Essays in Empirical Finance

by

James Christopher Morley

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

1999

Program Authorized to Offer Degree: Economics

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

James Christopher Morley

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of Supervisory Committee:

____
Charles R. Nelson

Reading Committee:

____
Richard Startz

____
Charles Engel

____
Eric Zivot

Date: ______________________

Doctoral Dissertation

University of Washington

Abstract

Essays in Empirical Finance

by James Christopher Morley

Chairperson of the Supervisory Committee
Professor Charles R. Nelson
Department of Economics

The essays are set out in three chapters investigating stock market efficiency, the relationship between stock market volatility and the market risk premium, and the Expectations Hypothesis of the term structure of interest rates, respectively. For each topic, advanced techniques in time-series economics are employed to challenge established findings on these classic issues of empirical finance. In the first chapter, we employ a time-varying parameter model and a Markov-switching model of the market risk premium to show that mean reversion of stock prices is exclusively a pre-World War II phenomenon driven by large and persistent changes in the level of market volatility in the 1930s. These findings reject the notion that mean reversion is inherent in post-War stock prices and challenge the idea that the pre-War evidence provides support for the presence of large fads in prices. In the second chapter, we further develop the Markov-switching model of the market risk premium employed in the first chapter and find that, when volatility feedback is properly accounted for, the empirical evidence suggests a statistically significant positive relationship between stock market volatility and expected excess returns. In the third chapter, we employ a state-space model with independently switching errors to estimate the behaviour of Canadian term premia over time. Contrary to the Expectations Hypothesis, the estimates suggest significant time variation corresponding to changes in interest rate volatility and the Canadian political climate.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful to many people in the Economics Department of the University of Washington for their help and guidance. In particular, I would like to thank Dick Startz, Charles Engel, and Eric Zivot for their encouragement and insightful suggestions on earlier drafts of the essays that comprise this manuscript. I would also like to thank Ming Chien Lo, Georgios Sakoulis, and Jeremy Piger for their friendship and countless hours of conversation, including many on the ideas presented here. Of course, the usual disclaimer applies.

My appreciation also extends to the Financial Markets Department at the Bank of Canada, especially Toni Gravelle, although the views expressed in this manuscript should in no way be attributed to the Bank of Canada.

I gratefully acknowledge financial support from Grover and Creta Ensley and the Ford and Louisa Van Voorhis endowment.

I would like to express my deepest gratitude to Charles Nelson and Chang-Jin Kim for their wisdom, guidance, and support. My passion for empirical economics arises from the profound influence of their graduate courses, while my understanding of the ideas presented here arises from the opportunity to proofread their highly instructive textbook.

Finally, I would like to thank Jennifer Milne for her patience and love.

For my parents, Jane and Terry

and

For Jennifer

OVERVIEW

The following three chapters investigate stock market efficiency, the relationship between stock market volatility and the market risk premium, and the Expectations Hypothesis of the term structure of interest rates, respectively. For each topic, advanced techniques in time-series economics are employed to challenge established findings on these classic issues of empirical finance.

The first chapter investigates whether changes in the market risk premium related to dramatic shifts in pre-World War II return volatility are responsible for the empirical evidence of mean reversion in stock prices. Notably, the timing of such changes reconciles previous findings of economically significant mean reversion with our findings using a time-varying parameter model of a disappearance of any underlying tendency for mean reversion during the post-War period. Furthermore, when we directly account for these changes with a Markov-switching risk premium model of mean reversion, we find that the remaining predictability of risk-adjusted returns is both economically and statistically insignificant. Thus, we conclude that the empirical evidence of mean reversion is a consequence of rational changes in the market risk premium, not persistent swings in market prices away from fundamental values.

The second chapter develops an extended version of Turner, Startz, and Nelson's (1989) Markov-switching model of stock returns. The model is motivated as an alternative version of Campbell and Hentschel's (1992) volatility feedback model, with news about future dividends subject to a two-state Markov-switching variance. We are able to identify a volatility feedback effect by assuming that economic agents acquire information about market volatility that is not directly available to econometricians. We find evidence for a positive relationship between volatility and the market risk premium, with agents' learning about the true volatility regime occurring at a similar speed throughout the entire 1926-96 sample period.

