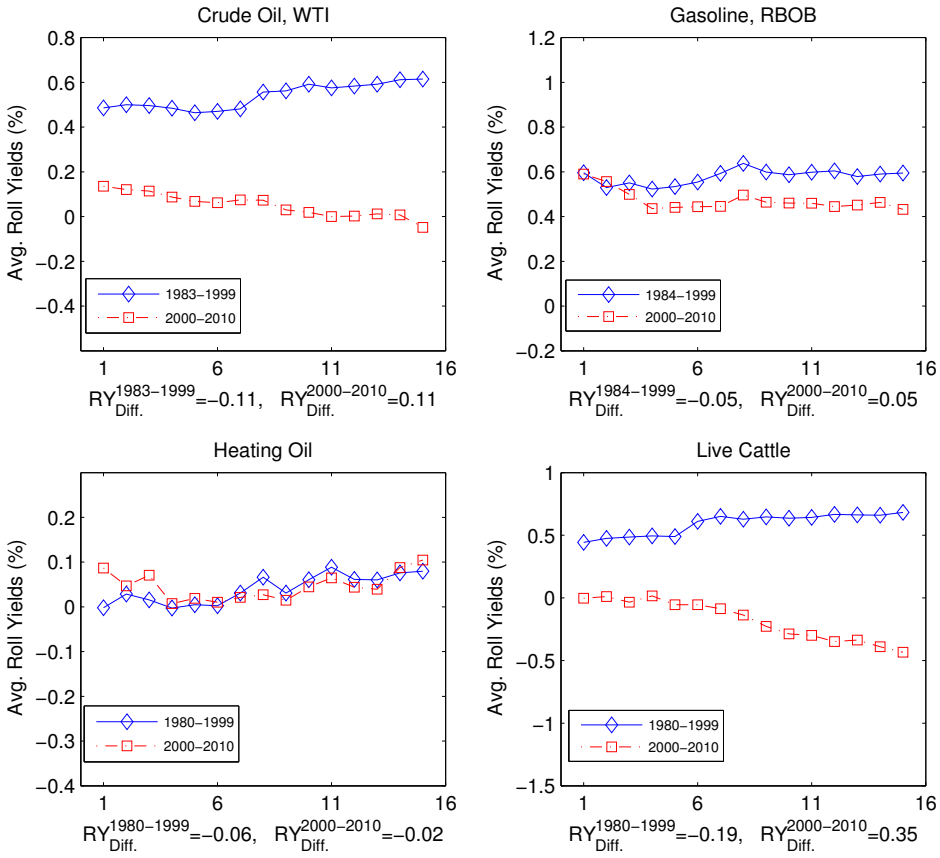

Figure 4.2: Average Roll Yield of Index Commodities over an Alternative 15-day Window

Finally, to further confirm that the unique pattern is caused by the price impact of the Goldman roll, I pick four representative out-of-index commodities from the control group and examine the change of roll yields in the rolling window. These four commodities are soybean meal, pork belly, propane and copper, one from each sector. As shown in 4.3,

the results form clear contrasts to the results of index commodities in the rolling window, but are very similar to the results of index commodities in the alternative window. For all 4 commodities in both periods, there were no clear trends and no significant differences between the average roll yields in the first and last 5 days. Also there were no clear drops of roll yields when entering the rolling period for all 4 commodities in the period 2000-2010.

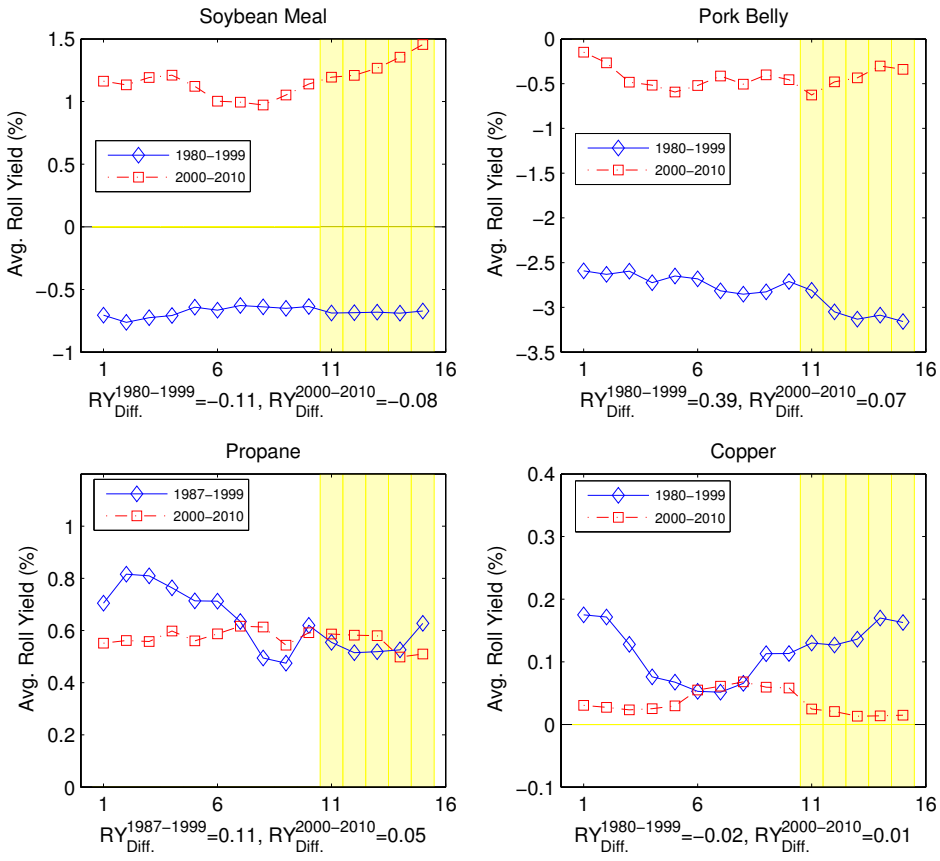

Figure 4.3: Average Roll Yield of Out-of-Index Commodities over the Rolling Window

In sum, the time-series and cross-sectional evidence above is very supportive of the existence of the price impact due to the Goldman roll. To provide further and more rigorous evidence, I will design two simple trading strategies to capture the price impact in the next section and show how both statistically and economically significant the price impact was.

4.2 Front-Running the Goldman Roll

The idea is that since the Goldman roll would cause the maturing contracts to be temporarily underpriced and the deferred contracts to be overpriced, we can create long-short positions to capture this price impact. One can either front-run by creating the positions before the Goldman roll or back-run by creating positions at the same time as the Goldman roll. Because there is liquidity concern of maturing contracts after the Goldman roll and the front-running offers more flexibility, I will only focus on front-running strategies.

Assuming that the price of the maturing contract (being rolled from) is F_{t,T_1}, and the deferred contract (being rolled into) has price F_{t,T_2}, then the spread

$$SP_t^{T_1,T_2} = F_{t,T_1} - F_{t,T_2} \tag{5}$$

is the amount of gain (or loss) per unit of the commodity when rolling futures forward. It is also the value of a calendar spread position which shorts one unit of the maturing contract, and longs one unit of the deferred contract. This long-short spread position is not subject to the change in absolute price level, and is ideal to capture the full impact of price pressures exerted by the Goldman roll in both directions.

Without price impact, the spread $SP_t^{T_1,T_2}$ should be roughly the same over a short time window. With price impact, the spread should decrease in the rolling period because the maturing contract's price F_{t,T_1} would be pushed down and the deferred contract's price F_{t,T_2} would be pushed up. The front-running strategy is designed to capture this drop of spread by shorting the maturing contracts and longing the deferred contracts before the SP-GSCI's rolling period. The spread positions are then unwound and exactly offset the SP-GSCI's positions when it roll futures forward[5].

I focus on the rolling window analyzed in the last section[6]. The 15-day window is equally divided into three groups. The formal trading strategies are designed as follows.

[5]One can also create a butterfly spread position to reduce some exposure to the slope of the futures curve. The butterfly spread position will capture the change in the convexity of the curve, and consists of long positions in the deferred contracts and short positions in the maturing contracts and contracts with maturities later than that of the deferred contracts.

[6]From 4.2, we can see that moving further ahead of the rolling window would not help the performance a lot.

With **Strategy 1**in each month, I first identify the commodities that the SP-GSCI will roll forward. For such commodities, calendar spread positions are created on each day in the first group, which runs from 10 to 6 business days before the SP-GSCI's first rolling date. The calendar spread position involves shorting the maturing contracts that the SP-GSCI is currently holding and longing the deferred contracts that it will roll into. In this way, I create the same spread positions as the Goldman roll, except I do it 10 days earlier. The calendar spread positions will be unwound in the SP-GSCI's rolling period. Like the SP-GSCI, I create 20% of the total spread positions each day and also unwind 20% each day.

Strategy 2 follows the same methodology except front-running the Goldman roll by just 5 days. The spread positions are created in the second group of days, which runs from 5 to 1 business day before the first rolling date of the SP-GSCI. Basically, Strategy 1 captures the spread change in 10 days and Strategy 2 captures the spread change in 5 days. Since both strategies are implemented in very short periods, if they earn very high abnormal excess returns, it is very unlikely to be caused by factors other than the price impact of the Goldman roll. There are multiple ways to improve the simple strategies, but the idea here is to show how the most simple and straightforward strategy would perform.

For commodity i, the excess return of Strategy j ($j = 1, 2$), from day t_j when the spread position is created to day $t^{'}$ when the position is unwound, is defined as follows

$$r_t^{i,j} = \frac{SP_{t_j}^{i,T_1,T_2} - SP_{t^{'}}^{i,T_1,T_2}}{(F_{t_j,T_1}^i + F_{t_j,T_2}^i)/2} = \frac{(F_{t_j,T_1}^i - F_{t_j,T_2}^i) - (F_{t^{'},T_1}^i - F_{t^{'},T_2}^i)}{(F_{t_j,T_1}^i + F_{t_j,T_2}^i)/2}. \tag{6}$$

This return is an excess return because the collateral earns the interest of risk-free rates. I also assume that both strategies invest the capital in the risk-free asset when they are not front-running the Goldman roll, so if the SP-GSCI rolls commodity i forward in the month, the monthly excess return of investing in commodity i with Strategy j is just the 5-day average of $r_t^{i,j}$, otherwise the monthly excess return is zero.

The 19 commodities are grouped by sector to form equally weighted portfolios (agriculture, energy, livestock and metals), and a total portfolio using all commodities. In each month, the portfolio's return is the average return of the commodities that the SP-GSCI rolls forward in this portfolio during the month. Equation (6) indicates that the calendar

spread position is fully collateralized, so the excess return $r_t^{i,j}$ involves no leverage. In practice, the margin requirement is about 10-15% of the nominal value for creating an outright futures position, and only 2-4% for initiating a calendar spread position, so both strategies can be easily implemented using very high leverage in the real world.

4.2.1 Performance of the Strategies

Similar to the previous analysis, I divided the full sample period into two sub-periods: 1980-1999 and 2000-2010. 4.2 reports the summary statistics of the five portfolios' monthly excess returns (in percentage). The difference of performances in the two periods is striking. Let us first discuss Strategy 1.

First, the mean excess returns of all 5 portfolios were very significantly positive in the period 2000-2010, and much larger than the mean excess returns before 2000. In the period 1980-1999, besides the metals portfolio, the mean excess returns ranged from -0.006% (energy) to 0.13% (agriculture) monthly, while in the period 2000-2010, the mean excess returns increased to a range of 0.31% (total) to 0.42% (livestock) monthly. The mean excess return was relatively small in magnitude for the metals portfolio, but still it increased from -0.028% before 2000 to 0.033% since 2000 (monthly).

Second, the monthly Sharpe ratios surged to very high levels in the period 2000-2010, ranging from 0.32 (agriculture) to 0.79 (total). In the period 1980-1999, besides the agriculture portfolio, the monthly Sharpe ratios of the other 4 portfolios were typically not high or even negative, ranging from -0.14 (metals) to 0.15 (total). The jumps in monthly Sharpe ratios were especially striking for three portfolios: energy portfolio (from -0.007 to 0.64), metals portfolio (from -0.14 to 0.55) and total portfolio (from 0.15 to 0.79).

Third, except for the agriculture portfolio, 4 portfolios' excess returns were positively skewed in the period 2000-2010, with skewness ranging from 0.13 (total) to 2.23 (metals). This makes it more difficult to explain the high Sharpe ratios with risk based theories. In contrast, in the period 1980-1999, the skewness was slightly positive 0.19 for the livestock portfolio, and negative for the other 3 portfolios, ranging from -3.12 (metals) to -0.17 (total).

	Agriculture		Energy		Livestock		Metals		Total	
	1980-1999	2000-2010	1980-1999	2000-2010	1980-1999	2000-2010	1980-1999	2000-2010	1980-1999	2000-2010
Strategy 1										
Mean	0.13	0.23	-0.006	0.37	0.12	0.42	-0.03	0.033	0.06	0.31
T-stat	3.29	3.50	-0.11	7.05	1.96	5.34	-2.17	6.14	2.36	8.81
Std	0.61	0.73	0.81	0.58	0.93	0.86	0.20	0.06	0.41	0.39
Skewness	1.78	-2.61	-0.71	0.88	0.19	0.74	-3.12	2.23	-0.17	0.13
Kurtosis	9.93	35.1	6.69	4.45	3.83	5.55	24.6	12.8	3.32	5.75
Min	-1.84	-5.39	-3.82	-0.88	-2.93	-1.72	-1.66	-0.087	-1.19	-1.22
Max	2.73	3.73	2.66	2.38	3.22	4.32	0.69	0.39	1.06	1.70
Sharpe Ratio	0.21	0.32	-0.007	0.64	0.13	0.48	-0.14	0.55	0.15	0.79
Max Drawdown	2.09	5.39	22.40	0.94	8.73	7.53	7.40	0.09	5.25	1.44
# of obs	240	123	240	123	240	123	240	123	240	123
Strategy 2										
Mean	0.05	0.07	0.027	0.21	0.01	0.13	-0.013	0.019	0.02	0.13
T-stat	2.13	1.47	0.66	4.81	0.25	2.33	-1.65	5.71	0.97	4.53
Std	0.38	0.51	0.64	0.50	0.62	0.61	0.12	0.04	0.31	0.31
Skewness	-1.15	-1.40	0.09	2.45	-0.07	0.12	-3.34	0.76	-0.16	-0.61
Kurtosis	12.9	33.4	7.13	15.8	4.38	3.97	21.2	5.40	4.34	11.0
Min	-1.91	-3.63	-2.44	-0.81	-2.14	-1.69	-0.85	-0.08	-0.94	-1.53
Max	1.54	3.06	2.85	3.37	2.07	2.25	0.3	0.15	1.27	1.35
Sharpe Ratio	0.17	0.16	0.04	0.43	0.02	0.21	-0.12	0.62	0.06	0.41
Max Drawdown	2.87	3.63	15.09	2.61	17.26	5.19	3.40	0.08	9.34	2.05
# of obs	240	123	240	123	240	123	240	123	240	123

Table 4.2: Summary Statistics of Monthly Excess Returns with Two Trading Strategies

Finally, in the period 2000-2010, 4 portfolios experienced big drops in the maximum drawdown. The most dramatic ones are energy and metals portfolio, whose maximum drawdowns dropped from 22.4% before 2000 to only 0.94% and from 7.4% to 0.09% respectively.

The results for Strategy 2 are similar, and even stronger in some cases. Besides the agriculture portfolio, the mean excess returns of the other 4 portfolios were not significantly different from zero before 2000, ranging from -0.013% (metals) to 0.027% (energy), but they became very positive and highly significant in the period 2000-2010, ranging from 0.019% (metals) to 0.22% (energy). The monthly Sharpe ratios of these 4 portfolios ranged from -0.11 (metals) to 0.06 (total) before 2000, and increased to the range of 0.21 (livestock) to 0.52 (metals) since 2000. The skewness of excess return also increased a lot for the energy, livestock and metals portfolios, among which the energy portfolio experienced a jump from 0.09 before 2000 to 2.45 since 2000 in skewness.

Panel A of 4.3 reports the summary statistics of the portfolios' annualized excess returns in the period 2000-2010. The annual Sharpe ratios ranged from 1.09 (agriculture) to 2.75 (total) with Strategy 1, and from 0.46 (agriculture) to 1.78 (metals) with Strategy 2. So far the capital is assumed to be invested in the risk-free assets when not utilized for front-running. However, a large hedge fund could use the capital to invest in other assets and trading strategies, so the fund manager may only care about the performance in the period when the capital is actually used. The excess returns with Strategy 1 were actually 10-day returns and should be annualized by multiplying by a factor of 252/10. Similarly, the excess returns with Strategy 2 were 5-day returns and should be annualized by a factor of 252/5. As reported in Panel B of 4.3, the annualized Sharpe ratios now are much higher, ranging from 2.0 (agriculture) to 3.99 (total) with Strategy 1 and from 1.16 (agriculture) to 4.39 (metals) with Strategy 2. Besides the metals portfolio, the means of *unlevered* annual excess returns ranged from 7.8% (total) to 10.47% (livestock) with Strategy 1, and ranged from 5.16% (agriculture) to 10.8% (energy) with Strategy 2. Therefore, the strategies' performance is much better from the perspective of a money manager with multiple investing opportunities.

The CRB data set does not have data on the bid-ask-spreads, so I can not incorporate transaction costs into the evaluation of the strategies. However, since the index commodities

Panel A: Annualized by Month

	Agriculture	Energy	Livestock	Metals	Total
Strategy 1					
Mean	2.74%	4.43%	4.99%	0.40%	3.71%
Std	2.51%	2.01%	2.99%	0.21%	1.35%
Skewness	−2.61	0.88	0.74	2.23	0.13
Sharpe Ratio	1.09	2.20	1.67	1.92	2.75
Strategy 2					
Mean	0.81%	2.58%	1.53%	0.23%	1.52%
Std	1.76%	1.72%	2.11%	0.13%	1.08%
Skewness	−1.40	2.45	0.12	0.76	−0.61
Sharpe Ratio	0.46	1.50	0.73	1.78	1.41

Panel B: Annualized by Trading Days

	Agriculture	Energy	Livestock	Metals	Total
Strategy 1					
Mean	8.74%	9.30%	10.47%	1.12%	7.80%
Std	4.38%	2.92%	4.34%	0.33%	1.95%
Skewness	−2.69	0.88	0.74	1.82	0.13
Sharpe Ratio	2.00	3.19	2.41	3.36	3.99
Strategy 2					
Mean	5.16%	10.82%	6.43%	1.29%	6.40%
Std	4.44%	3.52%	4.32%	0.29%	2.21%
Skewness	−1.32	2.45	0.12	0.34	−0.61
Sharpe Ratio	1.16	3.08	1.49	4.39	2.90

Table 4.3: Summary Statistics of Annualized Excess Returns in 2000-2010

have the most liquid markets among all commodities, and the contracts involved in the Goldman roll are also the most liquid contracts in each market, the transaction costs are quite low. The typical bid-ask-spread is only a few bps (basis points) of the futures price. For crude oil (WTI), the bid-ask-spread is often just 1 bp. In addition, since the trading volumes tend to increase a lot in the SP-GSCI's rolling period, the bid-ask-spread can be even lower when the strategies unwind the spread positions. Therefore, the strategies should still be very profitable even after taking into account the transaction costs, especially in the most liquid energy sector.

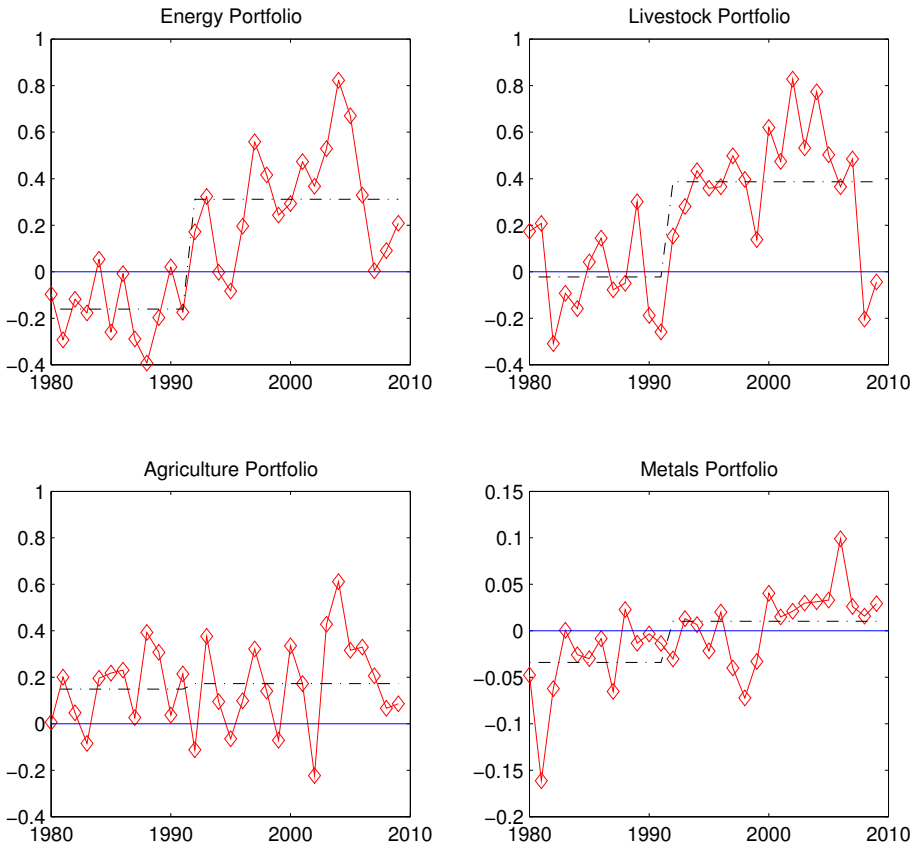

Figure 4.4: Average Monthly Excess Returns of the Four Sector Portfolios with Strategy 1

Now let us focus on Strategy 1 and take a closer look at the excess returns year by year. 4.4 shows each year the average monthly excess returns (in percentage) of the 4

sector portfolios. The energy and livestock portfolios actually had mostly positive excess returns as early as 1992, right after the launch of the SP-GSCI. For the metals portfolio, the average excess returns were mostly negative before 2000, and then stayed positive every year from 2000. The agriculture portfolio is quite different from the other 3 portfolios. The average excess returns have been mostly positive in the whole sample period, and there was a cyclical pattern before 2003. However, since 2003, the cyclical pattern disappeared and the average excess returns have stayed positive every year. The plots indicate that the exact choice of the cutoff year is not very important, and the results could be even better if the cutoff year is moved a few years earlier.

As a comparison, the same trading strategies are applied to the control group with the 18 out-of-index commodities. Similarly, four equally weighted sector portfolios and one total portfolio are formed. 4.4 reports the summary statistics of these 5 portfolios' monthly excess returns in the same two periods: 1980-1999 and 2000-2010. The results form a very clear contrast to the results in Table 3. With both strategies in both periods, most of the 5 portfolios' mean excess returns were not significantly different from 0, or even significantly negative in some cases. The monthly Sharpe ratios were all either negative or close to zero, with a maximum of 0.09 obtained by the energy portfolio with Strategy 1 before 2000. What is more, with Strategy 1, except for the livestock portfolio, the mean excess returns and Sharpe ratios actually dropped in the period 2000-2010 for 4 portfolios. With Strategy 2, there were also 4 portfolios whose mean excess returns and Sharpe ratios decreased in the period 2000-2010.

	Agriculture		Energy		Livestock		Metals		Total	
	1980-1999	2000-2010	1980-1999	2000-2010	1980-1999	2000-2010	1980-1999	2000-2010	1980-1999	2000-2010
Strategy 1										
Mean	0.02	-0.04	0.12	-0.31	-0.06	0.05	-0.06	-0.10	-0.013	-0.13
T-stat	0.64	-0.97	1.07	-1.81	-0.49	0.25	-1.55	-3.10	-0.25	-1.57
Std	0.51	0.47	1.34	1.93	1.87	2.07	0.59	0.35	0.83	0.90
Skewness	-0.73	-0.64	0.76	-0.42	1.50	-2.69	-0.84	0.06	1.27	-0.90
Kurtosis	8.17	6.16	11.0	4.42	19.8	16.7	10.9	11.2	13.3	5.63
Min	-2.48	-2.09	-5.46	-7.91	-7.50	-12.8	-2.70	-1.40	-3.57	-4.18
Max	1.70	1.15	6.39	4.66	12.1	4.17	3.34	1.73	5.13	1.80
Sharpe Ratio	0.04	-0.09	0.09	-0.16	-0.03	0.02	-0.10	-0.28	-0.02	-0.14
# of obs	240	123	145	123	240	123	240	123	240	123
Strategy 2										
Mean	0.003	-0.07	0.032	-0.10	-0.004	0.03	-0.035	-0.04	0.006	-0.06
T-stat	0.11	-2.26	0.37	-0.76	-0.05	0.25	-1.32	-2.15	0.15	-0.91
Std	0.39	0.35	1.04	1.52	1.36	1.54	0.41	0.20	0.62	0.71
Skewness	-1.24	-1.04	0.36	-0.26	2.38	-2.14	1.24	2.03	1.94	-0.83
Kurtosis	12.4	9.34	12.2	5.21	24.0	16.1	14.6	19.0	16.6	6.47
Min	-2.54	-1.85	-2.44	-5.97	-5.74	-8.64	-1.94	-0.74	-2.57	-3.10
Max	1.44	1.29	5.32	5.14	9.04	5.77	2.69	1.28	4.32	2.17
Sharpe Ratio	0.01	-0.20	0.03	-0.07	-0.003	0.02	-0.09	-0.19	0.01	-0.08
# of obs	240	123	145	123	240	123	240	123	240	123

Table 4.4: Summary Statistics of Monthly Excess Returns with Two Strategies using Out-of-Index Commodities

To further confirm the results, I perform a panel regression which is specified as follows

$$Ret^{i,t} = \alpha + \beta_1 I^i_{IndexCom} + \beta_2 I^{i,t}_{inIndex} + Controls + u_{i,t}. \quad (7)$$

where the dependent variable $Ret^{i,t}$ is Strategy's average excess return in the trading period of commodity i in year t and $u_{i,t}$ is the random error. $I^i_{IndexCom}$ is an indicator variable, which is equal to 1 if commodity i is an index commodity and 0 if it is an out-of-index commodity. $I^{i,t}_{inIndex}$ is also an indicate variable, which is equal to 1 if commodity i is actually included in the SP-GSCI in year t and 0 if otherwise. Since the SP-GSCI was launched at the end of 1991, $I^{i,t}_{inIndex} = 0$ for all index commodities before 1992. Among the 19 index commodities, natural gas was added to the SP-GSCI in 1994. Crude oil (Brent), gasoil and Kansas wheat were included into the SP-GSCI in 1999, and feeder cattle was included in 2002. All other 14 commodities were added before 1992.

To control for the macroeconomic demand-and-supply conditions and business cycle, the contemporaneous GDP growth and inflation in year t are included in the regressions. I also include a control variable that is specific to each commodity in each year. This variable is the average roll yield of commodity i in year t. This control variable summarizes the commodity-specific demand-and-supply condition and the term structure feature. All control variables are demeaned.

The coefficients of interests are α, β_1 and β_2. α is the average of $Ret^{i,t}$ for out-of-index commodities. For index commodities, $\alpha + \beta_1$ is the average of $Ret^{i,t}$ before they were included in the SP-GSCI (or the launch of the SP-GSCI), while $\alpha + \beta_1 + \beta_2$ is the average of $Ret^{i,t}$ after the inclusions. The expected values of α and β_1 are: $\alpha = 0$ and $\beta_1 = 0$, which means that without index investment, the strategy's excess return is 0. If the Goldman roll had price impact, we should expect $\beta_2 > 0$. As reported in Column 1 and 3 of 4.5, the coefficients α and β_1 are not statistically different from 0 for both strategies. After inclusion in the SP-GSCI, Strategy 1 yielded an average excess return of 0.35% in the trading period of 10 days, while Strategy 2 has an average excess return of 0.24% in the 5-day trading period. Both are statistically significant at the 1% level. Column 2 and 4 of 4.5 indicates that the results are robust if we only consider index commodities ($I^i_{IndexCom} = 1$).

For the control variables, GDP growth and inflation were both positively correlated

Dependent variable: $Ret^{i,t}$				
	Strategy 1		Strategy 2	
	All	$I^i_{IndexCom} = 1$	*All*	$I^i_{IndexCom} = 1$
	1	2	3	4
Constant	−0.028	0.011	−0.026	−0.030
	(0.040)	(0.022)	(0.024)	(0.022)
$I^i_{IndexCom}$	0.029		−0.016	
	(0.044)		(0.034)	
$I^{i,t}_{inIndex}$	0.35***	0.33***	0.24***	0.22***
	(0.057)	(0.060)	(0.036)	(0.034)
Controls				
$RY^{i,t}$	−0.031	−0.022	−0.008	0.002
	(0.019)	(0.016)	(0.011)	(0.012)
$growth^t_{GDP}$	0.034***	0.033**	0.030***	0.035***
	(0.011)	(0.013)	(0.007)	(0.009)
$Inflation^t$	0.032***	0.021**	0.019***	0.005
	(0.008)	(0.009)	(0.006)	(0.010)
R^2_{adj}	8.89%	10.25%	7.82%	10.72%
obs	956	537	956	537

Table 4.5: Regressions on the Trading Strategies' Excess Returns 1

with the dependent variable and statistically significant. The commodity-specific control variable–average roll yield of commodity i in year t–is insignificant, which means that the strategies' excess returns are not related the slope of the terms structure.

In sum, the results above indicate that the price impact of the Goldman roll is both statistically and economically significant. The Goldman roll effectively created a large market anomaly and a great trading opportunity for arbitragers.

4.3 Limits to Arbitrage

All information about the Goldman roll is publicly available. Compared to equity and bond markets, futures markets have much fewer barriers for arbitrage. There is no short-sell constraints, and high leverage can be easily obtained through low margin requirement. The transaction cost is also very low, and the trading strategies are very easy to implement. Therefore, if the market is well arbitraged, we should not expect to see such great performance of front-running the Goldman roll as any market anomaly would be quickly arbitraged away. The fact that the strategies worked so well in the last decade suggests that there are some limits to arbitrage. The performance of front-running is largely determined by two opposite forces. The positive one is the size of index investment, while the negative one is the size of arbitrage capital utilized to take advantage of the price impact.

From 1986, the CFTC started to publish weekly Commitment of Traders (COT) reports, which includes the aggregate number of spread positions taken by "Noncommercial" traders. These traders are mainly money managers and labeled *speculators* in the literature. Since to capture the price impact, the arbitrageurs have to create spread positions, the number of spread positions held by speculators serves as a good approximation, although the nature of these spread positions can not be identified.

4.5 shows each year the average spread positions taken by speculators and also their ratios relative to total open interests in the markets of 9 index commodities[7]. For most commodities, there was very few spread positions and also little growth until 2003, espe-

[7]Due to limit of space and the large number of commodities, I only report these 9 commodities. The plots for other 8 commodities have similar pattern, and are available upon request.

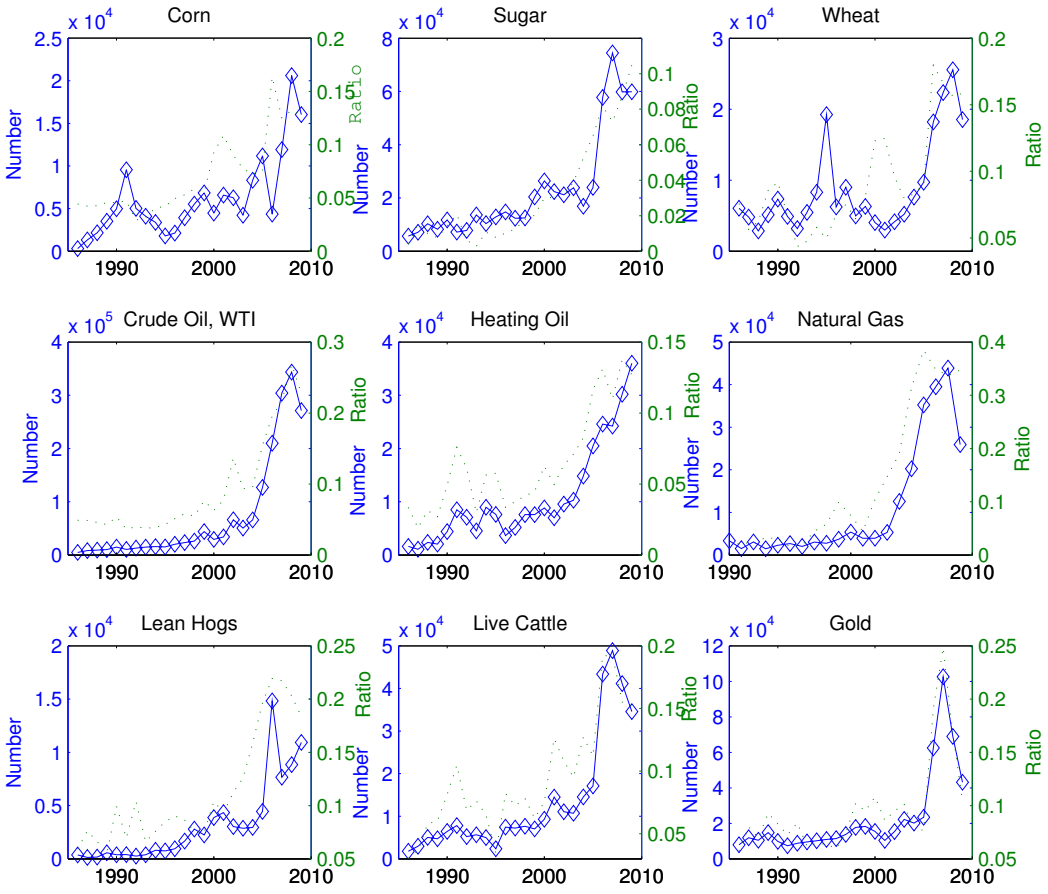

Figure 4.5: Average Number of Spread Position Taken by Speculators

cially in energy and livestock sectors, which front-running strategies yielded the best performances. However, the positions started to growth dramatically from 2003 and reached peaks in 2008 for many commodities. The growth was typically more than 5-fold. The plots suggest that very few arbitrage capital was used to exploit the price impact before 2003, and then as the arbitrageurs became more aware of this trading opportunity, more capital is utilized to exploit this market anomaly. This is consistent with the theory of Duffie (2010) that arbitrage capital can be slow-moving due to arbitrageurs' inattention to a particular market and particular strategy. Before 2003, commodity was not a popular asset class and commodity index investment was rarely known among the investment communities.

As shown in 4.4, the 4 sector portfolios enjoyed the best gains in the period 2003-2005, when commodity index investment started the most dramatic growth and there were not many arbitrageurs. During three years, the average of unlevered annual excess return was 8.09% for the energy portfolio, 7.18% for the livestock portfolio, 5.62% for the agriculture portfolio and 0.28% for the metals portfolio. However, the performance of the 4 portfolios has been declining since 2006, and the average excess returns dropped to levels close to 0. The livestock portfolio even experienced negative average excess returns since 2008.

Part of the reason is the increasing arbitrage capital, but another cause is that many investors might have moved their assets away from these commodity index investments. When the commodity prices collapsed in the middle of 2008, commodity index investment reduced a lot. The data from CFTC's supplement reports shows that the total long positions held by index investors dropped 30-50% from their peaks for many agriculture and livestock commodities in 2008. During this period, many portfolios also experienced their maximum drawdowns. Also, a new generation of commodity indices emerged since 2006 with more intelligent rolling methodologies. Many investments moved from the old generation of indices to the new generation. Instead of just focusing on contracts with short maturities, new commodity indices search the full term structure, and choose maturities as far as one year ahead. The exact maturity choice usually depends on the term structure of the current market. If the term structure is in contango, they roll into contracts with long maturities to reduce the frequency of rolling and thus the roll cost. If the term structure is in backwardation, they roll into the contracts with close maturities to take advantage of

the positive roll yields.

This is consistent with the classic limits to arbitrage theory by Shleifer and Vishny (1997). The arbitrage profit is lower when there is a reduction in size of index investment and an increase in the amount of arbitrage capital in the futures markets. The performance of front-running the Goldman roll is determined by the net result of two opposite forces. To confirm this correlation, I run the following panel regressions for index commodities:

$$Ret^{i,t} = \alpha + \beta_2 I^{i,t}_{inIndex} + \beta_3 I^{i,t}_{inIndex} \times NetRatio^{i,t} + Controls + u_{i,t}. \quad (8)$$

where the dependent variable $Ret^{i,t}$ is Strategy's average excess return in the trading period of commodity i in year t and $I^{i,t}_{inIndex}$ is the indicator variable specified in the last section, which is equal to 1 if commodity i is actually included in the SP-GSCI in year t and 0 if otherwise. $NetRatio^{i,t} = IndexRatio^{i,t} - SpreadRatio^{i,t}$ measures net result of the two forces, where $IndexRatio^{i,t}$ is the average ratio of index investment in commodity i relative to the value of its total open interest and $SpreadRatio^{i,t}$ is the average ratio of spread position held by speculators relative to total open interest.

The data on investment tied to the SP-GSCI and DJ-UBSCI are not publicly available. Master and White (2008) use sources of Bloomberg, Goldman Sachs and CFTC reports to construct an annual series of estimated investment tied to the two indices from 1991 to 2008 (first half year). In addition, they estimate that the SP-GSCI had about 63% market share and the DJ-UBSCI had about 32% market share in 2008. Another important data source is the quarterly CFTC reports of index investments starting from the fourth quarter of 2007, which have data on the values of total index investment. I only consider the value of long positions in the CFTC's reports, and the quarterly data is converted into annual data by using the average of four quarters in one calendar year. Using the estimated market shares, I construct the values of investment tied to the SP-GSCI and DJ-UBSCI in 2008 and 2009. For each index, total value of investment tied to it is then allocated to individual commodity according to its weighting scheme each year, and for individual commodity, the total value of index investment is equal to sum of investment from the two indices. The variable $IndexRatio^{i,t}$ is equal to the value of index investment in commodity i in year t divided by the commodity's total market value in year t, which is average value of total

open interests in year t. The data on total open interests and spread positions held by noncommercial traders can be obtained from the CFTC's COT reports.

As reported in Column 1 and 3 of 4.6, the coefficient β_3 is statistically positive for both strategies, especially for Strategy 1, whose average excess return increases by 0.96 bps with 1% increase in the net ratio. Column 2 and 4 of 4.6 shows that the results are robust if we only consider index commodities after they were included in the SP-GSCI.

To conclude, the exercise provides empirical evidence that a market anomaly can exist and persist due to slow-moving arbitrage capital and the resulting delayed arbitrage. As more people become aware of the price impact, more arbitragers will exploit it and index investors will also move their investments into better designed commodity indices.

4.4 Cost of the Price Impact

It has been very profitable to exploit the price impact of the Goldman roll, but from the perspective of index investors, how costly was the price impact? In this section, I will estimate the cost of the price impact by comparing two excess return indices. Since the SP-GSCI was launched at the end of 1991, I consider the period starting from 1992 for the estimation.

On January 2 1992, $100 dollars were assumed to be invested in futures contracts of the 19 index commodities. The investment that each commodity receives from the $100 is proportional to its SP-GSCI weight in 2010. To focus on the cost of the price impact, there is no re-balancing and the choice of futures contracts to hold is exactly the same as the SP-GSCI. I construct two indices with different rolling periods. One index rolls the futures forward in the SP-GSCI's rolling period, and is labeled "SP-GSCI Roll" index, so this index rolls exactly the same as the Goldman roll. The other index rolls just 10 business days earlier, in the first 5 days of the 15-day rolling window we discussed previously, and is labeled "Earlier Roll" index. The interest earned on collateral is not considered, so the indices are excess return indices.

As shown in Panel A of 4.6, the values of the two indices closely tracked each other before 2000, and then started to deviate far away. Although the two indices still shared the same

Dependent variable: $Ret^{i,t}$				
	Strategy 1		Strategy 2	
	$I^i_{IndexCom}=1$	$I^{i,t}_{inIndex}=1$	$I^i_{IndexCom}=1$	$I^{i,t}_{inIndex}=1$
	1	2	3	4
$Constant$	0.062**	0.24***	−0.013	0.15***
	(0.029)	(0.05)	(0.026)	(0.039)
$I^{i,t}_{inIndex}$	0.20***		0.18***	
	(0.055)		(0.037)	
$NetRatio^* I^{i,t}_{inIndex}$	0.96***	1.04***	0.33*	0.39**
	(0.25)	(0.24)	(0.17)	(0.17)
$Controls$				
$RY^{i,t}$	0.002	0.002	0.021	0.027
	(0.019)	(0.025)	(0.013)	(0.018)
$growth^t_{GDP}$	0.064***	0.078**	0.052***	0.061***
	(0.021)	(0.024)	(0.010)	(0.013)
$Inflation^t$	0.012	−0.009	0.002	−0.007
	(0.022)	(0.049)	(0.019)	(0.041)
R^2_{adj}	12.16%	9.26%	11.12%	8.40%
obs	404	287	404	287

Table 4.6: Regressions on the Trading Strategys' Excess Returns 2

pattern in the period 2000-2010 due to the same exposure to the spot returns, the "Earlier Roll" index outperformed the "SP-GSCI Roll" index because its roll yields were higher. When commodity prices reached heights in mid-2008, the "SP-GSCI Roll" index reached a peak value \$725, while the "Earlier Roll" index reached \$1099, with out-performance of \$374.

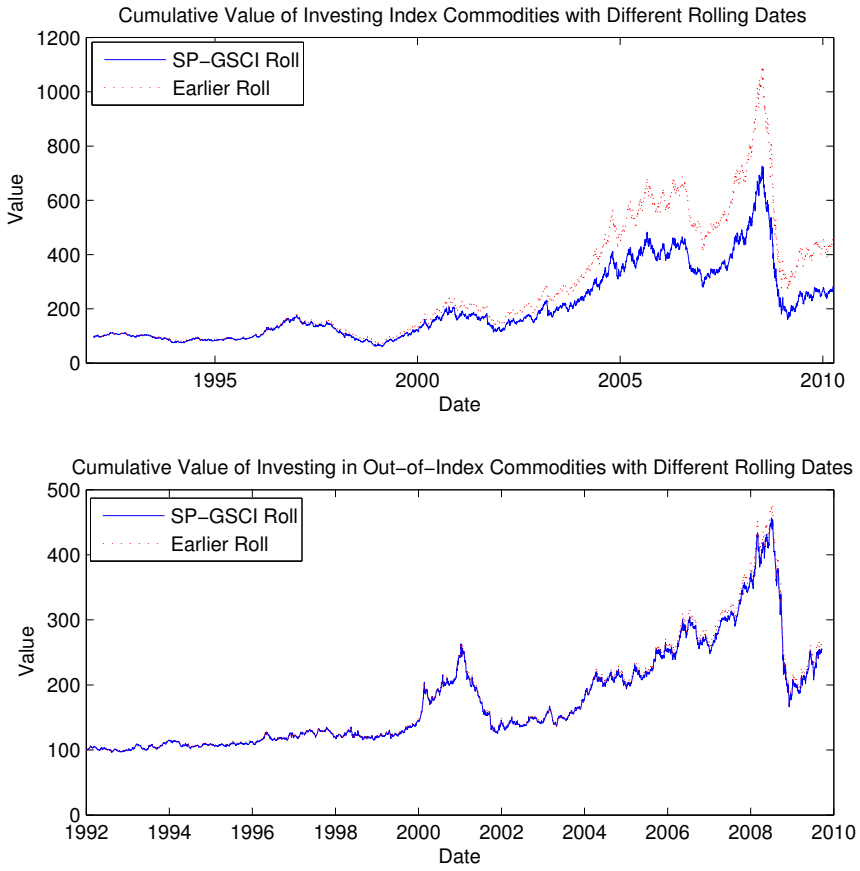

Figure 4.6: Value of Two Indices with Different Rolling Dates

As a comparison, I also picked from the control group 12 out-of-index commodities that have data back to 1992 and until 2009. Since there are no reference weights, equal weighting is applied to each commodity. The same two rolling rules are applied to form the same two indices: "SP-GSCI Roll" and "Earlier Roll". As shown in Panel B of 4.6, there is no detectable difference between the values of two indices in the whole period 1992-2009.

The maximum difference between the two indices was only about $6.

4.7 reports the summary statistics of the two indices' annualized excess returns. The full period is divided into two sub-periods: 1992-1999 and 2000-2009[8]. For the 19 index commodities, the excess returns of the two indices had almost the same standard deviations and skewness in both periods, but the means are quite different. The "SP-GSCI Roll" index yielded an annual excess return of 2.31% before 2000 and 7.93% since 2000, while the "Earlier Roll" index outperformed it annually by 1.66% and 3.59% respectively. Therefore, the Sharpe ratio of the "Earlier Roll" index was 82% higher in the period 1992-1999 and 48% higher in the period 2000-2009. In addition, the difference in excess returns had a positive skewness 0.43 before 2000, and 0.79 from 2000, which indicates the arbitrage opportunity induced by the price impact. It is also statistically significant that the mean difference in excess returns in the period 2000-2009 is larger than the mean difference in the period 1992-1999, which suggests that when index investment grew larger, index investors endured a higher cost of the price impact.

In a clear contrast, for the 12 out-of-index commodities, all the summary statistics of the two indices are roughly the same in both periods. Although the "Earlier Roll" index was still slightly better, the out-performance was very small, only about 0.25%, and the difference of excess returns were not always positively skewed.

In order to estimate the cost of the price impact in absolute amount, I collect the data on total commodity index investment from Masters and White (2008) and the CFTC's reports of index investment. All investments are assumed to be tied to the "SP-GSCI Roll" index. Each year, the cost due to the price impact is estimated by the size of index investment multiplied by the average difference of excess returns between the "SP-GSCI Roll" index and "Earlier Roll" index in this year. As shown in 4.7, as the index investment grew, the cost also grew fast. From 2004, investing in the "SP-GSCI Roll" index lost over $2 billion every year to the "Earlier Roll" index, and in 2009, the loss reached a maximum of $8.4 billion.

[8]The returns in 2010 are excluded because the propane data ends in Sep 2009 and I want to include one energy commodity in the group of out-of-index commodities. However, the results are very similar if data of 2010 is included and propane is excluded.

	1992-1999			2000-2009			
	SP-GSCI Roll	Earlier Roll	Diff.	SP-GSCI Roll	Earlier Roll	Diff.	DID
19 Index Commodities							
Mean	2.31%	3.97%	1.66%	7.93%	11.52%	3.59%	1.93%**
Sd	21.5%	20.3%	2.20%	34.4%	34.1%	2.36%	
Skewness	0.02	0.05	0.43	−0.3	−0.3	0.79	
Sharpe Ratio	0.11	0.20		0.23	0.34		
12 Out-of-Index Commodities							
Mean	4.67%	4.90%	0.23%	5.61%	5.87%	0.26%	0.03%
Sd	11.4%	11.4%	1.14%	20.1%	20.2%	1.02%	
Skewness	0.008	−0.001	2.19	−0.2	−0.2	−0.31	
Sharpe Ratio	0.41	0.43		0.28	0.29		

Table 4.7: Summary Statistics of Two Indices with Different Rolling Periods

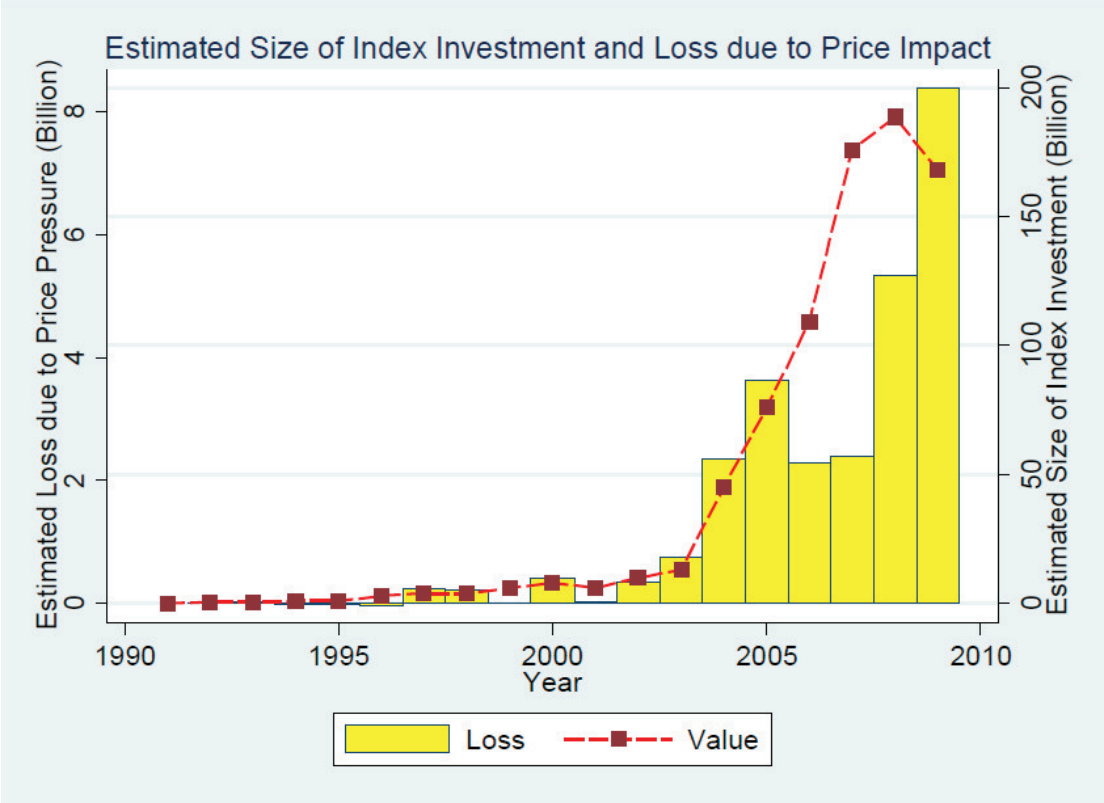

Figure 4.7: Estimated Size of Index Investment and Loss due to Price Impact

In sum, because the massive shorting and longing of futures contracts exerts very high price pressure in the rolling period, the resulting price impact has been very costly to index investors in terms of both forgone excess return and absolute amount of loss.