The third chapter develops a state-space model of the term structure of interest rates to test the Expectations Hypothesis. Our model employs the assumption of rational expectations to identify term premia from forward interest rates. Using Canadian data, we reject the Expectations Hypothesis. Instead, we find some evidence of a positive tradeoff between volatility and term premia. We also find evidence that Canadian term premia were negative during the late 1980s.

CHAPTER 1: Do Changes in the Market Risk Premium Explain the Empirical Evidence of Mean Reversion in Stock Prices?*

1.1 Introduction

Early empirical studies of market efficiency, surveyed by Fama (1970), found that daily and weekly price movements for individual securities show little evidence of economically significant predictability. Such findings led to Jensen's (1978) statement that "there is no other proposition in economics which has more solid empirical evidence supporting it than the Efficient Market Hypothesis." But, the evidence supporting the constant expected returns version of market efficiency seemed to be dramatically overturned when Fama and French (1988) and Poterba and Summers (1988) found that price movements for the stock market as a whole tend to be at least partially offset over longer horizons of months and years. This behaviour, labeled "mean reversion," is consistent with models of inefficient stock prices, such as Summers' (1986) *fads* model, in which market prices take persistent but temporary swings away from fundamental values. While Kim, Nelson, and Startz (1991), Richardson (1993), and Richardson and Stock (1989) have questioned the statistical significance of these findings, tests employed in studies of mean reversion ultimately have little power against alternatives like Summers' *fads* model. Thus, economically large but statistically insignificant point estimates, which according to Fama and French imply that as much as "25-45 percent of the variation of 3–5-year stock returns is predictable from past returns," provide a serious challenge to market efficiency.

In this chapter, however, we question the economic significance of the mean reversion evidence and its implications for market efficiency. First, we investigate whether point estimates for a time-varying parameter version of a regression model of mean reversion due to Jegadeesh (1991) have consistently remained economically significant throughout the past 70 years. Previous studies, including Fama and French (1988), Kim, Nelson, and Startz (1991), and Poterba and Summers (1988), have found

---

* This chapter contains materials first presented in Kim, Morley, and Nelson (1998).

that large point estimates supporting mean reversion are dependent upon the inclusion of pre-World War II data in statistical tests. We motivate our approach, then, as an attempt to distinguish between two plausible interpretations of this sensitivity to sample period. One interpretation, put forth by Poterba and Summers, is that the sensitivity does not reflect an underlying change in stock price behaviour. Instead, mean reversion is only statistically more evident in the pre-War years due to large movements in prices during the 1930s. If this interpretation is correct, time-varying parameter estimates of the mean reversion coefficient should change very little in the post-War period. The other interpretation, put forth by Kim, Nelson, and Startz, is that the sensitivity reflects an underlying change in stock price behaviour following the resolution of uncertainties surrounding the Great Depression and the War. If this interpretation is correct, time-varying parameter estimates of the mean reversion coefficient should change significantly during the post-War period as price behaviour contradicts the predictions of pre-War estimates. We find that point estimates do change significantly, with their path implying a disappearance of any underlying tendency for mean reversion during the post-War period.

We then investigate whether evidence of mean reversion remains economically significant when the market risk premium, instead of the regression coefficient, is allowed to change over time. For this task, we employ an extended version of the Jegadeesh (1991) regression model that incorporates a Markov-switching specification of time-varying expected returns due to Turner, Startz, and Nelson (1989). An important aspect of this specification is that it allows for a learning process to reflect the response of economic agents to new information about return volatility. When we account for mean reversion due to changes in the market risk premium in this way, we find little or no remaining evidence of mean reversion from other sources. Specifically, point estimates of the predictability of risk-adjusted returns are both economically and statistically insignificant, even though we include pre-War data in estimation. Our approach also provides us with inferences about the exact timing of changes in the market risk premium. These suggest that large changes in the market risk premium were closely related to the dramatic shifts in pre-War return volatility previously documented by French, Schwert, and Stambaugh (1987), Officer (1973) and Schwert (1989a,b). Notably, the timing of these changes is consistent with evidence of predictability for unadjusted returns at the 3–5-year horizon found in previous studies. But, the timing also implies that any underlying tendency for mean reversion at this horizon is purely a pre-War phenomenon, thus supporting our time-varying parameter results.