Chapter 5

Conclusions

Since index funds have very low management fees, investors usually perceive index investment as an inexpensive way to gain broad market exposure. While it seems to be true for the equity index funds, this paper shows that index investment can be very expensive in the commodity markets due to the large price impact of index investors' mechanical rolling forward of futures contracts. Equity index funds invest directly in the underlying assets, so the fund managers rarely need to change positions besides the inflow and outflow of new funds. While there are some documented inefficiencies in equity investment, like the inclusion effect, the resulted costs are quite small, because the inefficiencies only happen at very low frequency and arbitrageurs in the equity markets are very competitive. Commodity index investment is very different, because investors take long positions in commodity futures contracts. Since futures contracts have expiration dates, commodity index investors have to roll their entire positions forward at monthly frequency, which resulted a very high cost due to the large price impact of this rolling activity. Commodity index investors lost on average 3.6% annual excess returns due to the price impact. In absolute terms, the estimated loss amounted to a total of $26 billion over the period 2000 to 2009, while the estimated total management fee was only about $5 billion. This magnitude of economic loss dwarfs the cost of price impact in the S&P 500 equity index due to the inclusion and exclusion effect, which was about 0.21-0.28% each year on average from 1990 to 2005 estimated by Petajisto (2010). In absolute terms, Petajisto assumed total assets of $1.2 trillion tied to the S&P 500, and the estimated annual average costs were $2.5-3.4 billion.

Concern about price impact motivated some second-generation commodity indices to have longer rolling periods so that the price pressure on each rolling date is very small. One example is the UBS Constant Maturity Commodity Index (CMCI), which roll futures forward at daily frequency. However, it seems that many other new indices still did not recognize the possible price impact, because they still have very short rolling periods, and the new rolling methodologies are mainly designed to reduce the roll cost in the current contango markets. As discussed, these indices tend to roll into contracts with long maturities, but these contracts are not as liquid as the contracts with short maturities, so the price impact of the rolling activity could be quite large even though the investment tied to these indices is not very large. As these indices get more popular, the price impact and the resulted cost can be even larger.

This extends to a more general question of security design. Commodity indices are very different from the traditional securities, because investing in them requires continuous management due to the special rolling requirement. Therefore, the designer has to think about the possible negative effects of fixed management actions when the assets under management grow larger, and whether the designed index will be immune to these effects. Another problem is that as the designer tries to minimize the potential negative effects, the management rules could become very complicated. Since the complexity of the index increases, the cost of replicating it and thus the management fee increases, and investors may feel that it is more difficult to understand and analyze the index. There is a balance between the potential benefits and costs associated with the complexity of securities. These problems also apply to the design of exchange-traded funds (ETFs), which are becoming more and more popular among investors.

Although the market anomaly created by the Goldman roll can be arbitraged away by enough arbitrageurs, the impact of index investment will not disappear. As more and more arbitrageurs try to front-run the Goldman roll and also each other, they can spread the price impact out to other dates and also other maturities. This can have a profound effect on the term structure of commodity futures markets, and may potentially be one of the reasons why the term structures of many index commodities have moved from backwardation towards contango in recent years. Further research can investigate this hypothesis and look at the

impact of index investment on commodity term structure.

Bibliography

[1] Abreu, D. and M. Brunnermeier (2002), "Synchronization Risk and Delayed Arbitrage," *Journal of Financial Economics* 66, 341-360.

[2] Acharya, V., L. Lochstoer and T. Ramadorai (2010), "Limits to Arbitrage and Hedging: Evidence from Commodity Markets," Working Paper.

[3] Anson, M. (1998), "Spot Returns, Roll Yield, and Diversification with Commodity Futures," *Journal of Alternative Investment*, Winter, 16-32.

[4] Barberis, N. and R. Thaler (2001), "A Survey of Behavior Finance," In: Constantinides, G., Harris, M. and R. Stulz, Handbook of the Economics of Finance. Elsevier.

[5] Bessembinder, H. (1992), "Systematic Risk, Hedging Pressure, and Risk Premiums in Futures Markets," *Review of Economic Studies* 5, 637-667.

[6] Bessembinder, H., J. Coughenour, P. Seguin and M. Smoller (1995), "Mean Reversion in Equilibrium Asset Prices: Evidence from the Futures Term Structure," *Journal of Finance* 50, 361-375.

[7] Bessembinder, H. and M. Lemmon (2002), "Equilibrium Pricing and Optimal Hedging in Electricity Forward Markets," *Journal of Finance* 62, 1347-1382.

[8] Buyuksahin, B. and J. Harris (2009), "The Role of Speculators in the Crude Oil Futures Market", *Energy Journal*, forthcoming.

[9] Buyuksahin, B., M. Haigh, J. Harris, J. Overdahl and M. Robe (2008), "Fundamentals, Trader Activity and Derivative Pricing," Technical Report, Commodity Futures Trading Commission.

[10] Casassus, J. and P. Collin-Dufresne (2005), "Stochastic Convenience Yield Implied from Commodity Futures and Interest Rates," *Journal of Finance* 60, 463-520.

[11] Chang, E. (1985), "Returns to Speculators and the Theory of Normal Backwardation," *Journal of Finance* 40(1), 193-208.

[12] Chen, H., G. Noronha and V. Singal (2004), "The Price Response to S&P 500 Index Additions and Deletions: Evidence of Asymmetry and a New Explanation," *Journal of Finance* 59, 1901-1929.

[13] Cootner, P. (1960), "Return to Speculators: Telser vs. Keynes," *Journal of Political Economy* 68, 396-404.

[14] Cremers, M., A. Petajisto and E. Zitzewitz (2010), "Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation," Working Paper.

[15] Deaton, A. and G. Laroque (1992), "On the Behaviour of Commodity Prices," *Review of Economic Studies* 59, 1-23.

[16] DeLong, J., Shleifer, A., Summers, L. and R. Waldmann (1990), "Noise Trader Risk in Financial Markets," *Journal of Political Economy* 98(4), 703-738.

[17] De Roon, F., T. Nijman and C. Veld (2000), "Hedging Pressure Effects in Futures Markets," *Journal of Finance* 55, 1437-1456.

[18] Duffie, D. (2010), "Presidential Address: Asset Price Dynamics with Slow-Moving Capital," *Journal of Finance* 65, 1237-1267.

[19] Feldman, B. and H. Till (2006), "Backwardation and Commodity Futures Performance: Evidence from Evolving Agricultural Markets," *Journal of Alternative Investment*, Winter, 24-39.

[20] Gorton, G. and G. Rouwenhorst (2006), "Facts and Fantasies about Commodity Futures," *Financial Analysts Journal* 62(2), 47-68.

[21] Hamilton, J. (2009), "Causes and Consequences of the Oil Shock of 2007-2008," Working Paper, UC San Diego.

[22] Harris, L. and E. Gurel (1986), "Price and Volume Effects Associated with Changes in the S&P 500: New Evidence for the Existence of Price Pressures," *Journal of Finance* 41, 815-829.

[23] Hicks, J. (1939), Value and Capital (Lodon: Oxford University Press).

[24] Hirshleifer, D. (1988), "Residual Risk, Trading Costs, and Commodity Futures Risk Premia," *Review of Financial Studies* 1, 173-193.

[25] Hirshleifer, D. (1990), "Hedging Pressure and Futures Price Movements in a General Equilibrium Model," *Econometrica* 58, 411-428.

[26] Idzorek, T. (2006), "Strategic Asset Allocation and Commodities," Ibbotson Associates White Paper.

[27] Kaul, A., V. Mehrotra and R. Morck (2000), "Demand Curves for Stocks Do Slope Down: New Evidence from an Index Weights Adjustment," *Journal of Finance* 55, 893-912.

[28] Keynes, J. (1930), A Treatise on Money, vol. 2 (London: Macmillan).

[29] Kilian, Lutz (2009), "Not All Oil Price Shocks are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review* 99, 1053-1069.

[30] Lynch, A. and R. Mendenhall (1997), "New Evidence on Stock Price Effects Associated with Changes in the S&P 500 Index," *Journal of Business* 70, 351-383.

[31] Masters, M. and A. White (2008), "The Accidental Hunt Brothers: How Institutional Investors are Driving up Food and Energy Prices."

[32] Morck, R. and F. Yang (2001), "The Mysterious Growing Value of S&P Index Membership," NBER Working Paper.

[33] Nash, D. (2001), "Long-Term Investing in Commodities," *Global Pensions Quarterly*, January, 30-36.

[34] Petajisto, A. (2010), "The Index Premium and Its Hidden Cost for Index Funds," *Journal of Empirical Finance*, forthcoming.

[35] Routledge, B., D. Seppi and C. Spatt (2000), "Equilibrium Forward Curves for Commodities," *Journal of Finance* 55, 1297-1338.

[36] Schwartz, E. (1997), "The Stochastic Behavior of Commodity Prices: Implication for Valuation and Hedging," *Journal of Finance* 52, 923-972.

[37] Shleifer, A. (1986), "Do Demand Curves for Stocks Slope Down?" *Journal of Finance* 41, 579-590.

[38] Shleifer, A. (2000), "Arbitrage and Finance–An Introduction to Behavioral Finance," Oxford University Press, Oxford.

[39] Shleifer, A. and R. Vishny (1997), "The Limits of Arbitrage," *Journal of Finance* 52, 35-56.

[40] Singleton, K. (2010), "The 2008 Boom/Bust in Oil Prices," Research Report.

[41] Stoll, H. and R. Whaley (2010), "Commodity Index Investing and Commodity Futures Prices," *Journal of Applied Finance* 20(1), 7-46.

[42] Tang, K. and X. Wei (2010), "Index Investing and the Financialization of Commodities," Working Paper.

Part II

Learning about Consumption Dynamics

Chapter 6

Abstract

This paper studies the asset pricing implications of Bayesian learning about the parameters, states, and models determining aggregate consumption dynamics. Our approach is empirical and focuses on the quantitative implications of learning in real-time using post World War II consumption data. We characterize this learning process and find that revisions in beliefs stemming from parameter and model uncertainty are significantly related to realized aggregate equity returns. This evidence is novel, providing strong support for a learning-based story. Further, we show that beliefs regarding the conditional moments of consumption growth are strongly time-varying and exhibit business cycle and/or long-run fluctuations. Much of the long-run behavior is unanticipated ex ante. We embed these subjective beliefs in a general equilibrium model to investigate further asset pricing implications. We find that learning significantly improves the model's ability to fit standard asset pricing moments, relative to benchmark model with fixed parameters. This provides additional evidence supporting the importance of learning.

Chapter 7

Introduction

This paper studies the asset pricing implications of learning about aggregate consumption dynamics. We are motivated by practical difficulties generated from the use of complicated consumption-based asset pricing models with many difficult-to-estimate parameters and latent states. For example, parameters or states controlling long-run consumption growth are at once extremely important for asset pricing and particularly difficult to estimate. Thus, we are interested in studying an economic agent who is burdened with some of the same econometric problems faced by researchers, a problem suggested by Hansen (2007).[1]

A large existing literature studies asset pricing implications of statistical learning – the process of updating beliefs about uncertain parameters, state variables, or even model specifications. Pastor and Veronesi (2009) provide a recent survey. In theory, learning can generate a wide range of implications relating to stock valuation, levels and variation in expected returns and volatility, and time series predictability, with many of the results focussed on the implications of learning about dividend dynamics.

Our analysis differs from existing work along three key dimensions. First, we focus

[1]Hansen (2007) states: *"In actual decision making, we may be required to learn about moving targets, to make parametric inferences, to compare model performance, or to gauge the importance of long-run components of uncertainty. As the statistical problem that agents confront in our model is made complex, rational expectations' presumed confidence in their knowledge of the probability specification becomes more tenuous. This leads me to ask: (a) how can we burden the investors with some of the specification problems that challenge the econometrician, and (b) when would doing so have important quantitative implications" (p.2).*

on the empirical implications of simultaneously learning about parameters, state variables, and even model specifications. Most existing work focuses on learning a single parameter or state variable. Learning about multiple unknowns is more difficult as additional unknowns often confounds inference, slowing the learning process. Second, we focus on the specific implications of real-time learning about consumption dynamics from macroeconomic data during the U.S. post World War II experience. Thus, we are not expressly interested in general asset pricing implications of learning in repeated sampling settings, but rather the *specific* implications generated by the historical macroeconomic shocks realized in the United States over the last 65 years. Third, we use a new and stringent test of learning that relates updates in investor beliefs to contemporaneous, realized equity returns.

In studying the implications of learning, we focus on the following types of questions. Could an agent who updates his beliefs rationally detect non-i.i.d. consumption growth dynamics in real time? How rapidly does the agent learn about parameters and models? Are the revisions in beliefs about consumption moments correlated with asset returns, as a learning story would require? Is there evidence that learning effects can help us understand standard asset pricing puzzles, such as the high equity premium, return volatility, and degree of return predictability?

One of the key implications of learning is that the agent's beliefs are nonstationary. For example, the agent may gradually learn that one model fits the data better than an alternative model or that a parameter value is higher or lower than previously thought, both of which generate nonstationarity in beliefs. The easiest way to see this is to note that the posterior mean of a parameter, $\mathbb{E}\left[\theta|y^t\right]$, where y^t is data up to time t, is trivially a martingale. Thus revisions in beliefs represent permanent, nonstationary shocks, that can have important asset pricing implications. For instance, nonstationary dynamics can generate a quantitatively important wedge between *ex post* outcomes and *ex ante* beliefs, providing an alternative explanation for standard asset pricing quantities such as the observed equity premium or excess return predictability.[2]

We study learning in the context of three standard Markov switching models of consumption growth: unrestricted 2- and 3-state models and a restricted 2-state model that

[2]See also Cogley and Sargent (2008), Timmermann (1993), and Lewellen and Shanken (2002).

generates i.i.d. consumption growth. The states capture business cycle fluctuations and can be labeled as expansion and recession in 2-state models, with an additional 'disaster' state in 3-state models.[3] Our key assumption is that the agent views the parameters, states, and even models as unknowns, using Bayes rule to update beliefs using consumption data, as well as additional macroeconomic data such as GDP growth in extensions.

To focus on different aspects of learning, we consider three sets of initial parameter beliefs. The first, the 'historical prior,' trains the prior using Shiller's consumption data from 1889 until 1946, a common approach to generate 'objective' priors.[4] The second, the 'look-ahead prior,' sets prior parameter means to full-sample maximum likelihood point estimates using post World War II data. We embed substantial uncertainty around these estimates to study the effect of parameter uncertainty. This is often called an 'empirical Bayes' approach. The third, the fixed parameter prior, is a rational expectations benchmark with dogmatic beliefs that are fixed at the end-of-sample parameter estimates. Thus, there is no parameter uncertainty. There is state uncertainty, however, which allows us to separate the effects of parameter and state uncertainty.

Our first results characterize the beliefs about parameters, states, models, and future consumption dynamics (e.g., moments) through the sample. The perceived dynamic behavior of aggregate consumption is at the heart of consumption-based asset pricing as it, jointly with preferences, determines the dynamic properties of the pricing kernel. In terms of beliefs, we compute at each point in time the posterior distribution of parameters, states, and models. As new data arrives, we update beliefs using Bayes rule. In addition to usual sum-

[3]Markov switching models for consumption or dividends are a benchmark specification in the literature, see, e.g., Mehra and Prescott (1985), Rietz (1988), Cecchetti, Lam, and Mark (1990, 1993), Whitelaw (2000), Cagetti, Hansen, Sargent, and Williams (2002), Barro (2006), Barro and Ursua (2008), Chen (2008), Bhamra, Kuehn, and Strebulaev (2008), Barro, Nakamura, Steinsson and Ursua (2009), Backus, Chernov, and Martin (2009), and Gabaix (2009). Rietz (1988) and, more recently, Barro (2006, 2009) argue that consumption disaster risk can help explain some of the standard macro-finance asset pricing puzzles.

[4]We do account for measurement error, which likely increased reported macroeconomic volatility during the pre-war period, as argued in Romer (1989). Malmendier and Nagel (2011) present evidence that the experience of the Great Depression affected investors' subsequent beliefs about risk and return, broadly consistent with the Historical prior calibration approach.

maries of parameters and states, we also compute model probabilities and perform 'model monitoring' in real time as new data arrives. We find that the posterior probability of the i.i.d. model falls dramatically over time, provided the prior weight is less than one. Thus our agent is able to learn in real-time that consumption growth is not i.i.d., but has persistent components.[5] The agent believes that expected consumption growth is low in recessions and high in expansions, with the opposite pattern for consumption growth volatility. The 2-state model quickly emerges as the most likely, but the 3-state model with a disaster state has $5-10\%$ probability at the end of the sample. At the onset of the financial crisis in 2008, the probability of the disaster model increases.[6]

There is significant learning about the expansion state parameters, slower learning about the recession state, and almost no learning about the disaster state, as it is rarely, if ever, visited. Thus, there is an observed differential in the speed of learning. Standard large sample theory implies that all parameters converge at the same rate, but the realized convergence rate depends on the actual observed sample path. There is also strong evidence for nonstationary time-variation in the conditional means and variances of consumption growth, as well as measures of non-normality such as skewness and kurtosis. For both the historical and the look-ahead priors, the agent's perception of the long-run mean (volatility) of consumption growth generally increases (decreases) over the sample. The perceived persistence of recessions (expansions) decreases (increases).[7] As the agent's beliefs about these parameters and moments change, asset prices and risk premia will also change.

The first formal test of the importance of learning regresses contemporaneous excess stock market returns on revisions in beliefs about expected consumption growth. This test, which to our knowledge is new to the literature, is a fundamental implication of any learning-based explanation: for learning to matter, unexpected revisions in beliefs

[5]This result is robust to persistence induced by time-aggregation of the consumption data (see Working (1960)).

[6]The posterior probability of the three-state model would change dramatically, if visited. For example, if a -3% quarterly consumption growth shock were realized at the end of the sample, the posterior probability of the three-state model would increase to almost 50%.

[7]All of the results described in the current and previous paragraphs are robust to learning from additional GDP growth data.

about expected consumption growth should be reflected in the unexpected aggregate equity returns.[8]. We find strong statistical evidence that this relationship is positive, and the results are similar for both the historical and the look-ahead prior. To disentangle parameter from state learning, we include revisions in beliefs generated by the fixed parameter prior as a control. Revisions in beliefs obtained using the historical and look-ahead priors remain statistically significant, but revisions in beliefs generated by models with known parameters are statistically insignificant.

These results imply that learning about parameters and models is a statistically significant determinant of asset returns in our sample, confirming our main hypothesis. This result is strengthened if the agent learns from both consumption and GDP growth. It is important to note that our agent only learns in real-time and from macroeconomic fundamentals, as no asset price data (such as the dividend-price ratio) is used when forming beliefs. Since the revisions in beliefs obtained from the models with fixed parameters are statistically insignificant, the evidence questions the standard full-information, rational expectations implementation of the standard consumption-based model, at least for the models of consumption dynamics that we consider.[9]

As mentioned earlier, parameter and model learning generate nonstationary dynamics and permanent shocks that could have important implications. To investigate these implications, we consider a formal asset pricing exercise assuming Epstein-Zin preferences. Because the specific time-path of beliefs is important, the usual calibration and simulation approach used in the literature is not applicable, and we consider the following alternative pricing procedure. At time t, given beliefs over parameters, models, and states, our agent prices a levered claim to a future consumption stream, computing quantities such as *ex-*

[8]The sign of the effect would in a model depend on the elasticity of intertemporal substitution, and also on the other moments that change at the same time (volatility, skewness, kurtosis, etc.). In the model section, we show that this positive relation is consistent with a model with an elasticity of intertemporal substitution greater than 1.

[9]Parameter and model learning, on the one hand, and state learning on the other hand are distinct in our setting because the former generates a non-stationary path of beliefs, while the latter, after an initial burn-in period, is stationary.

ante expected returns and dividend-price ratios.[10] Then, at time $t+1$, our agent updates beliefs using new macro realizations at time $t+1$, recomputes prices, expected returns and dividend-price ratios. From this time series of prices, we compute realized equity returns, volatilities, etc. Thus, we feed historically realized macroeconomic data into the model and analyze the asset pricing implications for various models and prior specifications. This process is required when the time path matters and was previously used in, for example, Campbell and Cochrane (1999), where habit is a function of past consumption growth. We use standard preference parameters taken from Bansal and Yaron (2004).

Solving the full pricing problem with priced parameter uncertainty is computationally prohibitive, as the dimensionality of the problem is too large.[11] To price assets in a tractable way, while still incorporating learning, we follow Piazzesi and Schneider (2010) and Cogley and Sargent (2009) and use a version of Kreps' (1994) anticipated utility. This implies that our agent prices claims at each point in time using current posterior means for the parameters and model probabilities, assuming those values will persist into the indefinite future. We do account for state uncertainty when pricing.

This pricing experiment provides additional evidence, along multiple dimensions, for the importance of learning. Focussing on the 3-state model, we first note that the model with parameters fixed at the full-sample values has a difficult time with standard asset pricing moments: the realized equity premium and Sharpe ratio are less than half the values observed in the data. The volatility of the price-dividend ratio is eighty percent less than the observed value. Parameter learning uniformly improves *all* of these statistics, bringing them close to observed values. The results are, after a burn-in period, similar for the look-ahead and the historical prior as the agent quickly unlearns the mean parameter beliefs of the look-ahead prior early in the sample. It is important to note that this is not a calibration

[10] We do price a levered consumption claim and introduce idiosyncratic noise to break the perfect relationship between consumption and dividend growth. The dividends are calibrated to match the volatility of dividend growth and the correlation between dividend and consumption growth.

[11] As an example, for the 3-state model there are twelve parameters, each with two hyperparameters characterizing the posteriors. This implies that we would have to have to solve numerically for prices on a very high dimensional grid, which is infeasible. There are additional difficult technical issues associated with priced parameter uncertainty, as noted by Geweke (2001) and Weitzman (2007).

exercise – we did not choose the structural parameters to generate these returns.

The increase in the realized equity premium and return volatility is due to unexpected revisions in beliefs resulting from the parameter and model learning. In particular, the average annualized *ex ante* quarterly risk premium is similar across the models at about 1.7%, but the models with uncertain parameters generate a higher realized equity premium of about 3.8% to 4.2%, close to the 4.7% observed over the sample. This documents a dramatic impact of the specific time path of beliefs about parameters and models for standard asset pricing statistics, at least relative to the fixed parameter, rational expectations benchmark. This also implies, looking forward, that the perceived equity premium is much smaller than the realized equity premium over the post World War II period. These points are consistent with the results in Cogley and Sargent (2008).[12]

In terms of predictability, the returns generated by learning over time closely match the data. For the historical and look-ahead priors and for forecasting excess market returns with the lagged log dividend-price ratio, the generated regression coefficients and R^2's are increasing with the forecasting horizon and similar to those found in the data. The fixed parameters case, however, does not deliver significant *ex post* predictability, although the *ex ante* risk premium is in fact time-varying in these models as well because the risk premium time-variation assuming fixed parameters is too small relative to the volatility of realized returns to result in significant t-statistics. The intuition for why *in-sample* predictability occurs when agents are uncertain about parameters and models is the same as in Timmermann (1993) and Lewellen and Shanken (2002) – unexpected updates in growth and discount rates impact the dividend-price ratio and returns in opposite directions leading to the observed positive in-sample relation. Thus, in-sample predictability can be expected with parameter and model learning. The quantitatively large degree of in-sample relative

[12]Cogley and Sargent (2008) assume negatively biased beliefs about the consumption dynamics to highlight the same mechanism and also consider the role of robustness. In their model, the subjective probability of recessions is higher than the 'objective' estimate from the data. The results we present here are consistent with their conclusions, but our models are estimated from fundamentals in real-time, which allows for an out-of-sample examination of the time-series of revisions in beliefs. Further, we allow for learning over different models of the data generating process, as well as *all* the parameters of each model.

to out-of-sample predictability we find is consistent with the literature.[13]

We also note that the model exhibits volatile long maturity risk-free yields, consistent with the data. Learning about fixed quantities such as models or parameters generate permanent shocks that affect agents' expectations of the long-run (infinite-horizon) distribution of consumption growth. This is different from existing asset pricing models where only stationary variables affect marginal utility growth (see, e.g., Bansal and Yaron (2004), and Wachter's (2005) extension of Campbell and Cochrane (1999) model, as well as our fixed parameters benchmark model). In these models, long-run (infinite-horizon) risk-free yields are constant as the transitory shocks to marginal utility growth die out in the long run. This is additional evidence supporting a learning-based explanation relative to the fixed parameters alternative.

In conclusion, our results strongly support the importance of parameter and model learning for understanding the joint behavior of consumption and asset prices in the U.S. post World War II sample. First, parameter and model learning leads to a time path of belief revisions that are correlated with realized equity returns, controlling for realized consumption growth. Second, the time series of beliefs help explain the time-series of the price level of the market (the time-series of the price-dividend ratio) in a general equilibrium model. Third, beliefs display strong nonstationarity over time, driving a wedge between *ex ante* beliefs and *ex post* realizations that is absent in rational expectations models. Fourth, permanent shocks to beliefs generate permanent shocks to marginal utility growth. These features help explain common asset pricing puzzles such as excess return volatility, the high sample equity premium, the high degree of in-sample return predictability, and the high volatility of long-run yields, all relative to a fixed parameter alternative. The results are generated by real-time learning from consumption (and GDP growth), using standard preference parameters without directly calibrating to asset returns. In this sense the results are entirely "out-of-sample."

[13]For example, Fama and French (1988) document a high degree of in-sample predictability of excess (long-horizon) stock market returns using the price-dividend ratio as the predictive variable. On the other hand, Goyal and Welch (2008) and Ang and Bekaert (2007) document poor *out-of-sample* performance of these regressions in the data, and the historical and look-ahead prior learning models presented here are consistent with this evidence.

Chapter 8

The Environment

8.1 Model

We follow a large literature and assume an exogenous Markov or regime switching process for aggregate, real, per capita consumption growth dynamics. Log consumption growth, Δc_t, evolves via:

$$\Delta c_t = \mu_{s_t} + \sigma_{s_t}\varepsilon_t, \tag{8.1}$$

where ε_t are i.i.d. standard normal shocks, $s_t \in \{1, ..., N\}$ is a discretely-valued Markov state variable, and $\left(\mu_{s_t}, \sigma^2_{s_t}\right)$ are the Markov state-dependent mean and variance of consumption growth. The Markov chain evolves via a $N \times N$ transition matrix Π with elements π_{ij} such that $\text{Prob}[s_t = j | s_{t-1} = i] = \pi_{ij}$, with the restriction that $\sum_{j=1}^{N} \pi_{ij} = 1$. The fixed parameters of the N-state model contain the means and variances in each state, $\left\{\mu_n, \sigma^2_n\right\}_{n=1}^{N}$ as well as the elements of the transition matrix. The transition matrix controls the persistence of the Markov state.

Markov switching models are flexible and tractable and have been widely used since Mehra and Prescott (1985) and Rietz (1988). By varying the number, persistence, and distribution of the states, the model can generate a wide range of economically interesting and statistically flexible distributions. Although the ε_t's are i.i.d. normal and the distribution of consumption growth, conditional on s_t and parameters, is normally distributed, the distribution of future consumption growth is neither i.i.d. nor normal due to the shifting Markov state. This time-variation induces very flexible marginal and predictive distribu-

tions for consumption growth. These models are also tractable, as it is possible to compute likelihood functions and filtering distributions, given parameters.

We consider two and three state models and also consider a restricted version of the two state model generating i.i.d consumption growth by imposing the restriction $\pi_{11} = \pi_{21}$ and $\pi_{22} = \pi_{12} = 1 - \pi_{11}$. Under this assumption, consumption growth is an i.i.d. mixture of two normal distributions, essentially a discrete-time version of Merton's (1976) mixture model. The general two and 3-state models have 6 and 12 parameters, respectively. The i.i.d. two state model has 5 parameters ($\mu_1, \mu_2, \sigma_1, \sigma_2$ and π_{11}).

It is common in these models to provide business cycle labels to the states. In a 2-state model, we interpret the two states as 'recession' and 'expansion,' while the three state model additionally allows for a 'disaster' state.[1] Although rare event models have been used for understanding equity valuation since Rietz (1988), there has been a recent resurgence in research using these models (see, e.g., Barro (2006, 2009), Barro and Ursua (2008), Barro, Nakamura, Steinsson and Ursua (2009), Backus, Chernov, and Martin (2009), and Gabaix (2009)).

8.2 Information and learning

To operationalize the model, additional assumptions are required regarding the economic agent's information set. Since we want to model learning similar to that faced by the econometrician, we assume agents observe aggregate consumption growth, but are uncertain about the Markov state, the parameters, and the total number of Markov states. We label these unknowns as state, parameter, and model uncertainty, respectively. We assume agents are Bayesian, which means they update initial beliefs via Bayes' rule as data arrives. Later in the paper, we develop an extension to this model where agents can also learn from a vector of additional macro variables and consider the case of additional learning from GDP growth data.

[1] We do not consider, for instance, 1- or 4-state models as the Likelihood ratios of these relative to the 2- or 3-state model show that the 2- and 3-state models better describe the data. As we will show, however, there is some time-variation in whether a 2- or 3-state model matches the data better, which is one of the reasons we entertain both of these as alternative models.

The learning problem is as follows. We consider $k = 1, ..., K$ models, $\{\mathcal{M}_k\}_{k=1}^K$, and in model $\mathcal{M}_k$, the state variables and parameters are denoted as s_t and θ, respectively.[2] The distribution $p\left(\theta, s_t, \mathcal{M}_k | y^t\right)$ summarizes beliefs after observing data $y^t = (y_1, ...y_t)$. To understand the components of the learning problem, we can decompose the posterior as:

$$p\left(\theta, s_t, \mathcal{M}_k | y^t\right) = p\left(\theta, s_t | \mathcal{M}_k, y^t\right) p\left(\mathcal{M}_k | y^t\right). \tag{8.2}$$

$p\left(\theta, s_t | \mathcal{M}_k, y^t\right)$ solves the parameter and state "estimation" problem conditional on a model and $p\left(\mathcal{M}_k | y^t\right)$ provides model probabilities. It is important to note that this is a non-trivial, high-dimensional learning problem, as posterior beliefs depend in a complicated manner on past data and can vary substantially over time. The dimensionality of the posterior can be high, in our case more than 10 dimensions.

One of our primary goals is to characterize and understand the asset pricing implications of the transient process of learning about the parameters, states, and models.[3] Learning generates a form of nonstationarity, since parameter estimates and model probabilities are changing through the sample. When pricing assets, this can lead to large differences between *ex ante* beliefs and *ex post* outcomes, as shown in Cogley and Sargent (2008). Given this nonstationarity, we are concerned with understanding the implications of learning based on the specific experience of the U.S. post-war economy.[4]

To operationalize the learning problem, we need to specify the prior distribution, the data the agent uses to update beliefs, and develop an econometric method for sampling from the posterior distribution. In terms of data, we in a benchmark case assume that agents learn only from observing past and current consumption growth, a common assumption in

[2]This is a notational abuse. In general, the state and dimension of the parameter vector should depend on the model, thus we should superscript the parameters and states by 'k', θ^k and s_t^k. For notational simplicity, we drop the model dependence and denote the parameters and states as θ and s_t, respectively.

[3]These type of problems received quite a bit of theoretical attention early in the rational expectations paradigm - see for example Bray and Savin (1986) for a discussion of model specification and convergence to rational expectations equilibria by learning from observed outcomes.

[4]This is different from the standard practice of looking at population or average small-sample unconditional asset price and consumption growth moments from a model calibrated to the U.S. postwar data – we are looking at a single outcome corresponding to the U.S. post-war economy.

the learning literature (see, e.g., Cogley and Sargent (2008) and Hansen and Sargent (2009)). The primary data used is the 'standard' data set consisting of real, per capita quarterly consumption growth observations obtained from the Bureau of Economic Analysis (the National Income and Product Account tables) from 1947:$Q1$ until 2009:$Q1$.

8.3 Initial beliefs

The learning process begins with initial beliefs or the prior distribution. In terms of functional forms, we assume proper, conjugate prior distributions (Raiffa and Schlaifer (1956)). One alternative would be flat or 'uninformative' priors, but this is not possible in Markov switching models, as this creates identification issues (the label switching problem) and causes problems sampling from the posterior.[5] Conjugate priors imply that the functional form of beliefs is the same before and after sampling, are analytically tractable for econometric implementation, and are flexible enough to express a wide range initial beliefs.

For the mean and variance parameters in each state, $\left(\mu_i, \sigma_i^2\right)$, the conjugate prior is $p(\mu_i|\sigma_i^2)p(\sigma_i^2) \sim \mathcal{NIG}(a_i, A_i, b_i, B_i)$, where $\mathcal{NIG}$ is the normal/inverse gamma distribution. The transition probabilities are assumed to follow a Beta distribution in 2-state specification and its generalization, the Dirichlet distribution, in models with three states. Calibration of the hyperparameters completes the specification.

We endow our agent with economically motivated initial beliefs to study how learning proceeds from various starting points. We consider three prior distributions and use an 'objective' approach to calibrate the prior parameters. The first, the 'historical prior,' uses a training sample to calibrate the prior distribution. Training samples are the most common way of generating objective prior distributions (see, e.g., O'Hagan (1994)). In this case, an

[5]The label switching problem refers to the fact that the likelihood function is invariant to a relabeling of the components. For example, in a two-state model, it is possible to swap the definitions of the first and second states and the associated parameters without changing the value of the likelihood. The solution is to impose parameter constraints in optimization for MLE or to use informative prior distributions for Bayesian approaches. These constraints/information often take the form of an ordering of the means or variances of the parameters. For example in a two state model, it is common to impose that $\mu_1 < \mu_2$ and/or $\sigma_1 < \sigma_2$ to breaks the symmetry of the likelihood function.

initial data set is used to provide information on the location and scale of the parameters. In our application, we use the annual consumption data from Shiller from 1889 until 1946. Given the prior generated from the training sample, learning proceeds on the second data set – in our case, the post World War II sample.[6]

The second is called the 'look-ahead prior.' This prior sets the prior mean for each parameter equal to full-sample maximum likelihood estimates using the post World War II sample, similar to the procedure employed in an 'Empirical Bayes' approach. The prior variances are chosen to be relatively flat around these full-sample estimates, in order to allow for meaningful learning about the parameters as new data arrives, without running into label-switching identification problems. This approach violates the central idea of the Bayesian approach, as the prior contains information from the sample, but it is useful for analyzing the evolution of parameter uncertainty through the post World War II sample. The main differences between the historical and the look-ahead priors are that the historical priors have on average higher consumption growth volatility, shorter expansions, and longer recessions. For the 3-state model, the disaster state is also more severe in the historical prior, reflecting the Great Depression.

The third is called the 'fixed parameter' prior. This is a point-mass prior located at the end-of-sample estimates. In this case, the agent only learns about the latent Markov state. This prior mimics the typical rational expectations approach and allows us to separately identify the role of state and parameter learning, since the other priors have both state and parameter learning.

The details of the priors, the specific prior parameters chosen, as well as a description of the econometric technique we apply to solve this high-dimensional learning problem (particle filtering) are given in the Appendix.

---

[6]Romer (1989) presents evidence that a substantial fraction of the volatility of macro variables such as consumption growth pre-WW2 is due to measurement error. To alleviate this concern, we set the prior mean over the variance parameters to a quarter of the value estimated over the training sample. See the Appendix for further details.

Chapter 9

Time-series of subjective beliefs

This section characterizes the learning process. We first discuss state, parameter, and model learning and their implications for the time series of conditional consumption moments, as perceived by the Bayesian agent. Next, we empirically investigate how revisions in the agent's beliefs are related to stock market returns. We also consider the case of learning from GDP data, in addition to consumption data. In the following section, we embed these beliefs in a general equilibrium model and discuss the asset pricing implications in more detail.

9.1 State and parameter learning

Conditional on a model specification, our agent learns about the Markov state and the parameters, with revisions in beliefs generated by a combination of data, model specification, and initial beliefs. To start, consider the agent's beliefs about the current state of the economy, s_t, where state 1 is an 'expansion' state, state 2 the 'contraction' state and, if a 3-state model, state 3 the 'disaster' state. Estimates are given by

$$E\left[s_t|\mathcal{M}_k, y^t\right] = \int s_t p\left(\theta, s_t|\mathcal{M}_k, y^t\right) d\theta ds_t.$$

Note that these are marginal *mean* state beliefs, as parameter uncertainty is integrated out. Although s_t is discrete, the mean estimates need not be integer valued. Figure 9.1 displays the posterior state beliefs over time, for each model and for different priors.

There are a number of notable features of these beliefs. NBER recessions (shaded yellow) and expansions are clearly identified in the models. The only exceptions are the recessions in the late 1960s and 2001, which were not associated with substantial consumption declines. Comparing the panels, one area in which the models generate strong differences is persistence of the states. The i.i.d. model identifies recessions as a one-off negative shock, but since shocks are i.i.d., the agent does not forecast that the recession state will persist with high likelihood. In contrast, the 2- and 3-state models clearly show the persistence of the recession states. Disaster states are rare – after the initial transient post war period, there are only really two observations that place even modest probability on the disaster state – the recession in 1981 and the financial crisis at the end of 2008. This implies that disaster states are nearly 'Peso' events in the post WW2 sample.

The agent's beliefs are quite volatile early in the sample in all of the models. This is not surprising. Since initial parameter beliefs are highly uncertain, the agent has a difficult time discerning the current state as parameter uncertainty exacerbates state uncertainty. As the agent learns, parameter uncertainty decreases and state identification is easier. It is important to note that even with full knowledge of the parameters, the agent will never be able to perfectly identify the state.[1] The results also show that the priors do not have a large impact on the mean state beliefs, at least for the unrestricted 2- and 3-state models, as the posterior beliefs are roughly similar for the historical and look-ahead priors.

Next, consider beliefs over parameters. Due to the large number of parameters and in the interests of parsimony, we focus on a few of the more economically interesting and important parameters. For the 2-state models, the top panels of Figure 9.2 display posterior means of the beliefs over σ_1 and σ_2. Notice that for the Historical prior the conditional volatilities slowly decrease, after a short (about 5 year) burn-in period, essentially throughout the sample. This is a combination of the Great Moderation (realized consumption volatility did decrease over the post-war sample) and the initial beliefs, which based on the historical experience expected higher consumption growth volatility. Interestingly, for the look-ahead

[1]The posterior variance of the state, $var\left[s_t|\mathcal{M}_k, y^t\right]$, does decline over time due to decreasing parameter uncertainty. This will be discussed further when we use GDP growth as an additional observation to help identify the state.

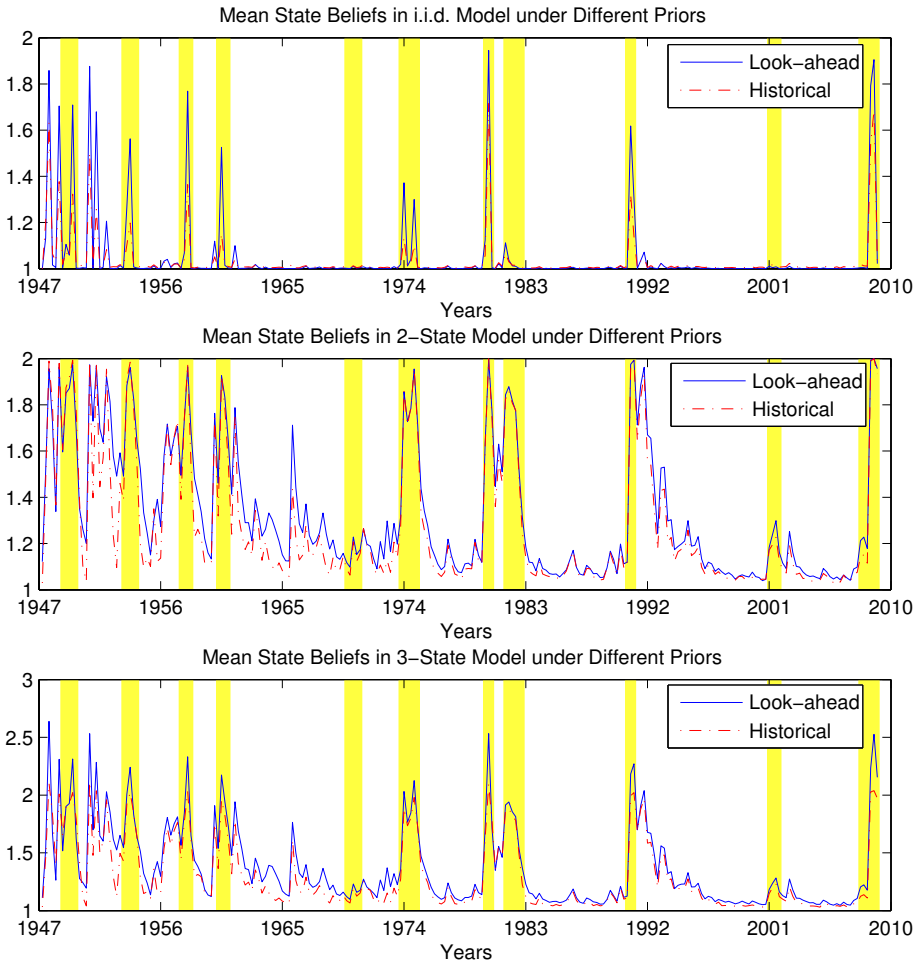

Figure 9.1: Evolution of Posterior Mean State Beliefs

prior, which is centered at the end of sample posterior values, the agents quickly unlearns the low full sample consumption growth volatility, and after about 5-year burn-in, the volatility is close to that observed for the historical prior. This occur because volatility was higher in the first portion of the sample. The subsequent decline in the volatility in the good state is quantitatively large (about a 30% drop).

The lower panels in Figure 9.2 display the transition probabilities, π_{11} and π_{22}. After the burn-in period, the first is essentially increasing over the sample, while the latter is decreasing. That is, 50 years of, on average, long expansions and high consumption growth leads to revisions in beliefs that are manifested in higher probabilities of staying in the good state and lower probabilities of staying in recession state. The probability of staying in a recession, conditional on being in a recession, goes down from about 0.85 to 0.75. Clearly, such positive shocks to the agents' perception of the data generating process will lead to higher ex post equity returns than compared to ex ante expectations.

The first three panels of Figure 9.3 displays estimates of the mean parameters, $E\left[\mu_i|\mathcal{M}_k, y^t\right]$ for $i = 1, 2, 3$, as well as a posterior two standard deviation band for the 3-state model using the historical prior. Learning is most apparent in the good state and least apparent in the disaster state. This is intuitive, since the economy spends most of its time in the good state and little, if any, time in the disaster state. This provides empirical evidence supporting the argument that a high level of parameter uncertainty is a likely feature of a model with a rarely observed state and is an important feature for disaster risk models (see also Chen, Joslin, and Tran, 2010).

The fourth, lower right panel shows how the speed of learning differs in the three models we consider. We use the conditional variance over the infinite horizon mean of quarterly consumption growth, $Var\left(E\left[\Delta c_{t+\infty}\right]|y^t\right)$, as a measure of the amount of parameter uncertainty (with no parameter uncertainty, the long-run mean of consumption growth is constant in all models), and show this variance for the unrestricted 2- and 3-state models normalized by the variance from the simpler i.i.d. model. The plot shows that learning happens faster in the simpler i.i.d. model in that both the variance ratios quickly increases. The unrestricted 2-state model settles at a variance about 50% higher than for the i.i.d. 2-state model, while the 3-state model increases its relative amount of parameter uncertainty

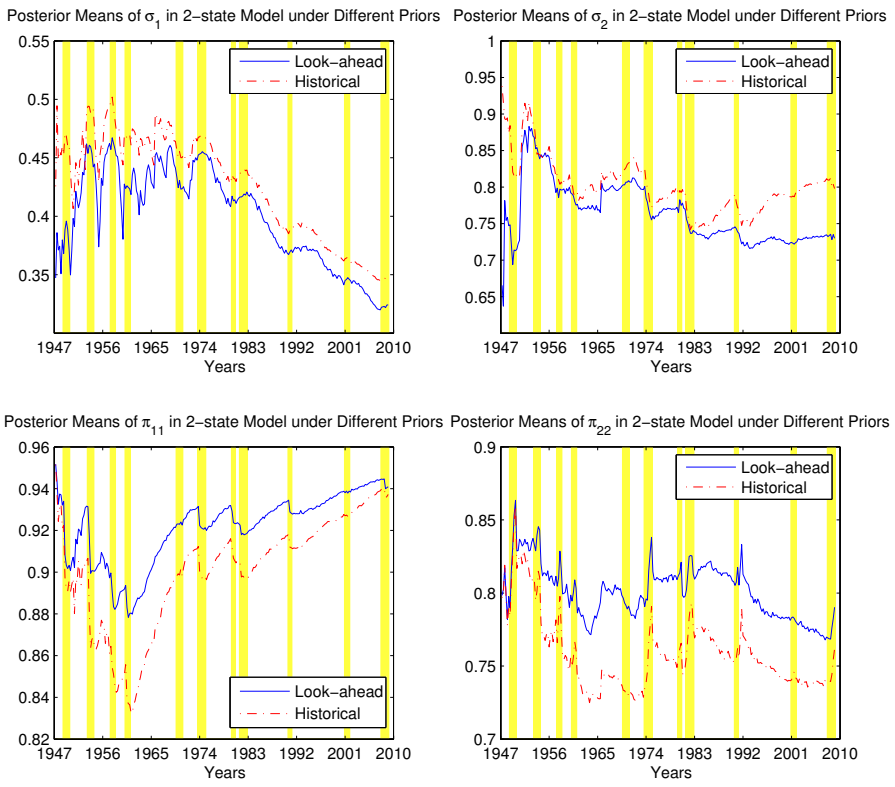

Figure 9.2: Mean Parameter Beliefs of the Volatility and Transition Probabilities

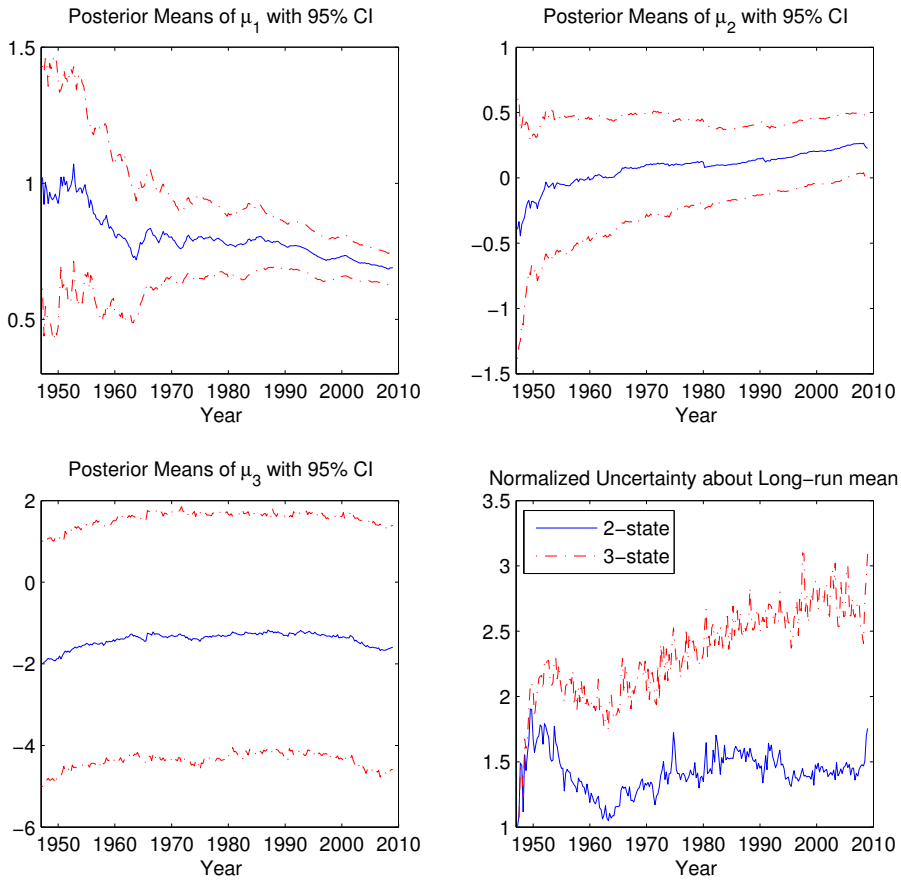

Figure 9.3: Speed of Learning

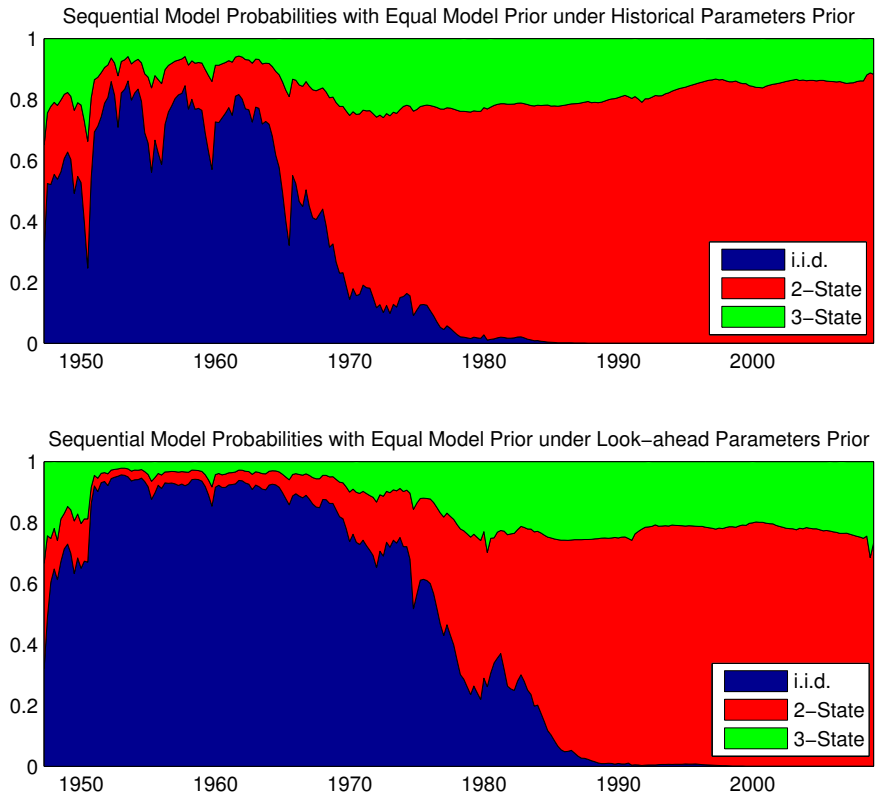

Figure 9.4: Marginal Model Probabilities under Different Priors

to about 3 times that of the i.i.d. model at the end of the sample. This is due to the very slow learning about the disaster state and the difficulty present in learning the transition probabilities.

There is additional interesting time-variation in beliefs about the parameters, but this time-variation is best summarized via the total impact across all parameters, which is measured via predictive moments and discussed in the next section.