We interpret these findings as providing support for market efficiency. First, time-varying estimates of mean reversion imply post-War price behaviour that is entirely consistent with the traditional Efficient Market Hypothesis. Second, the statistically and economically weaker evidence for mean reversion using risk-adjusted returns suggests that much of the apparent pre-War predictability of stock prices is a consequence of a failure of the constant expected returns assumption, not deviations from fundamental values.

The rest of this chapter is organized as follows. Section 1.2 presents the details of the regression models employed in this chapter. Section 1.3 reports estimates for these models using monthly returns from the CRSP data file. Section 1.4 provides a brief discussion of our main findings and their implications. Section 1.5 concludes. Tables and figures follow the conclusion.

1.2 Models

The models we use to test for mean reversion are all variations of a simple regression model of stock returns due to Jegadeesh (1991). In each model, the dependent variable is one period returns, the independent variable is lagged multi-period returns, and the coefficient on lagged returns is negative under the alternative hypothesis of mean reversion. Richardson and Smith (1994) demonstrate the equivalence of this test for mean reversion with other tests, including the overlapping autoregression employed by Fama and French (1988) and the variance ratio test employed by Poterba and Summers (1988). In particular, the test statistics can all be represented as weighted-averages of sample autocorrelations, with the difference between tests being the weights.

Given this equivalence, there are two reasons why we choose the Jegadeesh model as our benchmark. First, Jegadeesh (1991) shows that, within a class of model specifications that includes both his model and the overlapping autoregression, his model has the highest asymptotic power against Summers' (1986) *fads* model.[1] Second, implementation of the Kalman filter is straightforward for estimation of a time-varying coefficient version of the Jegadeesh model, while it is cumbersome for an overlapping autoregression due to an imposed MA error structure.

The four model specifications employed in this chapter are presented below. Briefly, the first model is the basic Jegadeesh model, which replicates the stylized facts of the mean reversion literature and provides a benchmark for the other models. The second model allows for a time-varying regression coefficient to see how the point estimates of mean reversion have changed over time. The third model adds regime switching in the volatility of returns to account for heteroskedasticity. The fourth model allows for time-varying expected returns to reflect the response of economic agents to new information about return volatility. We use this last model to directly test whether changes in the market risk premium are responsible for any evidence of mean reversion.

1.2.1 The Basic Jegadeesh Model

The basic Jegadeesh (1991) model of mean reversion is given by the following equation:

$$r_t - m = b(k)\sum_{j=1}^{k}(r_{t-j} - m) + e_t \ , \tag{1.1}$$

where r_t is a one-period continuously compounded return, m is the unconditional mean of r_t, k is the holding period for lagged returns, and e_t is a serially uncorrelated error

---

[1] See Jegadeesh (1991) for details. Briefly, he uses the approximate slope criterion to determine the optimal aggregation intervals for the dependent and independent variables in terms of power against mean reversion. For Summers' (1986) *fads* model with a variety of parameter values, the optimal aggregation interval for the dependent variable is always one month.

term. The null hypothesis $H_0: b(k) = 0$ is that returns are serially uncorrelated with expected value m. The alternative hypothesis $H_A: b(k) < 0$ corresponds to the presence of mean reversion in stock prices.

We estimate $b(k)$ using ordinary least squares (OLS), although under the null hypothesis there is a small sample bias equal to $-1/(T-k)$, where T is the sample size (Jegadeesh, 1991; also, see Kendall, 1954, and Marriot and Pope, 1954). Following Jegadeesh, we use White's (1980) heteroskedasticity-consistent standard errors for our inferences.

An important note is that $b(k)$ is most comparable to the regression coefficient from a ½*k-period overlapping autoregression since the test statistics incorporate roughly the same sample autocorrelations (Richardson and Smith, 1994). Thus, given that Fama and French (1988) find the most evidence for mean reversion with 3–5-year overlapping autoregressions, we should expect to find the most evidence at holding periods of 6–10-years. By focusing exclusively on holding periods in that range and by ignoring the negative bias in OLS estimates, we stack the evidence in favour of finding mean reversion, thus making any findings of no mean reversion more convincing.

1.2.2 The Time-Varying Coefficient Model

The time-varying coefficient model of mean reversion is given by adding a time subscript to $b(k)$ in equation (1.1):

$$r_t - m = b_t(k)\sum_{j=1}^{k}(r_{t-j} - m) + e_t \ . \tag{1.2}$$

However, in order to get estimates of $b_t(k)$, given that we only observe T observations of r_t, we must impose some structure on its evolution. We choose a random walk specification for the time-varying coefficient:

$$b_t(k) = b_{t-1}(k) + v_t \ , \tag{1.3}$$

where v_t is a serially random error term.