9.1.1 Beliefs about models and consumption dynamics

Figure 9.4 shows the marginal model probabilities, $p\left(\mathcal{M}_k|y^t\right)$, for each of the models we consider for the Historical and the Look-ahead priors, respectively.[2] For simplicity, the prior probability of each model was set to 1/3. Note first that the posterior probability of the i.i.d. model decreases towards zero for both priors. Thus, i.i.d. consumption growth is rejected by a Bayesian agent that updates by observing past realized consumption growth. Although not reported for brevity, this conclusion is robust even if the prior probability of the i.i.d. model is set to 0.95 - in this case it takes somewhat longer (but still just a little over half the sample) for the probability of the i.i.d. model to drop very close to zero. The 3-state model also sees a reduction in its likelihood and ends at about 10% and 20% probability levels at the end of the sample for the Historical and Look-ahead priors, respectively. The Look-ahead prior has a less severe disaster state, as it does not reflect the Great Depression, and this is why the probability of the 3-state model is higher in this case. As mentioned in the introduction, a single large negative consumption shock would quickly change these probabilities. In sum, we observe large changes in the model uncertainty over the sample.

The fact that the agent can learn that consumption growth is not i.i.d. is important. Many asset pricing models specify i.i.d. consumption growth with the implicit assumption that it is not possible or difficult to detect non-i.i.d. dynamics in consumption. Our results show that agents, using only consumption growth data, can detect non-i.i.d. dynamics, and can do so in real time, which is an even stronger result. The agent does not need to wait until the end of the sample. This result holds for various prior specifications and is robust to time-aggregation.[3]

[2]Note that marginal model probabilities (i.e., where parameter uncertainty is integrated out) penalizes extra parameters as more sources of parameter uncertainty tends to flatten the likelihood function. Thus, it is not the case, as we see an example of here, that a 3-state model always dominates a 2-state model in Bayesian model selection.

[3]In the Appendix, we show that taking out an autocorrelation of 0.25 from the consumption growth data, which is what time-aggregation of i.i.d. data predicts (see Working (1960)), does not qualitatively change these results - if anything it makes the rejection of the i.i.d. model occur sooner. The same is true if we

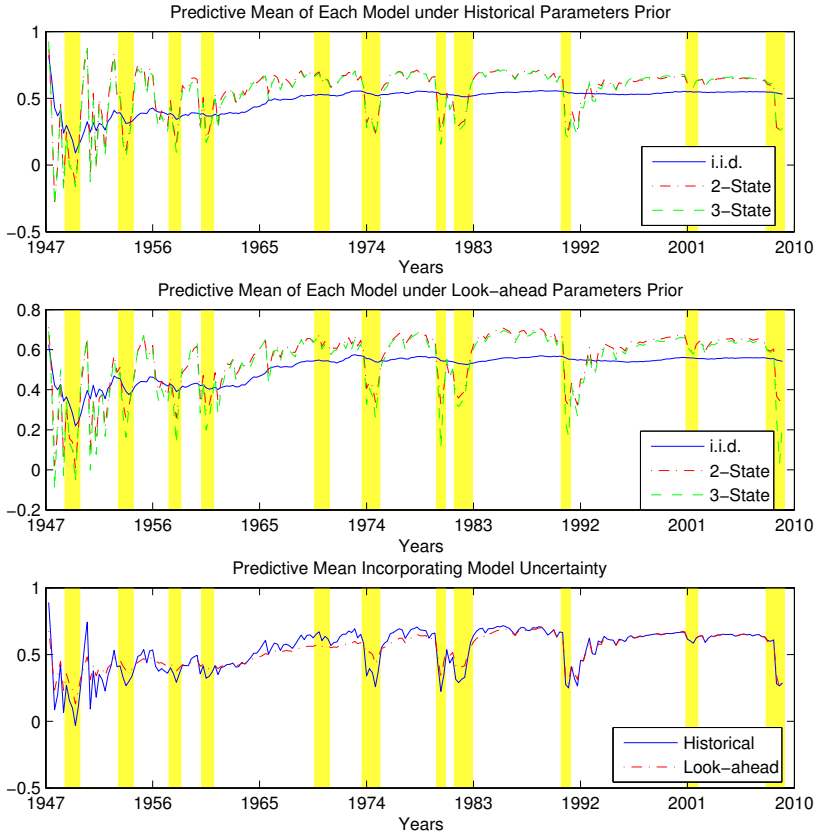

Figure 9.5: Quarterly Expected Consumption Growth

The results of the previous section indicate that beliefs about the parameters vary through the sample, even for the look-ahead prior, but it is not clear from this how much variation in conditional moments is present.[4] To provide asset-pricing relevant measures, we report the agent's beliefs regarding the first four moments of conditional consumption growth and model probabilities. All of these quantities are marginal, integrating out parameter, state, and/or model uncertainty. For example, the predictive mean for a given model, $\mathcal{M}_k$, is

$$E\left[\Delta c_{t+1}|\mathcal{M}_k, y^t\right] = \int \Delta c_{t+1} p\left(\Delta c_{t+1}|\theta, s_t, \mathcal{M}_k, y^t\right) p\left(\theta, s_t|\mathcal{M}_k, y^t\right) d\theta ds_t.$$

In describing these moments, we generally abstract from the first ten years and treat it is a 'burn-in' period, in order to allow the prior some time to adjust to the data, as there is some transient volatility over these first few years.

The top two panels in Figure 9.5 (for historical and look-ahead priors, respectively) display the conditional expected quarterly consumption growth for each model. The two and 3-state models generate relatively modest differences in this moment – both pick up business cycle fluctuations in expected consumption growth, with the 3-state model identifying the recessions in the early 80's and the financial crisis in '08 as severe. Persistent recessions are missing from the i.i.d. model, as expected. All three models exhibit a low frequency increase in expected consumption growth over the first half of the sample, due to parameter learning.

The bottom panel of Figures 9.5 shows *model averaged* expected quarterly consumption growth for the two priors. In the first third of the sample, the presence of the i.i.d. model smooths business cycle fluctuations in expected consumption growth. Thereafter, only the 2- and 3-state models are relevant and model uncertainty has a minor impact as the conditional expected growth is similar in these models. Overall, recessions are associated with a mean quarterly consumption growth of about 0.3%, while the mean consumption

purge the data of its full sample first order autocorrelation.

[4] As an example, consider the conditional volatility of consumption growth. A decrease in the probability of the bad state, which has higher consumption growth volatility, could be offset by an increase in the consumption volatility in the good state, σ_1, keeping the total conditional volatility of consumption growth constant.

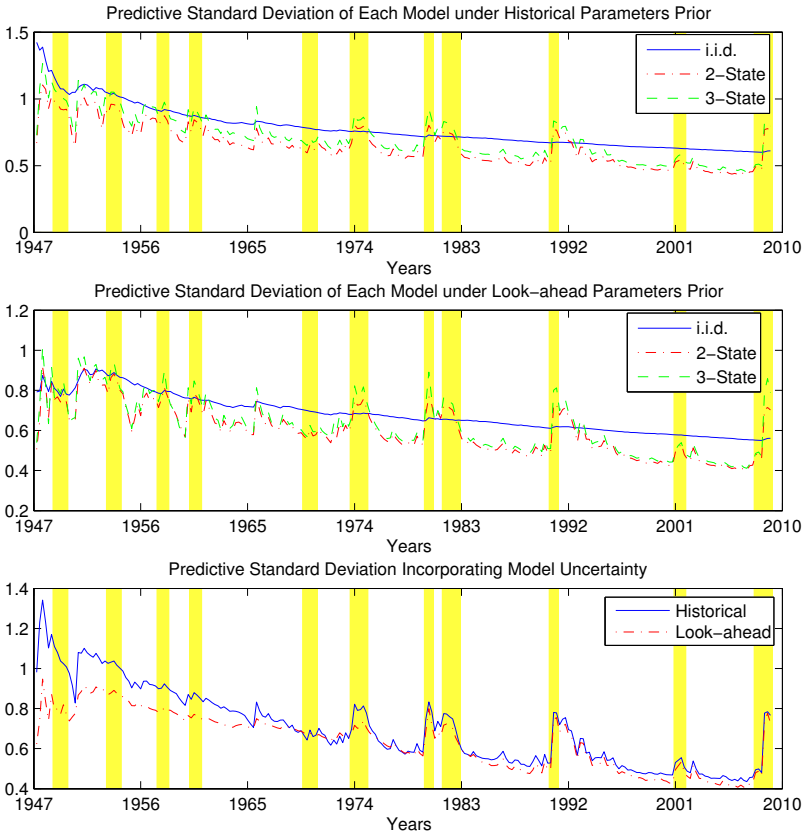

Figure 9.6: Quarterly Predictive Consumption Growth Standard Deviation

growth in expansions is about 0.6%. Since business cycles are relatively persistent, these fluctuations in conditional consumption growth are a source of long-run consumption risk, akin to that of Bansal and Yaron (2004). However, the lower frequency fluctuations we observe in expected consumption growth, which is due to parameter learning, constitute "truly" long-run risk, as shocks to parameter beliefs are permanent.

Turning to the conditional volatility of quarterly consumption growth, Figure 9.6 shows that for both priors there is a downward trend in consumption growth volatility through the sample, with marked increases during recessions for the non-i.i.d. models. Again, the bottom panel shows the belief about conditional standard deviation for each prior when

model uncertainty is integrated out. Model probabilities could be driven by unexpected volatility, but this does not appear to be a primary determinant. Conditional consumption growth volatility is not particularly affected by model uncertainty, since both the two and the 3-state models have similar volatility patterns, and since the i.i.d. model is essentially phased out in the first third of the sample.

The secular decline is largely driven by downward revisions in estimates of the volatility parameters as realized consumption growth was less volatile in the second half of this century. This is particularly strong for the historical prior, as the conditional volatility of consumption growth decreases from about 1% per quarter to about 0.5%. Interestingly, the look-ahead prior has a similar trend, after a short burn-in period, as the prior's low consumption growth volatility is quickly unlearned, though the size of the effect is about half as large. This is the Great Moderation - the fact that consumption volatility has decreased also over the post-war sample. In the models considered here, the agent learning in real-time perceives this decrease to happen gradually, in contrast to studies that find *ex post* evidence of structural breaks or regime shifts at certain dates.

Every recession is associated with higher consumption growth volatility, although the size of the increase varies. The largest increase, on a percentage basis, occurs with the financial crisis of 2008. The increase is largest in the 3-state model, as the mean state belief at this time approaches the third state, which has a very high volatility. There is little updating about the volatility of the disaster state through the sample, since there have been no prolonged visits to this state. Thus, this reflects the fear that prevailed in the fall of 2008 that the economy was potentially headed into a depression not seen since the 1930s. This econometric result squares nicely with anecdotes from the crisis.

Figure 9.7 shows the time-series of conditional consumption growth skewness for the both priors, again with the model averaged estimates in the bottom panel. The time-variation in the conditional skewness is dominated by business cycle variation for the two and 3-state models, and there is a slight downward trend, as the probability of a disaster and recession decrease. When the economy is in a recession, consumption growth is naturally less negatively skewed for two reasons: (1) there is a high probability that the economy

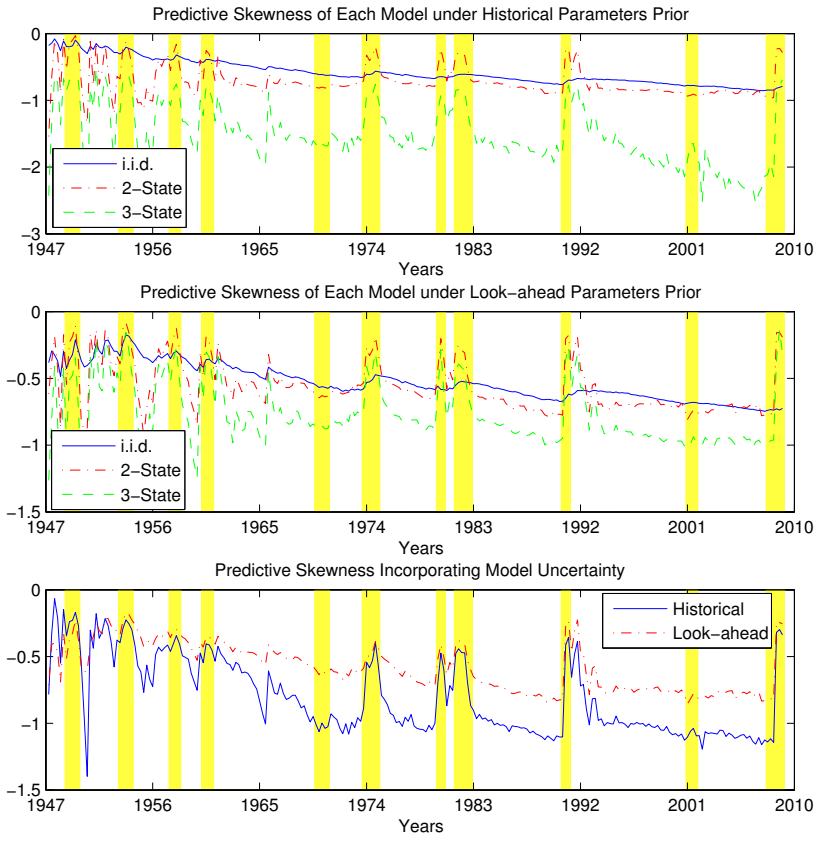

Figure 9.7: Quarterly Predictive Consumption Growth Skewness

jumps to a higher (i.e. better) state and (2) expected consumption volatility is high, which tends to decrease skewness. Note that in terms of skewness, the 3-state model, with its severe recession (disaster) state, is quite different from the 2-state model. Thus model uncertainty plays a larger role for the agent's overall consumption beliefs in terms of the skewness. The 3-state model, especially for the Historical prior, strongly impacts the total perception of conditional consumption growth skewness as given in the bottom panel.

Figure 9.8 shows the time-series of conditional consumption growth kurtosis for the both priors. Conditional kurtosis is lower in bad states as these states are the least persistent and volatility is highest. Large, rare, outcomes are more likely when the economy is in the good state. This has potentially interesting option pricing implications (see, e.g., Backus, Chernov, and Martin (2009)), as the skewness and kurtosis will be related to volatility smiles. It is worth noting that parameter uncertainty gives an extra 'kick' to conditional skewness and kurtosis measures relative to the case of fixed parameters, where the skewness and kurtosis both move little over time (the fixed parameter case is not reported here for brevity). Both for skewness and kurtosis, there is clear evidence of parameter learning over the business cycle: the skewness becomes more negative and the kurtosis higher the longer an expansion last, reflecting updating of the transition probabilities, which reflect business cycle dynamics. Similar to skewness, there are now relatively large differences between the 2- and 3-state models. The 3-state model has significantly higher conditional kurtosis than the 2-state model, due to the presence of the disaster-state. Interestingly, the differences are greater in expansions than in recessions, again due to the 'rare' nature of recessions and, especially, disasters. In terms of the conditional kurtosis after model uncertainty is integrated out (bottom panel), the 3-state model has large impact on kurtosis even at the end of the sample where the probability of this model being the right model is low. Thus, among the models considered here, model uncertainty and its dynamic behavior is likely to have the strongest implications for assets such as out-of-the-money options that are more sensitive to the tail behavior of consumption growth.

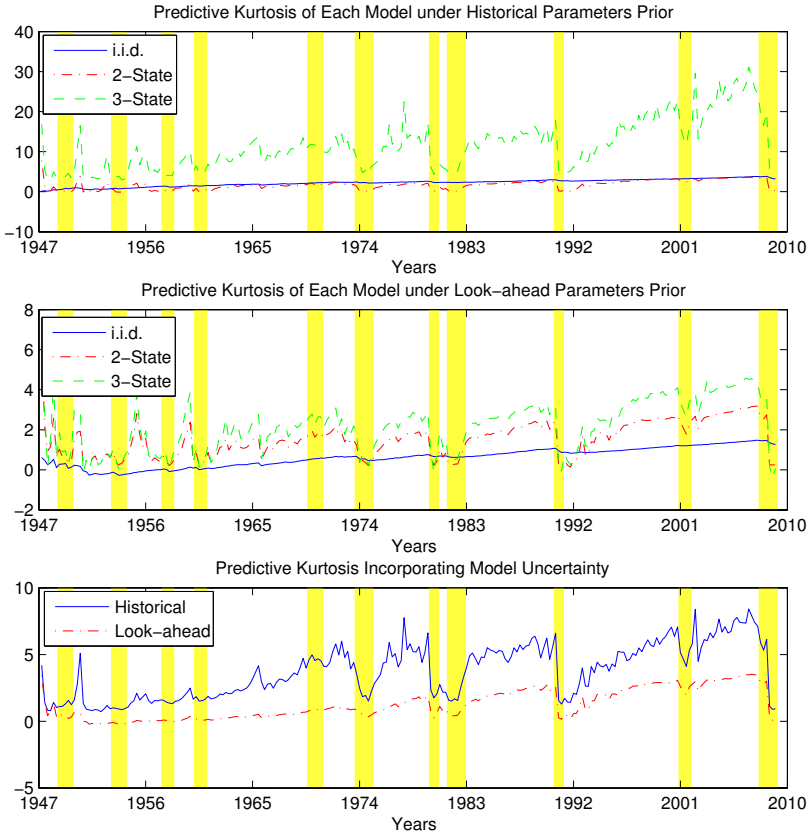

Figure 9.8: Quarterly Predictive Consumption Growth Kurtosis

Chapter 10

Does learning matter for asset prices?

10.1 A new test for the importance of learning

The previous results indicate that the agent's beliefs – about parameters, moments, and models – vary substantially at both very low frequencies and over the business cycle. If learning is an important determinant of asset prices, changes in beliefs should be a significant determinant of asset returns. This is a fundamental test of the importance of learning about the consumption dynamics. For example, if agents learn that expected consumption growth is higher than previously thought, this revision in beliefs will be reflected in the aggregate wealth-consumption ratio (if the elasticity of intertemporal substitution is different from one). In particular, if the substitution effect dominates, the wealth-consumption ratio will increase when agents revise their beliefs about the expected consumption growth rate upwards (see, e.g., Bansal and Yaron (2004)). As another example, if agents learn that aggregate risk (consumption growth volatility) is lower than previously thought, this will generally lead to a change in asset prices as both the risk premium and the risk-free rate are affected. In the Bansal and Yaron (2004) model, an increase in the aggregate volatility leads to a decrease in the stock market's price-dividend ratio.

To test this, we regress excess quarterly stock market returns (obtained from Kenneth French's web site) on changes in beliefs about expected consumption growth and expected

consumption growth variance. This is a particularly stringent test of learning, which to our knowledge has not been done in the previous literature. We use the beginning of period timing for the consumption data here and elsewhere in the paper.[1] The regressors are the shocks, $E_t(\Delta c_{t+1}) - E_{t-1}(\Delta c_{t+1})$ and $\sigma_t(\Delta c_{t+1}) - \sigma_{t-1}(\Delta c_{t+1})$. Notice that the only thing that is changing is the conditioning information set as we go from time $t-1$ to time t; the regressors are revisions in beliefs. We calculate these conditional moments for each prior integrating out state, model and parameter uncertainty. The first 10 years of the sample are used as a burn-in period to alleviate any prior misspecification (there is some excess volatility in state and parameter beliefs in these first years).

Separate regressions are run for the historical and look-ahead priors, and we control for contemporaneous consumption growth and lagged consumption growth (the direct cash flow effect). By controlling for realized consumption growth, we ensure that the results are driven by model-based revisions in beliefs, and not just the fact that realized consumption growth (a direct cash flow effect) was, for example, unexpectedly high. To separate out the effects of parameter from state learning, we use revisions in expected consumption growth beliefs computed from the 3-state model with fixed parameters (set to their full-sample values) as an additional control.[2]

Specifications 1 and 2 in Panel A (historical prior) and Panel B (look-ahead prior) in Table 10.1 show that increases in expected conditional consumption growth are positively and strongly significantly associated with excess contemporaneous stock returns for both priors. This result holds controlling for contemporaneous and lagged consumption growth (the direct cash flow effect), and so we can conclude that revisions in beliefs are significantly related to shocks to the price-dividend ratio. This is a very strong result, pointing to the importance of a learning-based explanation for realized stock returns. These results could

[1] Due to time-averaging (see Working, 1960), Campbell (1999) notes that one can use either beginning of period or end of period consumption in a given quarter as the consumption for that quarter. The beginning of period timing yields stronger results than using the end of period convention (although the signs are the same in the regressions). In principle, the results should be the same, so this is consistent with some information being impounded in stocks before the consumption data is revealed to the Bureau of Economic Analysis.

[2] Using the fixed parameter 2-state model as the control instead does not change the results.

Table 10.1: Updates in Beliefs versus Realized Stock Returns

Dependent variable: $r_{m,t+1} - r_{f,t+1}$ (excess market returns)

Panel A: Historical prior	1	2	3	4	5	6	7
$E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]$	40.43^{***}	26.00^{**}		42.41^{**}			
	(9.36)	(11.44)		(18.97)			
$\sigma^2_{t+1}[\Delta c_{t+2}] - \sigma^2_t[\Delta c_{t+2}]$						-36.34^{***}	-13.83
						(10.90)	(10.14)
Controls:							
Δc_{t+1}		2.02			7.94^{**}		3.76^{***}
		(1.51)			(3.12)		(1.37)
Δc_t		2.31^{*}			1.73		2.05
		(1.41)			(1.43)		(1.43)
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{3-state model}}_{\theta \text{ known}}$			24.98^{***}	-1.76			
			(9.36)	(12.82)			
					-449.17		
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{i.i.d. model}}_{\theta \text{ unknown}}$					(392.94)		
R^2_{adj}	8.8%	10.9%	5.9%	8.4%	9.5%	5.0%	9.7%

Panel B: Look-ahead prior	1	2	3	4	5	6	7
$E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]$	52.44^{***}	29.05^{**}		39.63^{**}			
	(11.71)	(13.62)		(18.42)			
$\sigma^2_{t+1}[\Delta c_{t+2}] - \sigma^2_t[\Delta c_{t+2}]$						-46.10^{***}	-23.26^{*}
						(12.53)	(13.43)
Controls:							
Δc_{t+1}		2.77^{*}			7.78^{**}		3.51^{**}
		(1.56)			(3.18)		(1.42)
Δc_t		2.25^{*}			1.74		2.16
		(1.38)			(1.43)		(1.39)
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{3-state model}}_{\theta \text{ known}}$			24.98^{***}	8.77			
			(9.36)	(10.62)			
					-427.05		
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{i.i.d. model}}_{\theta \text{ unknown}}$					(400.75)		
R^2_{adj}	7.3%	10.5%	5.9%	7.2%	9.5%	5.3%	10.2%

be driven by parameter or state learning.

Specification 3 shows that the updates in expected consumption growth derived from the model with fixed parameters (that is, a case with state learning only) are also significantly related to realized stock returns. The R^2, however, is lower than for the case of the full learning model, and when we include the revisions in beliefs about expected consumption growth from both the full learning model and the fixed parameters benchmark model in the regression (specification 4), the updates in expected consumption growth that arise in a model with fixed parameters are insignificant, while the belief revisions from the full learning model remain significant. That is, updates in expectations when learning about about parameters, states, and models are more closely related to realized stock market

returns than the corresponding updates in expectations based on a single model with known parameters but hidden states estimated on the full sample. To our knowledge, this is the first direct comparison of learning about models and parameters versus the traditional implementation of the rational expectation explanations in terms of explaining the time-series of realized stock returns using the actual sequence of realized macro shocks.

This result is driven by the nonlinear process of jointly learning about parameters and states. In particular, specification 5 shows that updates in beliefs from the i.i.d. model cannot be distinguished from the direct cash flow effect. The i.i.d. model captures parameter uncertainty about the long-run mean and variance, but not the state dynamics. The fixed parameter model (specification 4) captures the transitory state learning, but not the parameter dynamics.[3] Thus, it is the updates in beliefs stemming from the more complicated, non-i.i.d. models' learning problem that drives the increased correlation with stock returns, relative to the direct cash flow effect. Recall also that our agent quickly learned that the i.i.d. model is not likely, relative to the other specifications.

For the variance (regression specifications 6 and 7 in Table 10.1) we get the opposite result, as one would expect (at least with a high elasticity of intertemporal substitution, as we will use later in the paper): unexpected increases in conditional consumption growth variance are associated with negative contemporaneous stock returns. This result is not significant at the 5% level when including contemporaneous and lagged consumption growth in the regressions (specification 7). This does not mean there is no effect; we just cannot distinguish it from the direct cash flow effect when learning from consumption data alone.

To summarize, we find strong evidence that the updates in beliefs elicited from our model/prior combinations are associated with actual updates in agent beliefs at the time, as proxied by stock market returns. Again, it is important to recall that no asset price data was used to generate these belief revisions.

[3]One can show analytically that in a simple i.i.d. model, updates in expectations of consumption growth are very close to linear in the realized consumption growth.

10.2 Learning from additional macro variables

Agents have access to more than just aggregate consumption growth data when forming beliefs. Here we provide one approach for incorporating this additional information and apply this methodology to learning from quarterly GDP growth, in addition to consumption. Suppose x_t represents the common growth factor in the economy and evolves via:

$$x_t = \mu_{s_t} + \sigma_{s_t}\varepsilon_t, \tag{10.1}$$

where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, and s_t is the state of the economy, which follows the same Markov chains specified earlier. Consumption growth Δc and J additional variables $Y_t = [y_t^1, y_t^2, ..., y_t^J]'$ are assumed to follow:

$$\Delta c_t = x_t + \sigma_c \varepsilon_t^c, \tag{10.2}$$

where

$$y_t^j = \alpha_j + \beta_j x_t + \sigma_j \varepsilon_t^j, \quad for \quad j = 1, 2, ..., J \tag{10.3}$$

and $\varepsilon_t^c \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, and $\varepsilon_t^j \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ for any j. Note that the coefficients in equation (10.3) are not state dependent, which implies that the additional variables will primarily aid in state identification. The specification allows for the additional observation variables to be stronger or weaker signals of the underlying state of the economy than consumption growth. For the case of GDP growth, this captures the idea that investment is more cyclical than consumption, which makes GDP growth a better business cycle indicator. The linearity of the relationship is an assumption that is needed for conjugate priors.

The similar conjugate priors for the parameters are applied. For each state $s_t = i$, $p(\mu_i|\sigma_i^2)p(\sigma_i^2) \sim \mathcal{NIG}(a_i, A_i, b_i, B_i)$, where $\mathcal{NIG}$ is the normal/inverse gamma distribution. σ_c is assumed to follow an inverse gamma distribution $\mathcal{IG}(b_c, B_c)$, and for each $j = 1, 2, ..., J$, $p([\alpha_j, \beta_j]'|\sigma_j^2)p(\sigma_j^2) \sim \mathcal{NIG}(a_j, A_j, b_j, B_j)$, where $p([\alpha_j, \beta_j]'|\sigma_j^2)$ is a bivariate normal distribution $\mathcal{N}(a_j, A_j\sigma_j^2)$, a_j is a 2×1 vector and A_j is a 2×2 matrix. Particle filtering is straightforward to implement in this specification by modifying the algorithm described in the Appendix.

To analyze the implications of additional information, we consider learning using real, per capita U.S. GDP growth as an additional source of information. This exercise generates a battery of results: time series of parameter beliefs, conditional moments, and model probabilities. We report only a few interesting statistics in the interests of parsimony. Figure 10.1 shows that the state beliefs do not change dramatically, although GDP growth is typically thought of as more informative about business cycle fluctuations than consumption growth. To characterize how the additional data aids in state identification, we compute posterior standard deviations for the states, $std\left[s_t|\mathcal{M}_k, y^t\right]$, again integrating out parameter uncertainty. The top Panel of Figure 10.2, shows that indeed the uncertainty about the state is much lower (about half) than what was the case when using consumption growth as the only source of information. Thus, adding GDP growth to the agent's information set increases the precision of the state identification.[4] The increased certainty about the state improves parameter identification also, which is confirmed in the two lower Panels in Figure 10.2. Here the uncertainty about the good and bad states mean consumption growth rates is lower, after a 10-year burn-in, than in the case using consumption as the only source of information.

Figure 10.3 shows that the model specification results are similar, as the data again favors the 2-state model, leaving the 3-state model with a very low probability at the end of the sample. It is noteworthy, however, that the probability of the 3-state (disaster) model again increases at the onset of the financial crisis in 2008.

Adding GDP growth also results in a greater difference in expected consumption growth across the states. Figure 10.4 shows that the difference in the expected consumption growth rate in recessions versus expansions is about 0.6% per quarter, versus about 0.3% in the case of consumption information only (see Figure 9.5). The dynamic behavior of the conditional standard deviation of consumption growth is not significantly changed (not reported for brevity).

---

[4]It is technically feasible to impose cointegration between consumption and GDP by including the log consumption to GDP ratio on the right hand side of Equation (10.3). We thank Lars Hansen for pointing this out.

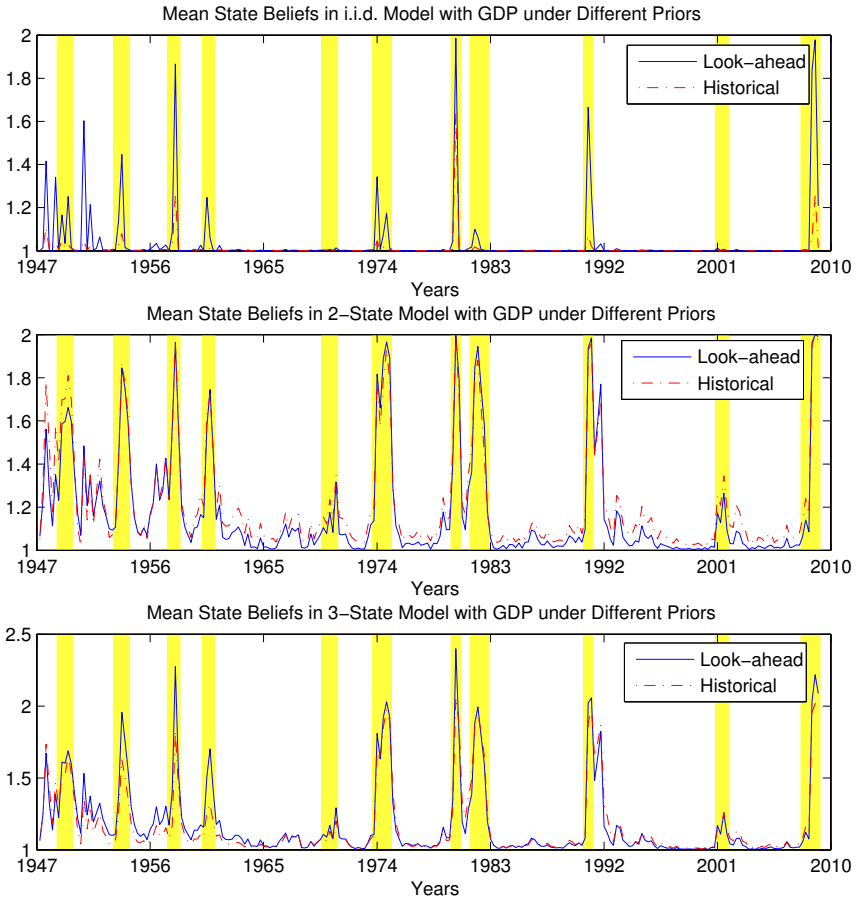

Figure 10.1: Evolution of Mean State Beliefs with GDP

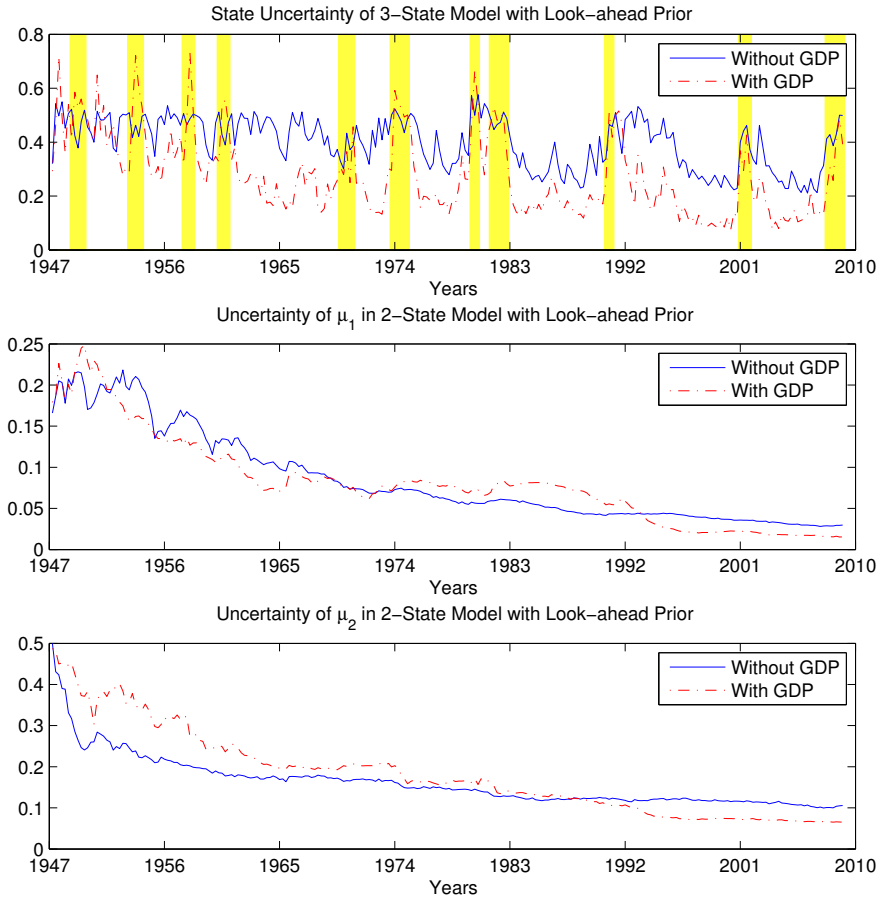

Figure 10.2: Uncertainty about state identification with/without GDP

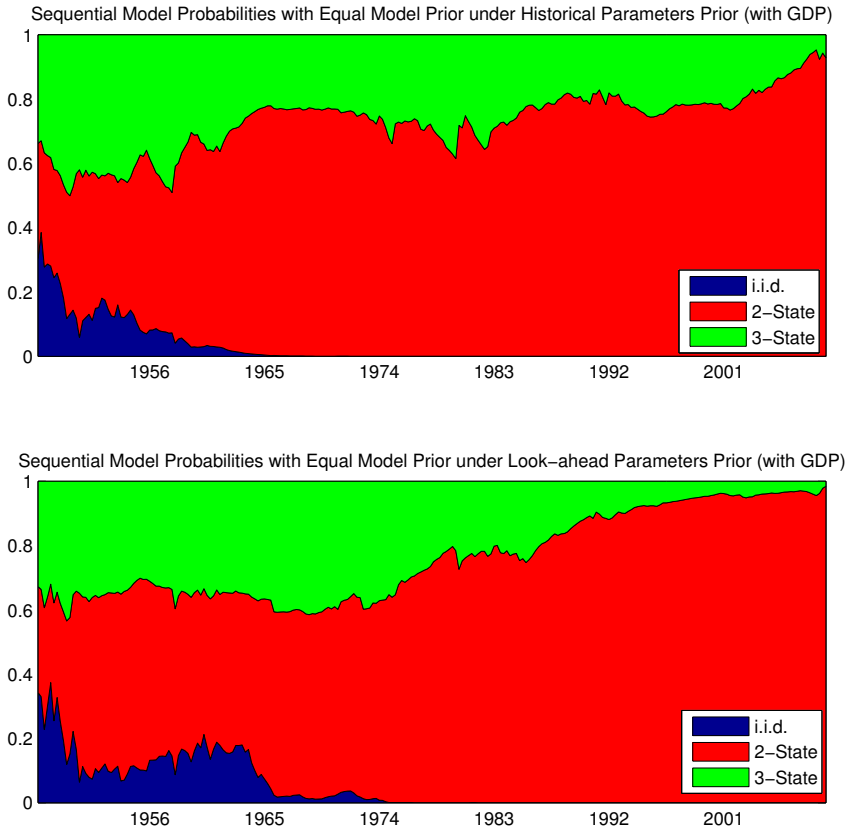

Figure 10.3: Marginal Model Probabilities with GDP.

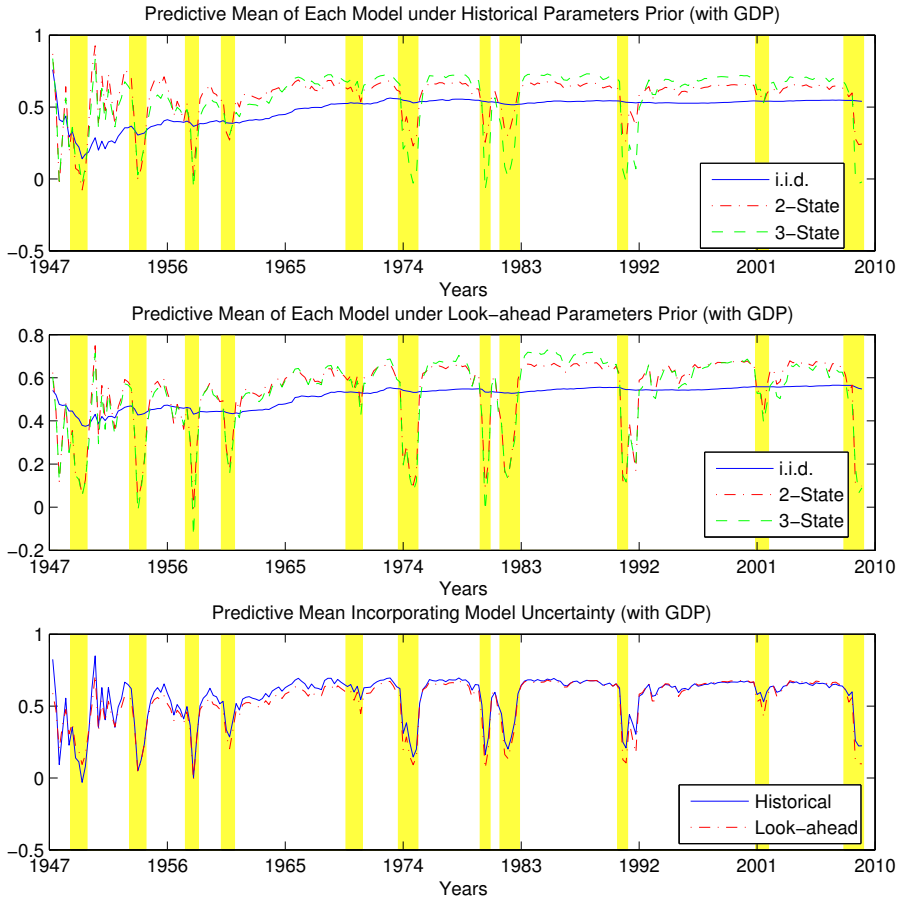

Figure 10.4: Conditional Expected Consumption Growth with GDP

Table 10.2 shows the regressions of contemporaneous stock returns and updates in agent beliefs about conditional expected consumption growth and consumption growth variance, as calculated from this extended model. The results are similar, but in fact overall *stronger* than the results using only consumption growth. Updates in agent expectations about these moments from the full learning model are significantly related to stock returns, also after controlling for contemporaneous and lagged consumption growth and updates in expected consumption growth derived from a model with fixed parameters. Again, this evidence indicates that learning about parameters and models is an important feature of the data.

10.3 Additional asset pricing implications

We now embed the beliefs of our learning agent in a general equilibrium asset pricing model. There are considerable computational and technical issues that need to be dealt with when considering such an exercise. First, the state space is prohibitively large. The 3-state model, as an example, have 12 parameters governing the exogenous consumption process, and the beliefs over each parameter are governed by 2 hyper-parameters. Thus, there are 24 state variables, in addition to beliefs over the state of the economy and the corresponding parameter and state beliefs for the i.i.d. and the general 2-state models. Second, as pointed out by Geweke (2001) and Weitzmann (2007), some parameter distributions must be truncated in order for utility to be finite. This introduces additional nuisance parameters.

Given the computational impediments, we follow Sargent and Cogley (2008) and Piazzesi and Schneider (2010) and apply the principle of "anticipated utility" to the pricing exercise (originally suggested by Kreps (1998)). Under this assumption, the agents maximize utility at each point in time assuming that the parameters and model probabilities are equal to the agents' current mean beliefs and will remain constant forever. Of course, at time $t+1$ the mean parameter beliefs will in general be different due to learning. While parameter and model uncertainty are not *priced* risk factors in this framework, they are nonetheless important for the time-series of asset prices as updates in mean parameter and model beliefs lead to changes in prices. We do integrate out state uncertainty in the pricing exercise, so state uncertainty is a priced risk factor (as in, e.g., Lettau, Ludvigson, and Wachter (2008)).

Table 10.2: Updates in Beliefs versus Realized Stock Returns with GDP

Dependent variable: $r_{m,t+1} - r_{f,t+1}$ (excess market returns)

Panel A: Historical Prior	1	2	3	4	5
$E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]$	40.68^{***}	40.52^{***}	39.77^{**}		
	(6.62)	(8.99)	(19.13)		
$\sigma^2_{t+1}[\Delta c_{t+2}] - \sigma^2_t[\Delta c_{t+2}]$				-56.94^{***}	-46.25^{***}
				(10.79)	(22.37)
Controls:					
Δc_{t+1}		-0.70			1.02
		(1.56)			(1.52)
Δc_t		1.93			2.36^{*}
		(1.33)			(1.41)
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{3-state model}}_{\theta \text{ known}}$			0.60		
			(10.47)		
R^2_{adj}	15.4%	15.6%	15.0%	11.9%	13.3%

Panel B: Look-ahead Prior	1	2	3	4	5
$E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]$	33.48^{***}	30.84^{***}	28.41^{**}		
	(5.56)	(7.24)	(14.03)		
$\sigma^2_{t+1}[\Delta c_{t+2}] - \sigma^2_t[\Delta c_{t+2}]$				-67.93^{***}	-48.59^{**}
				(13.76)	(17.94)
Controls:					
Δc_{t+1}		0.01			1.88
		(1.40)			(1.69)
Δc_t		2.11^{*}			2.50^{*}
		(1.33)			(1.41)
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{3-state model}}_{\theta \text{ known}}$			3.89		
			(9.20)		
R^2_{adj}	14.5%	15.0%	14.1%	9.3%	11.9%

The anticipated utility approach reduces the number of state variables to three (the belief about the state in the general 2-state model, and the 2-dimensional belief about the state in the 3-state model).[5]

The purpose of the pricing exercise is to examine what features of the post-WW2 U.S. aggregate consumption and asset price data a realistic, general learning problem can help explain. Since we do not integrate out the parameter and model uncertainty in the pricing exercise, we focus on two aspects of the model that are likely to be robust to the introduction of *priced* parameter and model uncertainty.

1. *Ex-ante* versus *ex post*

   With learning *ex ante* expectations need not in general equal average *ex post* outcomes, which is the assumption in the typical rational expectations implementation. In the following, we argue that substantial components of the observed equity premium, excess return volatility, the degree of in-sample excess return predictability, and the time-series of the aggregate price-dividend ratio can be explained by the (nonstationary) time-path of mean parameter beliefs.

2. Permanent versus transitory shocks

   The shocks to mean parameter beliefs are permanent shocks to investor information sets. This has implications for, for instance, the volatility of long-run bond yields, and is different from a model with transitory shocks to state variables (such as our state beliefs, the long-run risk variable in Bansal and Yaron (2004), or the surplus consumption ratio in Campbell and Cochrane (1999)).

10.3.1 The model

The model is solved at the quarterly frequency, and the representative agent is assumed to have Epstein and Zin (1989) preferences, which are defined recursively as:

$$U_t = \left\{ (1-\beta) C_t^{1-1/\psi} + \beta \left(E_t \left[U_{t+1}^{1-\gamma} \right] \right)^{\frac{1-1/\psi}{1-\gamma}} \right\}^{\frac{1}{1-1/\psi}}, \tag{10.4}$$

[5]It would be computationally feasible to account for model uncertainty or to focus on parameter uncertainty over one of the parameters, but we leave such considerations for future research.

where C_t is the consumption, $\psi \neq 1$ is the intertemporal elasticity of substitution (IES) in consumption, and $\gamma \neq 1$ is the coefficient of relative risk aversion. These preferences imply the stochastic discount factor:

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \left(\beta \frac{PC_{t+1} + 1}{PC_t} \right)^{\frac{1/\psi - \gamma}{1-1/\psi}}, \tag{10.5}$$

where PC_t is the wealth-consumption ratio – that is, the price-dividend ratio for the claim to the stream of future aggregate consumption. The first component of the pricing kernel is that which obtains under standard power utility, while the second component is present if the agent has a preference for the timing of the resolution of uncertainty (i.e., if $\gamma \neq 1/\psi$). As mentioned earlier, we consider an anticipated utility approach to the pricing problem in terms of parameter and model uncertainty, while state uncertainty is priced.[6] This corresponds to a world where investors understand and account for business cycle fluctuations, but where they simply use their best guess for the parameters governing these dynamics.

Our goal in this section is to, for reasonable preference parameters, understand how learning affects pricing relative to the benchmark case of fixed parameters. Given that the consumption dynamics are not *ex post* calibrated (in particular in the historical prior case) but estimated in real-time, we also do not calibrate preference parameters to match any particular moment(s). Instead, we simply use the preference parameters of Bansal and Yaron (2004). Thus, $\gamma = 10$, $\psi = 1.5$, and $\beta = 0.998\hat{}3$.

Following both Bansal and Yaron (2004) and Lettau, Ludvigson, and Wachter (2008), we price a levered claim to the consumption stream with a leverage factor λ of 4.5. The annual consumption volatility over the post-war sample is only 1.34%, and so the systematic

[6]The model is solved numerically through value function iteration *at each time t* in the sample, conditional on the mean parameter beliefs at time t, which gives the time t asset prices. The state variables when solving this model are the beliefs about the hidden states of the economy for each model under consideration. For a detailed description of the model solution algorithm, please refer to the Appendix.

Cogley and Sargent (2009) argue that anticipated utility approach is a close approximation to the true Bayesian approach, although their analysis is with respect to time-separable preferences. Piazzesi and Schneider (2010) is an example of a recent application of an anticipated utility pricing framework with Epstein-Zin preferences.

annual dividend volatility is therefore about 6%. Quarterly log dividend growth is defined as:

$$\Delta d_t = \lambda \Delta c_t + \varepsilon_{d,t}, \tag{10.6}$$

where $\varepsilon_{d,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(-\frac{1}{2}\sigma_d^2, \sigma_d^2\right)$ is the idiosyncratic component of dividend growth. σ_d is chosen to match the observed annual 11.5% volatility of dividend growth reported in Bansal and Yaron (2004). With these choices of λ and σ_d we also in fact closely match the sample correlation they report between annual consumption and dividend growth (0.55).[7]

Unconditional Moments Table 10.3 reports realized asset pricing moments in the data, and also those generated by our learning models over the same sample period. The first 10 years are removed as a burn-in period to reduce concerns with regards to prior misspecification. We consider cases with and without parameter learning.

The models with parameter uncertainty match the observed equity premium reasonably well: 4.7% in the data versus 3.8% and 3.4% for the consumption only historical and look-ahead priors, respectively. The models where GDP is used as an additional signal, which as reported earlier have a more severe recession state, have average sample excess equity returns of 4.2% and 4.0% for the historical and the look-ahead priors, respectively. This compares favorably to the benchmark fixed parameters two and 3-state models which sample equity premiums are 1.5% and 1.8%, respectively. Thus, allowing for parameter uncertainty more than doubles the sample risk premiums, despite the fact that parameter and model uncertainty are not priced risk factors in the anticipated utility pricing framework. The high sample equity premium arises because of the specific time path of beliefs, which we discuss next.

The table also reports the average *ex ante* equity risk premium ($E_T\left[E\left(R_{m,t+1}^{excess}|I_t\right)\right]$, where I_t denotes the information set (beliefs) of agents at time t and $E_T[\cdot]$ denotes the

[7]The dividend dynamics imply that consumption and dividends are not cointegrated, which is a common assumption (e.g., Campbell and Cochrane (1999), and Bansal and Yaron (2004)). One could impose cointegration between consumption and dividends, but at the cost of an additional state variable. Further, it is possible to also learn about λ and σ_d^2. However, quarterly dividends are highly seasonal, which would severely complicate such an analysis. Further, data on stock repurchases is mainly annual. We leave a rigorous treatment of these issues to future research.

Table 10.3: Asset Price Moments

	Data	Historical prior		Look-ahead prior		Fixed parameters	
Moments	1957:Q2-2009:Q1	Cons. only	Cons. + GDP	Cons. only	Cons. + GDP	2-state model	3-state model
The real risk-free rate:							
$E_T(r_t^f)$	1.6%	3.8%	3.7%	3.7%	3.7%	3.7%	3.7%
$\sigma_T(r_t^f)$	1.6%	0.8%	0.9%	0.6%	0.8%	0.7%	0.8%
The dividend claim: $d_t = \lambda c_t + \varepsilon_{d,t}$							
ex post:							
$E_T(r_t - r_t^f)$	4.7%	3.8%	4.2%	3.4%	4.0%	1.5%	1.8%
$\sigma_T(r_t - r_t^f)$	17.1%	15.6%	15.7%	15.5%	15.4%	12.2%	12.4%
Sharpe ratio	0.27	0.24	0.27	0.22	0.26	0.12	0.14
$\sigma_T(pd_t)$	0.38	0.26	0.28	0.26	0.29	0.06	0.07
$Corr_T(pd_t^{Model}, pd_t^{Data})$	n/a	0.37	0.53	0.31	0.52	0.24	0.25
ex ante:							
$E_T[E_t(r_{t+1} - r_{t+1}^f)]$	n/a	1.5%	1.7%	1.4%	1.6%	1.5%	1.8%

sample average). The cases with parameter and model learning have about the same *ex ante* risk premium. This implies that more than half of the excess returns achieved in these models occur due to *ex post* positive surprises in updates of beliefs. This is one of the primary implications of learning for this sample. Interestingly, after the burn-in period, this effect is also strong in the look-ahead prior. With parameter and model uncertainty, agents beliefs quickly deviate from their full sample estimates, highlighting the difficulty of learning in real-time, similar to the problem faced by an econometrician. In particular, the sequence of shocks realized over the post-war sample generate a times series of beliefs that have a systematic time series pattern: the initial low mean and high volatility of

consumption growth causes an upward revision in the mean growth rates and a negative revision in the volatility parameters, as described in Section 3. Fama and French (2002) reach a similar conclusion in terms of the *ex post* versus the *ex ante* risk premium when looking at the time-series of the aggregate price-earnings and price-dividend ratios. Sargent and Cogley (2008) assume negatively biased beliefs in their model to highlight the same mechanism. The results we present here are consistent with their conclusions, but our models are estimated from fundamentals alone.

The equity return volatility is, in all the cases permitting parameter and model uncertainty, close to the 17.1% annual return volatility in the data (from 15.4% to 15.7%). In contrast, the equity return volatility in the models with fixed parameters is about 12%, which is almost all cash flow volatility as the annual dividend growth volatility is 11.5%. Thus, the sample variation in discount and growth rates arising from updates in agents' beliefs cause excess return volatility (Shiller, 1980). This is reflected in the sample volatility of the log price-dividend ratio, which is 0.38 in the data. In the cases with parameter and model uncertainty the volatility of the log price-dividend ratio lies between 0.26 and 0.29.[8] While this is only about three quarters of its volatility in the data, it is 4 to 5 times the volatility of the log price-dividend ratio in the benchmark fixed parameters models (here the volatility of the log price-dividend ratio is 0.06 for the 2-state model and 0.07 for the 3-state model).

The sample correlation between the log price-dividend ratios from the model versus the data, is 0.53 and 0.52 for the models using both GDP and consumption to estimate beliefs and 0.31 and 0.37 for the models using consumption only to estimate beliefs. The models with fixed parameters have lower correlations, 0.24 for the 2-state model and 0.25 for the 3-state model. As an alternative measure of the fit between the time-series of the sample price-level in the data versus those in the models considered here, the highest covariance between the price-dividend ratio in the data and the models with parameter and model uncertainty is 0.0573, whereas the highest covariance between the price-dividend ratio in

[8]The price-dividend ratio in each model is calculated as the corresponding in the data by summing the last four quarters of payouts to get annual payout. The price-dividend ratio from the data includes share repurchases in its definition of total dividends.

the data and the models with fixed parameters is 0.0067 – a difference close to an order of magnitude. Thus, with parameter and model learning the model tracks the aggregate stock market price level (normalized by dividends) much more closely than either of the models we consider with fixed parameters. The price-level, a first order moment, is arguably even more important than matching the second order moments that usually are the focus in asset pricing.

As a formal test of the learning model's match of the aggregate stock price level (the log D/P ratio) relative to the fixed parameter benchmark model, we run the following regression:

$$dp_t^{data} = \alpha + \beta_1 dp_t^{ParModUnc} + \beta_2 dp_t^{FP3} + \varepsilon_t, \tag{10.7}$$

where dp_t^{data} refers to the historical quarterly log dividend price ratio of the market portfolio, $dp_t^{ParModUnc}$ refers to the log dividend price ratio from the model with parameter and model uncertainty, and dp_t^{FP3} refers to the log dividend price ratio from the fixed parameters, 3-state model. The first four columns of Table 10.4 shows that the regression coefficient on the model with parameter and model uncertainty (β_1) is significant at the 1% level for both the historical and look-ahead priors, as well as whether learning is from realized consumption growth only or also including realized GDP growth. The R^2 ranges from 12% to 26% and is the lowest for the look-ahead prior with learning from consumption only, and the highest for the historical prior with learning from both consumption and GDP growth. As before, the results are shown after a 10-year burn-in period, from 1957 to 2009. The coefficient on the dividend yield from the fixed parameters model is insignificant in all of these cases. The fifth column of Table 10.4 shows the regression with only the dividend yield from the fixed parameters model. It is significant in this case, but the R^2 is only 6%. Finally, the last column of the table shows the regression with both the dividend yield from the fixed parameter model and the dividend yield from the historical prior with learning from both GDP and consumption growth, but where the dividend yield from the model with parameter and model learning has been orthogonalized with respect to the dividend yield from the fixed parameter model. The coefficient on the orthogonalized dividend yield (β_1) is still significant at the 1% level which implies that including the dividend yield from the model with parameter and model learning leads to a statistically significant (at the 1%

level) increase in the R^2, relative to the fixed parameters benchmark case. The increase in fit from the full learning models stems from a better match of the business cycle fluctuations in the dividend yield, as well as low-frequency fluctuations. In particular, with parameter learning the dividend yield displays a downward trend over the sample, similar to that found in the data as documented by, for instance, Fama and French (2002).

In sum, including parameter and model uncertainty leads to not only better fit of the unconditional asset pricing moments, but a significantly better fit of the realized aggregate stock price level in the post-WW2 era.

Table 10.4: Dividend Yield Regression

	Historical Prior		Look-ahead Prior		Fixed parameters	Historical Prior
	Cons.	Cons.	Cons.	Cons.	3-state model	Cons. + GDP
Variables	only	+ GDP	Only	+ GDP	only	(orthogonal)
constant	0.82	0.19	1.25	0.16	1.86	1.86
	(1.82)	(1.72)	(1.91)	(1.71)	(2.06)	(1.86)
$pd^{ParModUnc}$	0.47^{***}	0.79^{***}	0.37^{***}	0.61^{***}		0.79^{***}
	(0.13)	(0.14)	(0.12)	(0.11)		(0.14)
pd^{FP3}	0.73	0.28	0.92^{*}	0.40	1.45^{***}	1.45^{***}
	(0.46)	(0.43)	(0.49)	(0.43)	(0.56)	(0.51)
R^2	15.0%	25.8%	11.7%	20.0%	6.2%	25.8%

Permanent shocks and the volatility of long-run yields. With parameter and model uncertainty, the updates in mean beliefs constitute permanent shocks to expectations about consumption growth rates, consumption growth volatility, and higher order moments. This is a distinguishing feature of models with learning about constant quantities relative to learning about or observing a stationary underlying process (such as our state of the Markov chain, long-run risk in Bansal and Yaron (2004), or the surplus consumption ratio in Campbell and Cochrane (1999)). The latter models have transitory variables only in marginal utility growth. Shocks to a transitory state variable eventually die out, and so (very) long-run expectations are constant. Shocks to, for instance, the mean belief about

the unconditional growth rate of consumption are, on the other hand, permanent, leading to permanent shocks to marginal utility growth. This has implications for all asset prices, but can be most clearly seen when considering the volatility of long-run default-free real yields, which can be readily calculated from our model. Table 10.5 shows the volatility of annualized yields for default-free real, zero-coupon bonds at different maturities. The data column gives the volatility of yields on U.S. TIPS, calculated from monthly data for the longest available sample, 2003 to 2011, from the Federal Reserve Board, along with the standard error of the volatility estimates. In the remaining columns, the corresponding model-implied yield volatilities, calculated from each of the models considered in this paper over the post-WW2 sample, are given.

First, the yield volatilities for the models with parameter and model uncertainty are substantially higher than the yield volatilities from the models with fixed parameters. The 2-year yields are twice as volatile, while the 10-year yields are an order of magnitude more volatile. This is a direct consequence of the permanent shocks to expectations resulting from parameter learning, whereas the models with fixed parameters have constant long-run consumption growth mean and volatility. Notably, the long maturity yields in the data have about the same yield volatility as in the models with parameter uncertainty, and so this is another dimension along which learning about parameters and models can help explain historical asset pricing behavior.

Table 10.5: Real Risk-free Yield Volatilities

TIPS	*Data*	*Consumption*		*Consumption, GDP*		*Fixed Parameters*	
(2003 – 2011)	(*s.e.*)	*Historical*	*Lookahead*	*Historical*	*Lookahead*	2-*state*	3-*state*
5-yr yield	0.75%	0.35%	0.30%	0.44%	0.39%	0.17%	0.19%
	(0.18%)						
10-yr yield	0.45%	0.31%	0.27%	0.42%	0.36%	0.09%	0.10%
	(0.11%)						
20-yr yield	0.30%	0.30%	0.26%	0.42%	0.35%	0.05%	0.06%
	(0.06%)						
30-yr yield	*n/a*	0.30%	0.25%	0.42%	0.35%	0.03%	0.03%

Return Predictability Lastly, we consider excess market return forecasting regression using the dividend yield as the predictive variable. These regressions have a long history in asset pricing and remain a feature of the data that asset pricing models typically aim to explain (e.g., Campbell and Cochrane (1999), Bansal and Yaron (2004)). However, the strength of the empirical evidence is under debate (see, e.g., Stambaugh (1999), Ang and Bekaert (2007), Boudoukh, Richardson and Whitelaw (2008), and Goyal and Welch (2008) for critical analyses). Here we run standard forecasting regressions overlapping at the quarterly frequency using the sample of market returns and dividend yields as implied by each of the models. Note that, as before, we are not looking at population moments or average small-sample moments, but the single sample generated by feeding the models the actual sample of realized consumption growth.

Table 10.6 shows the forecasting regressions over different return forecasting horizons from the data. We use both the market dividend yield and the approximation to the consumption-wealth ratio, *cay*, of Lettau and Ludvigson (2001) to show the amount of predictability implied by these regressions in the data. We then run the same regressions using model implied returns and dividend yields. The benchmark models with fixed parameters (bottom right in the table) show no evidence of return predictability at the 5% significance level and the R^2's are very small. These models do, in fact, feature time-variation in the equity risk premium, but the standard deviation of the risk premiums are only about 0.5% per year and so the signal-to-noise ratio in these regressions is too small to result in significant predictability in a sample of the length we consider here. The models with parameter uncertainty, however, display significant in-sample return predictability and the regression coefficients and the R^2's are large and increasing in the forecasting horizon similar to those in the data. The *ex ante* predictability in these models is in fact similar to that in the fixed parameters cases, but since the parameters are updated at each point in time, there is significant *ex post* predictability. For instance, an increase in the mean parameters of consumption growth leads to high returns and lower dividend yield. Thus, a high dividend yield *in sample* forecasts high excess returns *in sample*. This is the same effect of learning as that pointed out in Timmermann (1993) and Lewellen and Shanken (2002). The models here show that the significant regression coefficients in the classical forecasting regressions

show up in the sample only in the model where there is parameter learning which generates a significant difference between ex ante expected returns and ex post realizations. Thus, the model predicts that the amount of predictability is much smaller out-of-sample, consistent with the empirical evidence in Goyal and Welch (2008) and Ang and Bekaert (2007).

Table 10.6: Return Forecasting Regressions

$$r_{t,t+q} - r_{f,t,t+q} = \alpha_q + \beta_{q,dp} \ln(D_t/P_t) + \varepsilon_{t,t+q}$$

	Data				Historical prior			
	$\ln(D_t/P_t) := cay_t$		$\ln(D_t/P_t) := \ln \frac{\Sigma_{j=0}^{3} D_{t-j}^{Mkt.}}{P_t^{Mkt.}}$		Cons. only		Cons. and GDP	
q	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}
1	1.19^{***}	4.67%	0.03^{*}	1.6%	0.04	1.4%	0.03	1.3%
	(0.31)		(0.02)		(0.03)		(0.02)	
4	4.29^{***}	15.65%	0.11^{**}	6.6%	0.18^{**}	8.3%	0.14^{**}	6.8%
	(1.18)		(0.05)		(0.07)		(0.06)	
8	7.60^{***}	28.1%	0.17^{*}	8.5%	0.38^{***}	19.2%	0.28^{***}	13.7%
	(1.72)		(0.10)		(0.09)		(0.08)	
16	12.31^{***}	41.6%	0.22^{**}	9.5%	0.61^{***}	28.4%	0.44^{***}	17.9%
	(1.82)		(0.11)		(0.15)		(0.13)	

	Look-ahead prior				Fixed parameters			
	Cons. only		Cons. and GDP		2-state model		3-state model	
q	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}
1	0.03	1.3%	0.03	0.9%	-0.01	0.0%	0.004	0.0%
	(0.02)		(0.03)		(0.062)		(0.062)	
4	0.18^{**}	7.7%	0.15^{**}	5.4%	0.19	1.0%	0.20	1.2%
	(0.07)		(0.07)		(0.17)		(0.16)	
8	0.38^{***}	18.3%	0.29^{**}	10.8%	0.37^{*}	2.3%	0.41^{*}	2.7%
	(0.12)		(0.09)		(0.24)		(0.23)	
16	0.64^{***}	28.9%	0.42^{**}	13.0%	0.26	0.7%	0.28	0.8%
	(0.17)		(0.17)		(0.31)		(0.30)	

Chapter 11

Conclusion

This paper studies the statistical problem and asset pricing implications of learning about parameters, states, and models in a standard class of models for consumption dynamics. Our approach is empirical, focuses on the specific implications generated by learning about U.S. consumption dynamics during the post World War II period, and contributes to a growing empirical literature documenting the importance of learning for asset prices (e.g., Malmendier and Nagel (2011), and Pastor and Veronesi (2003)).

We find broad support for the importance of learning about parameters and models. Agents' beliefs about consumption growth dynamics are strongly time-varying, nonstationary, and help explain the realized time-series of equity returns and price-dividend ratio. In particular, the new and significant relationship we document between contemporaneous realized returns and revisions in beliefs is strong support for the importance of learning. Incorporating learning and our estimated time-series of beliefs in a general equilibrium model uniformly improves the model fit with respect to the standard asset pricing moments.

Taken together, this evidence questions the typical implementations of rational expectations consumption-based exchange economy models, in which agents know with certainty the data generating process for aggregate consumption growth. Further, the nonstationary dynamics induced by learning about fixed quantities such as parameters and models translates to nonstationary dynamics in marginal utility growth and asset valuation ratios. This, in turn, implies that standard econometric approaches to model tests and parameter estimation should be used with caution (see also Cogley and Sargent (2008)).

The procedure implemented in this paper can in a straightforward way be implemented for other countries or markets, or extended to multi-country or multi-asset settings. For instance, learning about the joint dynamics of dividends and consumption is an interesting exercise abstracted away from in this paper. In terms of other countries, it is clear that the post World War II experience of Japan would lead to a very different path of beliefs. Learning about the joint dynamics of, say, the U.S. and Japan's economies would have interesting implications, not only for their respective equity markets, but also for the real exchange rate dynamics. It will in future research be interesting to consider priced parameter uncertainty with Epstein-Zin preferences. Parameter and model uncertainty will be major sources of anxiety for agents with preferences for early resolution of uncertainty as these risks are nonstationary and thus truly "long-run." As in Bansal and Yaron (2004), these sources of uncertainty will likely command high risk prices.

Bibliography

[1] Ai, H. (2010), "Information about Long-Run Risk: Asset Pricing Implications," forthcoming, *Journal of Finance.*

[2] Ang, A. and G. Bekaert (2007), "Return predictability: Is it there?", *Review of Financial Studies*, 20, 3, 651-707.

[3] Backus, D., M. Chernov and I. Martin (2009), "Disasters Implied by Equity Index Options", *NYU Working Paper*

[4] Bakshi, G. and G. Skoulakis (2010), "Do Subjective Expectations Explain Asset Pricing Puzzles?", *Journal of Financial Economics,* December 2010, 117 - 140.

[5] Bansal, R. and A. Yaron (2004), "Risks for the Long-Run: A Potential Resolution of Asset Pricing Puzzles", *Journal of Finance* 59(4), 1481 - 1509

[6] Bansal, R. and I. Shaliestovich (2010), "Confidence Risk and Asset Prices," *American Economic Review P&P*, 100, 537 – 541.

[7] Barberis, N. (2000), "Investing for the Long Run When Returns Are Predictable", *Journal of Finance* 55(1), 225 - 264

[8] Barillas, F., Hansen, L. and T. Sargent, "Doubts or Variability?", *Journal of Economic Theory*, forthcoming.

[9] Barro, R. (2006), "Rare Disasters and Asset Markets in the Twentieth Century", *Quarterly Journal of Economics* 121, 823 - 866

[10] Barro, R., Nakamura, E., Steinsson, J., and J. Ursua (2009), "Crises and Recoveries in an Empirical Model of Consumption Disasters," Columbia Business School working paper

[11] Barro, R. and J. Ursua (2008), "Consumption Disasters since 1870", *Brookings Papers on Economic Activity*

[12] Bhamra, H., L. Kuehn and I. Strebulaev (2009), "The Levered Equity Risk Premium and Credit Spreads: A Unified Framework", *Review of Financial Studies* (forthcoming)

[13] Boguth, O. and L. Kuehn (2009), "Consumption Volatility Risk," Working Paper Carnegie Mellon University

[14] Boudoukh, J., M. Richardson, and R. Whitelaw (2008), "The Myth of Long Horizon Predictability," *Review of Financial Studies* 21, 1577 - 1605.

[15] Brandt, M. , Q. Zeng, and L. Zhang (2004), "Equilibrium stock return dynamics under alternative rules of learning about hidden states," *Journal of Economic Dynamics and Control*, 28, 1925 – 1954.

[16] Brennan, M. and Y. Xia (2002), "Dynamic Asset Allocation Under Inflation," *Journal of Finance* 57, 1201 – 1238.

[17] Bray, M. M. and N. E. Savin (1986), "Rational Expectations Equilibria, Learning, and Model Specification," *Econometrica* 54, 1129 - 1160.

[18] Cagetti, M., L. Hansen, T. Sargent and N. Williams (2002), "Robustness and Pricing with Uncertain Growth", *Review of Financial Studies* 15, 363 - 404

[19] Carvalho, Johannes, Lopes, and Polson (2010a), "Particle Learning and Smoothing," *Statistical Science*, 25, 88-106.

[20] Carvalho, Johannes, Lopes, and Polson (2010b – forthcoming), "Particle learning: Simulation-based Bayesian inference," *Bayesian Statistics* 9.

[21] Cecchetti, S., P. Lam and N. Mark (1990), "Mean Reversion in Equilibrium Asset Prices", *American Economic Review* 80, 398 - 418

[22] Cecchetti, S., P. Lam and N. Mark (1993), "The Equity Premium and the Risk Free Rate: Matching the Moments", *Journal of Monetary Economics* 31, 21 - 46

[23] Chen, H. (2009), "Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure", *Journal of Finance* (forthcoming)

[24] Chen, H., S. Joslin, and N. Tran (2010), "Rare Disasters and Risk Sharing with Heterogeneous Beliefs," MIT Working paper

[25] Cogley, T. and T. Sargent (2008), "The Market Price of Risk and the Equity Premium: A Legacy of the Great Depression?", *Journal of Monetary Economics* 55, 454 - 476

[26] Cogley, T. and T. Sargent (2009), "Anticipated Utility and Rational Expectations as Approximations of Bayesian Decision Making", *International Economic Review* 49, 185 - 221

[27] David, A. and P. Veronesi (2009), "What ties return volatilities to price valuations and fundamentals?" University of Calgary Working Paper.

[28] Detemple, J. (1986), "Asset pricing in a production economy with incomplete information." *Journal of Finance*, 41, 383–390.

[29] Dothan, M. U. and D. Feldman (1986), "Equilibrium interest rates and multiperiod bonds in a partially observable economy." *Journal of Finance*, 41, 369 – 382.

[30] Drechsler, I. and A. Yaron (2008), "What's Vol Got to Do with It?", Working Paper, Wharton School of Business, University of Pennsylvania

[31] Epstein, L. and S. Zin (1989), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework", *Econometrica* 57, 937 - 969

[32] Fama, E. and K. French (1988), "Dividend yields and expected stock returns," *Journal of Financial Economics* 22, 3 - 25.

[33] Fama, E. and K. French (2002), "The Equity Premium," *Journal of Finance* 57, 637 – 659.

[34] Froot, K. and S. Posner (2002), "The Pricing of Event Risks with Parameter Uncertainty", *GENEVA Papers on Risk and Insurance - Theory* 27, 153 - 165

[35] Hall, R. (1978), "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, December 1978, 86(6), pp. 971-987.

[36] Gabaix, X. (2009), "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance", *NYU Working Paper*

[37] Gennotte, G. (1986), "Optimal Portfolio Choice under Incomplete Information," *Journal of Finance*, 61, 733-749.

[38] Geweke, J. (2001), "A note on some limitations of CRRA utility," *Economics Letters* 71, 341 - 345.

[39] Goyal A. and I. Welch, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," July 2008, *Review of Financial Studies* 21(4) 1455-1508.

[40] Hansen, L. (2007), "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," Richard T. Ely Lecture, *The American Economic Review* 97, No. 2., 1 - 30

[41] Hansen, L. and T. Sargent (2009), "Robustness, Estimation and Detection," Working paper.

[42] Hansen, L. and T. Sargent (2010), "Fragile Beliefs and the Price of Uncertainty," *Quantitative Economics*, Vol. 1, Issue 1, pp. 129-162

[43] Johannes, M. and N. Polson (2006), "Particle Filtering," *Springer Verlag Handbook of Financial Time Series*, edited by Torben G. Andersen Richard A. Davis, Jens-Peter Kreiss, and Thomas Mikosch, September 2006.

[44] Kandel, S. and R. Stambaugh (1990), "Expectations and Volatility of Consumption and Asset Returns", *Review of Financial Studies* 2, 207 - 232

[45] Kandel, S. and R. Stambaugh (1996), "On the Predictability of Stock Returns: An Asset-Allocation Perspective", *Journal of Finance* 51, 66 - 74

[46] Kreps, D. (1998), "Anticipated Utility and Dynamic Choice", *Frontiers of Research in Economic Theory*, (Cambridge: Cambridge University Press) 242 - 274

[47] Lettau, M. and S. Ludvigson (2001), "Consumption, Aggregate Wealth and Stock Returns," *Journal of Finance* 56, 815 - 849

[48] Lettau, M., S. Ludvigson and J. Wachter (2008), "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?", *Review of Financial Studies* 21(4), 1653 - 1687

[49] Lewellen, J. and J. Shanken (2002), "Learning, Asset-Pricing Tests, and Market Efficiency," *Journal of Finance* (57(3), 1113 - 1145.

[50] Lucas, R. (1978), "Asset Prices in an Exchange Economy," *Econometrica* 46, 1429-1446

[51] Lucas, R. and T. Sargent (1979), "After Keynesian Macroeconomics," The Federal Reserve Bank of Minneapolis, *Quarterly Review* 321.

[52] Malmendier, U. and S. Nagel (2011), "Depression babies: Do Macroeconomic Experience Affect Risk-Taking?", *Quarterly Journal of Economics* 126, 373 – 416.

[53] Mehra, R. and E. Prescott (1985), "The Equity Premium: A Puzzle", *Journal of Monetary Economics* 15, 145 - 161

[54] Moore, B. and H. Schaller (1996), "Learning, Regime Switches, and Equilibrium Asset Pricing Dynamics", *Journal of Economic Dynamics and Control* 20, 979 - 1006

[55] Pastor, L. (2000), "Portfolio Selection and Asset Pricing Models," *Journal of Finance* 55, 179 - 223

[56] Pastor, L. and P. Veronesi (2003), "Stock valuation and learning about profitability," *Journal of Finance*, 58, 1749 – 1789.

[57] Pastor, L. and P. Veronesi (2006), "Was there a NASDAQ bubble in the late 1990s?," *Journal of Financial Economics*, 81, 61 – 100.

[58] Pastor, L. and P. Veronesi (2009), "Learning in Financial Markets," *Annual Review of Financial Economics.*

[59] Piazzesi, M. and M. Schneider (2010), "Trend and Cycle in Bond Risk Premia," Working Paper Stanford University

[60] Rietz, T. (1988), "The equity risk premium: A solution?", *Journal of Monetary Economics*, Volume 22, Issue 1, July 1988, Pages 117-131

[61] Romer, C. (1989), "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1869-1908", *Journal of Political Economy* 97, 1 - 37

[62] Shaliastovich, I. (2008), "Learning, Confidence and Option Prices," Working Paper, Duke University

[63] Stambaugh, R. (1999), "Predictive Regressions." *Journal of Financial Economics,* 1999, 54, pp. 375–421.

[64] Timmermann, A. (1993), "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices." *Quarterly Journal of Economics*, 1993, 108, 1135-1145.

[65] Veronesi, P. (1999), "Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model," *Review of Financial Studies*, 12, 5, Winter 1999

[66] Veronesi, P. (2000), "How does information quality affect stock returns?" *Journal of Finance*, 55, .

[67] Weitzman, M. (2007), "Subjective Expectations and Asset-Return Puzzles," *American Economic Review*, 97(4), 1102-1130.

[68] Whitelaw, R. (2000), "Stock Market Risk and Return: An Equilibrium Approach", *Review of Financial Studies* 13, 521 - 547.

[69] Working, H. (1960), "Note on the Correlation of First Differences of Averages in a Random Chain," *Econometrica* 28(4), 916 - 918.

[70] Xia, Y. (2001), "Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation", *Journal of Finance* 56(1), 205 - 246

Chapter 12

Appendix

12.1 Existing literature and alternative approaches for parameter, state, and model uncertainty.

Our paper is related to a large literature studying the asset pricing implications of parameter or state learning. Most of this literature focuses on learning about a single unknown parameter or state variable (assuming the other parameters and/or states are known) that determines dividend dynamics and power utility. For example, Timmerman (1993) considers the effect of uncertainty on the average level of dividend growth, assuming other parameters are known, and shows in simple discounted cash-flow setting that parameter learning generates excess volatility and patterns consistent with the predictability evidence (see also Timmerman 1996). Lewellen and Shanken (2002) study the impact of learning about mean cash-flow parameters with exponential utility with a particular focus on return predictability.

Veronesi (2000) considers the case of learning about mean-dividend growth rates in a model with underlying dividend dynamics with power utility and focuses on the role of signal precision or information quality. Pastor and Veronesi (2003, 2006) study uncertainty about a fixed dividend-growth rate or profitability levels with an exogenously specified pricing kernel, in part motivated in order to derive cross-sectional implications. Weitzman (2007) and Bakshi and Skoulakis (2009) consider uncertainty over volatility.

Cogley and Sargent (2008) consider a 2-state Markov-switching model, parameter uncertainty over one of the transition probabilities, tilt beliefs to generate robustness via pessimistic beliefs, and use power utility. After calibrating the priors to the 1930s experience, they simulate data from a true model calibrated to the post War experience to show how priced parameter uncertainty and concerns for robustness impact asset prices, in terms of the finite sample distribution over various moments.

A number of papers consider state uncertainty, where the state evolves discretely via a Markov switching model or smoothing via a Gaussian process. Moore and Shaller (1996) consider consumption/dividend based Markov switching models with state learning and power utility. Brennen and Xia (2001) consider the problem of learning about dividend growth which is not a fixed parameter but a mean-reverting stochastic process, with power utility. Veronesi (2004) studies the implications of learning about a peso state in a Markov switching model with power utility. David and Veronesi (2010) consider a Markov switching model with learning about states.

In the case of Epstein-Zin utility, Brandt, Zeng, and Zhang (2004) consider alternative rules for learning about an unknown Markov state, assuming all parameters and the model is known. Lettau, Ludvigson, and Wachter (2008) consider information structures where the economic agents observe the parameters but learn about states in Markov switching consumption based asset pricing model. Chen and Pakos (2008) consider learning about the mean of consumption growth which is a Markov switching process. Ai (2010) studies learning in a production-based long-run risks model with Kalman learning about a persistent latent state variable. Bansal and Shaliastovich (2008) and Shaliastovich (2010) consider learning about the persistent component in a Bansal and Yaron (2004) style model with sub-optimal Kalman learning.

Additionally, some papers consider combinations of parameter or model uncertainty and robustness, see, e.g., Hansen and Sargent (2000,2009) and Hansen (2008).

12.2 Econometrics

This section briefly reviews the mechanics of sequential Bayesian learning and introduces the econometric methods needed to solve the high-dimensional learning problem. For ease of exposition, we abstract here from the problem of model uncertainty and drop the dependence on the model specification. Model uncertainty can be dealt with easily in a fashion analogous to the problem considered here.

The agent begins with initial beliefs over the parameters and states, $p(\theta, s_t) = p(s_t|\theta)\, p(\theta)$, and then updates via Bayes' rule. If at time t the agent holds beliefs $p\left(\theta, s_t|y^t\right)$, then updating occurs in a two step process by first computing the predictive distribution, $p\left(\theta, s_{t+1}|y^t\right)$, and then updating via the likelihood function, $p(y_{t+1}|s_{t+1}, \theta)$:

$$p\left(\theta, s_{t+1}|y^{t+1}\right) \propto p\left(y_{t+1}|\theta, s_{t+1}\right) p\left(\theta, s_{t+1}|y^t\right).$$

The predictive distribution is

$$p\left(\theta, s_{t+1}|y^t\right) = \int p\left(s_{t+1}|s_t, \theta\right) p\left(\theta, s_t|y^t\right) ds_t,$$

which shows the recursive nature of Bayesian updating, as $p\left(\theta, s_{t+1}|y^{t+1}\right)$ is functionally dependent on $p\left(\theta, s_t|y^t\right)$.

The main difficulty is characterizing $p\left(\theta, s_t|y^t\right)$ for each t, which is needed for sequential learning. Unfortunately, even though s_t is discretely valued, there is no analytical form for $p\left(\theta, s_t|y^t\right)$, as it is high-dimensional and the dependence on the data is complicated and nonlinear. We use Monte Carlo methods called particle filters to generate approximate samples from $p\left(\theta, s_t|y^t\right)$. Johannes and Polson (2008) developed the general approach we use, and it was extended and applied to Markov switching models by Carvalho, Johannes, Lopes, and Polson (2010a, 2010b) and Carvalho, Lopes and Polson (2009). Details of the algorithms are given in those papers.

The first step of the approach, data augmentation, introduces a conditional sufficient statistics, T_t, for the parameters. Sufficient statistics imply that the full posterior distribution of the parameters conditional on entire history of latent states and data takes a known functional form conditional on a vector of sufficient statistics: $p\left(\theta|s^t, y^t\right) = p\left(\theta|T_t\right)$, where $p\left(\theta|T_t\right)$ is a known distribution. The conditional sufficient statistics are given by

$T_{t+1} = \mathcal{T}(T_t, s_{t+1}, y_{t+1})$, where the function $\mathcal{T}(\cdot)$ is analytically known, which implies the sufficient statistics can be recursively updated. For Markov switching models, the sufficient statistics contain random variables such as the number of times and duration of each state visit, the mean and variance of y_t in those visits, etc. This step requires conjugate priors.

The key is that it is easier to sample from $p\left(\theta, T_t, s_t|y^t\right)$ than $p\left(\theta, s_t|y^t\right)$, where

$$p\left(\theta, T_t, s_t|y^t\right) = p\left(\theta|T_t\right) p\left(T_t, s_t|y^t\right). \tag{12.1}$$

By the definition of sufficient statistics and the use of conjugate priors, $p\left(\theta|T_t\right)$ is a known distribution (e.g., normal). This transforms the problem of sequential learning of parameters and states into one of sequential learning of states and sufficient statistics, and then standard updating by drawing from $p\left(\theta|T_t\right)$. The dimensionality of the target distribution, $p\left(\theta, T_t, s_t|y^t\right)$, is fixed as the sample size increases.

An $N-$ particle approximation, $p^N(\theta, T_t, s_t|y^t)$, approximates $p\left(\theta, T_t, s_t|y^t\right)$ via 'particles' $\left\{(\theta, T_t, s_t)^{(i)}\right\}_{i=1}^{N}$ so that:

$$p^N(\theta, T_t, s_t|y^t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{(\theta, T_t, s_t)^{(i)}},$$

where δ is a Dirac mass. A particle filtering algorithm merely consists of a recursive algorithm for generating new particles, $(\theta, T_{t+1}, s_{t+1})^{(i)}$, given existing particles and a new observation, y_{t+1}. The approach developed in Johannes and Polson (2008) and Carvalho, Johannes, Lopes, and Polson (2009a, 2009b) generates a direct or exact sample from $p^N(\theta, T_t, s_t|y^t)$, without resorting to importance sampling or other approximate methods. The algorithm is straightforward to code and runs extremely quickly so that it is possible to run for large values N, which is required to keep the Monte Carlo error low. These draws can be used to estimate parameters and states variables.

In addition to sequential parameter estimation, particle filters can also be used for Bayesian model comparison. Bayesian model comparison and hypothesis testing utilizes the Bayes factor, essentially a likelihood ratio between competing specifications. Formally, given a number of competing model specifications, generically labeled as model $\mathcal{M}_k$ and

$\mathcal{M}_j$, the Bayesian approach computes the probability of model k as:

$$p\left(\mathcal{M}_k|y^t\right) = \frac{p(y^t|\mathcal{M}_k)p\left(\mathcal{M}_k\right)}{\sum_{j=1}^{N} p(y^t|\mathcal{M}_j)p\left(\mathcal{M}_j\right)},$$

where $p\left(\mathcal{M}_k\right)$ is the prior probability of model k,

$$p(y^{t+1}|\mathcal{M}_k) = p(y_{t+1}|y^t, \mathcal{M}_k)p\left(y^{t-1}|\mathcal{M}_k\right),$$

and

$$p(y_{t+1}|y^t, \mathcal{M}_i) = \int p\left(y_{t+1}|\theta, s_t, \mathcal{M}_i\right) p\left(\theta, s_t|y^t, \mathcal{M}_i\right) d\left(\theta, s_t\right)$$

is the marginal likelihood of observation y_{t+1}, given data up to time t in model k. Marginal likelihoods are not known analytically and are difficult to compute even using MCMC methods. Since our algorithm provides approximate samples from $p\left(s_t, \theta|y^t\right)$, it is straightforward to estimate marginal likelihoods via

$$p^N(y_{t+1}|y^t, \mathcal{M}_k) = \frac{1}{N}\sum_{i=1}^{N} p\left(y_{t+1}|\left(\theta, s_t\right)^{(i)}, \mathcal{M}_k\right).$$

For all of our empirical results, we ran particle filtering algorithms with $N = 10K$ particles. We performed extensive simulations to insure that this number of particles insured a low and negligible Monte Carlo error.

12.3 Priors

Table 12.1 shows the prior parameters for the three different models we consider. The historical and look-ahead priors are different along some important dimensions. In particular, pre-WW2 consumption data is a lot more volatile than the post-war data (annual standard deviation of 4.8% in the pre-WW2 data versus 1.36% in post-WW2 data). This has been, in part, attributed to inferior pre-war data that is more noisy and sample that contains a more cyclical component of the economy (Romer, 1989). What is true, nevertheless, is that recessions were more frequent and lasted longer in the pre-WW2 data, and that the Great Depression was a worse recession than ever experienced afterwards, current crisis included. This is reflected in the disaster state in the 3-state models, in particular for the historical prior, akin to the disaster risk considered in Barro (2008).

For the historical prior, we have estimated, respectively, the 2- and 3-state models starting with very flat priors on the annual Shiller data. The posterior obtained at the end of the pre-war sample is transformed into a prior for the quarterly post-WW2 sample by dividing the average expected means and standard deviations within each regime by 4, and the average transition probability matrix, Π, is taken to the power of $1/4$. This is of necessity somewhat ad hoc - first, a 2-state model on annual data does not imply a 2-state model on quarterly data; second, one would usually divide standard deviations by 2 to go from annual to quarterly. However, a large fraction of the pre-WW2 excess volatility is likely due to noisy data, which is not what we intend to capture with our prior. What is more, applying priors where the mean belief of the standard deviation of consumption growth within each regime is counter-factually high, leads to a state identification issue: the difference in the average beliefs of the mean within each state is too small relative to the volatilities and so the procedure cannot identify the separate states.

The look-ahead priors have mean values equal to the posterior from the corresponding historical priors in 2009:$Q1$. These are very close to what would be the maximum likelihood estimates obtained from estimating the 2- and 3-state models using the post-WW2 quarterly sample. The look-ahead priors have lower consumption growth volatility and higher persistence of the good state relative to the historical priors. Thus, the look-ahead prior reflects an expectation in 1947:$Q1$ of the world having higher growth and lower volatility than in the period before WW2. In terms of the tightness of the priors, the expansion state (always state 1), which has occurred the most, has the tightest priors, the recession state (state 2) has flatter priors as this state is visited less often, while the disaster state (state 3), for the 3-state models, has the flattest priors. This state is the one agents has the least information about, as it is a rare event.

For the extended model with both consumption and GDP growth, the priors are set to match the consumption-only model as much as possible to minimize the priors' effect on the comparison of the models. Since the means of the hidden state variable are equal to the means of the consumption growth in each state, the priors of these means are the same as in the consumption-only model. We also match the prior means of the total variance of consumption growth with similar flatness. However, since the specification allows for

idiosyncratic noise in consumption growth ($\sigma_c \varepsilon_t^c$), we set both the mean of the variance of the hidden state variable in each state and the mean of the variance of the noise component to half of the prior mean of the total variance of consumption growth, with similar flatness. This way, the total prior mean variance of consumption growth, is the same as in the consumption only case. The priors for the transition probabilities are the same as in the consumption only case. For α and β in the GDP growth equation, the prior mean is -0.2 for α and 1.2 for β, and prior standard deviation is 0.45 for both. Finally, the prior mean of the idiosyncratic component of the variance of GDP growth is set by matching the variance of the GDP growth in the post-war data.

12.4 Time-Averaging of Consumption Data and Model Probabilities

The aggregate consumption data is time-averaged, which has implications for the volatility and autocorrelation structure of measured consumption growth. In particular, Working (1960) shows that time-averaging of i.i.d. data leads to lower variance (the variance is decreased by a factor of 1.5) and an autocorrelation of 0.25. Time-averaging can therefore artificially lead us to conclude that consumption growth follows a non-i.i.d. process (e.g., as we would get in the 2-state model with persistent states). Further, Hall (1978) argues theoretically and empirically that consumption growth is close to i.i.d. To ensure the rejection of the i.i.d. model we document in the paper is not an artifact of the time-averaging, we here assume the null hypothesis that consumption growth is in fact i.i.d., and remove the autocorrelation induced by time-averaging by creating the following residuals:

$$\nu_{c,t} = \Delta c_t - 0.25 * \Delta c_{t-1}. \tag{12.2}$$

We then redo the filtration exercise (parameters and models) and assign a prior probability of the i.i.d. model of 0.95. Figure 12.1 shows that also in this case, even with the strong model prior imposed, the i.i.d. model is rejected by the Bayesian agent about half-way through the sample.

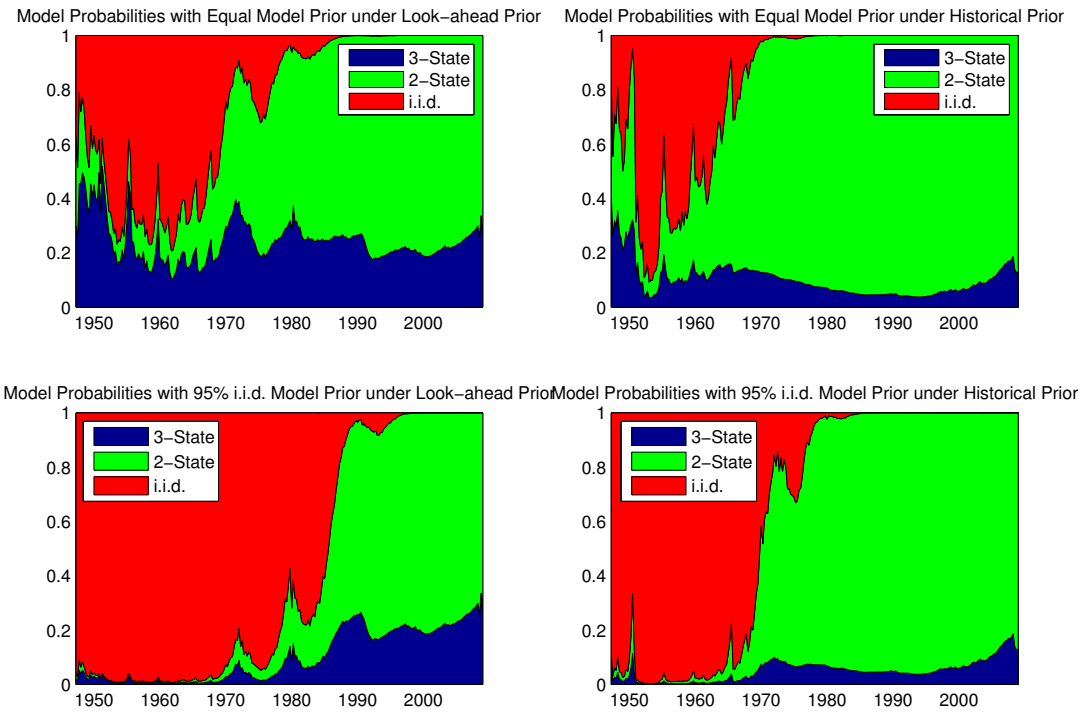

Figure 12.1: Model Probabilities and Time-Averaging of Consumption Data

12.5 Model solution and pricing

Here we give the details for how the prices of the consumption and aggregate equity claim in Section 4 are computed. At each point in time t, we price the equity claim given a set of model parameters, which are set equal to the mean beliefs at the time. The i.i.d. 2-state model, and the general 2- and 3-state models have parameters:

$$\begin{aligned}\theta^{(1)} &= \{\mu_1, \mu_2, \sigma_1, \sigma_2, \pi_{11}\} \\ \theta^{(2)} &= \{\mu_1, \mu_2, \sigma_1, \sigma_2, \pi_{11}, \pi_{22}\}, \\ \theta^{(3)} &= \{\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3, \pi_{11}, \pi_{12}, \pi_{22}, \pi_{23}, \pi_{13}, \pi_{33}\},\end{aligned}$$

respectively. In addition, there is the probability that the i.i.d 2-state model is the correct model, the probability that the general 2-state model is the correct model versus the residual probability of the 3-state model being the correct model. We also set these probabilities as constants when the agent prices the equity claim. Denote these probabilities p_1, p_2, and $p_3 = 1 - p_1 - p_2$. Thus, there is a total of 25 parameters that all are estimated using the particle filter and realized consumption (and GDP) data in real time. These mean parameter estimates will change at each time t, but we do not give the parameters time-subscripts to highlight that they are assumed to be constant following the anticipated utility framework in the pricing problem at each time t. In addition, there are the preference parameters γ, ψ, β, which are set to the values used in Bansal and Yaron (2004), and the leverage factor λ and the idiosyncratic dividend growth volatility σ_d. These parameters remain constant over the sample. When solving for the price-dividend ratio, we can and do ignore the idiosyncratic component of dividend growth.

First, we have to solve for the wealth-consumption ratio, PC. At each time t, the wealth-consumption ratio is solved using the recursion:

$$PC\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right) = \beta E\left[e^{(1-\gamma)\Delta c_{t+1}}\left(1 + PC\left(s_{t+1}^{(2)}, \tilde{s}_{t+1}^{(3)}\right)\right)^{\theta} | I_t\right]^{1/\theta}, \tag{12.3}$$

where the wealth-consumption ratio at time t is a function of the state-variables $s_t^{(2)}$ and $\tilde{s}_t^{(3)}$, and where I_t is the agent's information set which includes the mean parameter values used as constant parameters, as well as the mean state beliefs. The state-variable $s_t^{(2)}$ is the belief that the economy is in state 1 in the 2-state model. Remember that the states

are still hidden, even though all the parameters are set to constants, so this belief will have a support of $(0, 1)$. Similarly, $\tilde{s}_t^{(3)}$ is the 2×1 vector of state belief probabilities from the 3-state model – the probability of being in state 1 and the probability of being in state 2.

In the model solution, the agent updates beliefs about $s^{(2)}$ and $\tilde{s}^{(3)}$ only by observing realized consumption growth – he does not know which model is the true model, or which state is the current state, so this uncertainty must be integrated out in the model solution. Below is a conceptual algorithm for the model solution.[1]

1. Given a set of parameters, start with an initial guess of the function $PC\left(s^{(2)}, \tilde{s}^{(3)}\right)$ on a grid for the 3 state variables, which all have support $(0, 1)$.

2. For each value of $s^{(2)}$, $\tilde{s}^{(3)}$ on the grid, do points 3. – 8. below:

3. Draw a model (the i.i.d. 2-state mode, or the general 2-state or 3-state model) according to the model probabilities p_1, p_2, and p_3.

4. Draw the current state of this model (state 1, state 2 (or state 3)), using the state belief for the current values in the grid for $s_t^{(2)}$ or $\tilde{s}_t^{(3)}$. Note: this step is irrelevant for the i.i.d. 2-state model.

5. Given the model and the state, draw a random standard normal shock ε_{t+1}, and compute consumption growth as

$$\Delta c_{t+1} = \mu_{M,j} + \sigma_{M,j}\varepsilon_{t+1}, \tag{12.4}$$

   where the subscript M refers to the model and the subscript j refers to the state in the same model. The parameters are assumed known and constant as discussed above.

6. Given observed log consumption growth (Δc_{t+1}) (the agent does not observe the shock ε), update the agent's belief using Bayes' rule. When finding $s_{t+1}^{(2)}$, condition on the 2-state model being the correct model, and when finding $\tilde{s}_{t+1}^{(3)}$, condition on the 3-state

[1] In actually solving the model, we employ numerical integration and not Monte Carlo simulation to find the wealth-consumption ratio. We compute the price-dividend ratio by summing over zero-coupon dividend claims. While we implement the model solution in this way for faster and more accurate model solution, this additional level of detail is not necessary for conceptually understanding how prices are computed.

model being the correct model. See, e.g., Hamilton (1994) for how to update beliefs in switching regime models such as the ones considered here. Note that one has to update the belief for both models ($s^{(2)}$ and $\tilde{s}^{(3)}$), even though in the simulation of consumption growth we conditioned on one of the models, as the agent does not know the model.

7. Given $s_{t+1}^{(2)}$ and $\tilde{s}_{t+1}^{(3)}$ and the initial guess for PC, we have all we need to evaluate the expression inside the expectation of Equation (12.3).

8. Repeat 3.−7. *many* times and take the average of the different values calculated for the expression inside the expectation of Equation (12.3). Use this average as an estimate of the expectation in Equation (12.3). Store the resulting value for $PC\left(s^{(2)}, \tilde{s}^{(3)}\right)$ found for the current place in the grid for $s^{(2)}$ and $\tilde{s}^{(3)}$.

9. Once 3. – 8. has been implemented for all values of $s^{(2)}$ and $\tilde{s}^{(3)}$ on the grid, update the function $PC\left(s^{(2)}, \tilde{s}^{(3)}\right)$.

10. Iterate on 2. – 9. until a suitable convergence criterion for the PC function has been achieved.

Points 1. – 10. gives the wealth consumption ratio at time t. The pricing functional $PC\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right)$ must be computed in this way *for each* t, as the parameters will change at each time t. This is the anticipated utility component of the pricing. Denote the price-consumption ratio as a function of time t parameters as $PC_t\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right)$.

The price-dividend ratio can be found similarly, by iterating on the below expression in the same manner as above for each time t in the sample with its corresponding time t set of parameter values:

$$PD_t\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right) = E\left[\beta^{\theta} e^{(\lambda-\gamma)\Delta c_{t+1}} \left(\frac{PC_t\left(s_{t+1}^{(2)}, \tilde{s}_{t+1}^{(3)}\right)+1}{PC_t\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right)}\right)^{\theta-1} \left(1+PD_t\left(s_{t+1}^{(2)}, \tilde{s}_{t+1}^{(3)}\right)\right) | I_t\right]. \tag{12.5}$$

Finally, the returns to the equity claim are calculated as follows. For the return from time t to time $t+1$:

1. Set $s_t^{(2)}$ and $\tilde{s}_t^{(3)}$ equal to the mean state beliefs at time t (after parameter uncertainty is integrated out).

2. This gives the price dividend ratio at time t as $\frac{P_t}{D_t} = PD_t\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right)$.

3. Set $s_{t+1}^{(2)}$ and $\tilde{s}_{t+1}^{(3)}$ equal to the mean state beliefs at time $t+1$ (after parameter uncertainty is integrated out).

4. This gives the price dividend ratio at time $t+1$ as $\frac{P_{t+1}}{D_{t+1}} = PD_{t+1}\left(s_{t+1}^{(2)}, \tilde{s}_{t+1}^{(3)}\right)$.

5. Next, using realized (in the data) consumption growth, obtain dividend growth as:

$$\frac{D_{t+1}}{D_t} = \left(\frac{C_{t+1}}{C_t}\right)^{\lambda} e^{-\frac{1}{2}\sigma_d + \sigma_d \varepsilon_{t+1}}, \tag{12.6}$$

   where ε_{t+1} is a draw from a standard normal distribution independent of everything else. These simulated shocks are constrained to have mean zero and variance one over the sample, such that $E_T\left[e^{-\frac{1}{2}\sigma_{d.t+1}^2 + \sigma_d \varepsilon_{t+1}}\right] = 1$ (in practice, extremely close to 1). This is done to ensure that the level of the in-sample average equity return and equity return volatility are not affected by the (by chance) high or low draw of the idiosyncratic component of dividends, or (by chance) high or low volatility of idiosyncratic dividend growth.

6. Given this, the return is calculated as:

$$R_{t,t+1} = \frac{D_{t+1}}{D_t}\left(\frac{P_t}{D_t}\right)^{-1}\left(1 + \frac{P_{t+1}}{D_{t+1}}\right). \tag{12.7}$$

Table 12.1: Priors Specification

Historical priors

Priors for i.i.d. model			Priors for 2-state model			Priors for 3-state model		
Par.	Mean	St.Dev	Par.	Mean	St.Dev	Par.	Mean	St.Dev
μ	0.9%	0.5%	μ_1	1.0%	0.25%	μ_1	1.0%	0.25%
μ_J	-2.0%	0.5%	μ_2	-0.5%	0.5%	μ_2	-0.4%	0.5%
						μ_3	-2.0%	1.5%
σ^2	$(0.7\%)^2$	$(0.7\%)^2$	σ_1^2	$(0.5\%)^2$	$(0.5\%)^2$	σ_1^2	$(0.5\%)^2$	$(0.5\%)^2$
σ_J^2	$(1.0\%)^2$	$(1.0\%)^2$	σ_2^2	$(1.0\%)^2$	$(1.0\%)^2$	σ_2^2	$(1.0\%)^2$	$(1.0\%)^2$
						σ_3^2	$(1.5\%)^2$	$(1.5\%)^2$
λ	0.05	0.05	π_{11}	0.95	0.034	π_{11}	0.95	0.034
						$\hat{\pi}_{12}$	0.80	0.16
						$\hat{\pi}_{21}$	0.80	0.16
			π_{22}	0.80	0.16	π_{22}	0.75	0.19
						$\hat{\pi}_{31}$	0.33	0.24
						π_{33}	0.40	0.20

Look-ahead priors

Priors for i.i.d. model			Priors for 2-state model			Priors for 3-state model		
Par.	Mean	St.Dev	Par.	Mean	St.Dev	Par.	Mean	St.Dev
μ	0.63%	0.22%	μ_1	0.68%	0.18%	μ_1	0.68%	0.18%
μ_J	-1.2%	0.25%	μ_2	0.2%	0.5%	μ_2	0.3%	0.5%
						μ_3	-1.14%	0.5%
σ^2	$(0.45\%)^2$	$(0.45\%)^2$	σ_1^2	$(0.36\%)^2$	$(0.36\%)^2$	σ_1^2	$(0.35\%)^2$	$(0.35\%)^2$
σ_J^2	$(0.55\%)^2$	$(0.55\%)^2$	σ_2^2	$(0.7\%)^2$	$(0.7\%)^2$	σ_2^2	$(0.7\%)^2$	$(0.7\%)^2$
						σ_3^2	$(0.7\%)^2$	$(0.7\%)^2$
λ	0.05	0.05	π_{11}	0.95	0.034	π_{11}	0.95	0.034
						$\hat{\pi}_{12}$	0.83	0.14
						$\hat{\pi}_{21}$	0.67	0.24
			π_{22}	0.80	0.16	π_{22}	0.75	0.19
						$\hat{\pi}_{31}$	0.50	0.29
						π_{33}	0.33	0.24