The random walk specification is standard in the time-varying parameter literature.[2] Engle and Watson (1987) and Garbade (1977) argue that the random walk specification provides a good empirical model of the univariate behaviour of regression coefficients in many situations. Specifically, it represents a parsimonious way of allowing for permanent changes in regression coefficients, while at same time being fairly robust to misspecification as long as the true dynamics are highly autocorrelated.[3] Also, the model incorporates equation (1.1) as a special case when the variance of v_t collapses to zero.

We use the Kalman filter to make inferences about the time-varying coefficient. These inferences will be optimal given the following assumptions about the error processes in equations (1.2) and (1.3):

$$e_t \sim i.i.d.N(0, s_e^2)$$

$$v_t \sim i.i.d.N(0, s_v^2)$$

and

$$E[e_t v_t] = 0 \text{, for all } t \text{ and } t \ .$$

Estimates of the hyper-parameters m, s_e^2, and s_v^2 are obtained from maximum likelihood estimation based on the prediction error decomposition as discussed in Harvey (1990). Details of the Kalman filter, maximum likelihood estimation for this model, and calculation of inferences about $b_t(k)$ can be found in Appendix A.1.

---

[2] Examples include Engle and Watson (1987), Garbade (1977), Garbade and Watchel (1978), and Stock and Watson (1996). The classic reference in the literature is Cooley and Prescott (1976), although they incorporate a slightly more general random walk plus noise specification for the time-varying parameter.

[3] See Garbade (1977) for a Monte Carlo investigation of the consequences of misspecification. Briefly, he shows that a random walk specification is able to detect parameter instability even when the truth is a one-time discrete jump in the parameter or when the true parameter follows a persistent, but stationary, AR(1). In addition, he points out that the graphical representations of parameter estimates tend to reflect the true nature of instability, not just its presence.

1.2.3 The Time-Varying Coefficient with Markov-Switching Variance Model

Under the null hypothesis of no mean reversion, the previous model implies that stock returns are homoskedastic and Normal. However, French, Schwert, and Stambaugh (1987), Officer (1973) and Schwert (1989a,b) document large changes in the variance of stock returns over time. Their results suggest that these changes may be reasonably modeled as regime changes. Hamilton and Susmel (1994) confirm this by showing that, when regime switching is accounted for, persistent autoregressive conditional heteroskedasticity (ARCH) effects all but disappear at the longer horizons looked at in this chapter. Thus, we follow Schwert (1989b) and Turner, Startz, and Nelson (1989) by modeling the error term in equation (1.2) as having a 2-state Markov-switching variance.[4] Specifically, the time-varying coefficient with Markov-switching variance model is the same as the previous model with the following modification to the error term in equation (1.2):

$$e_t \sim i.i.d.N(0, s^2_{et})$$

$$s^2_{et} = s^2_{e0}(1 - S_t) + s^2_{e1}S_t$$

$$s^2_{e1} > s^2_{e0}.$$

That is, e_t is heteroskedastic, and its variance depends upon the unobserved 2-state first-order Markov-switching state variable S_t, which evolves according to the following transition probabilities:

$$\Pr[S_t = 0 | S_{t-1} = 0] = q \text{ and } \Pr[S_t = 1 | S_{t-1} = 1] = p.$$

[4] See Kim, Nelson, and Startz (1998) and Schaller and van Norden (1997) for more recent examples of this approach. We initially allowed both variances to switch, but found no evidence of switching for the error in equation (3). We also looked at 3-state Markov-switching, as in Kim, Nelson, and Startz (1998), but found that a third regime was unnecessary for the error in equation (2). In both cases, the time-varying parameter estimates were qualitatively unchanged. Results for these specifications are available from the authors upon request.

This model is operationally equivalent to the time-varying parameter model with Markov-switching heteroskedasticity introduced in Kim (1993), and can be estimated via approximate maximum likelihood as discussed in Kim (1994). Details can be found in Appendix A.2.

1.2.4 The Markov-Switching Risk Premium with Learning Model

The final model of mean reversion that we consider allows for time-varying expected returns stemming from changes in the market risk premium as return volatility switches between regimes. In particular, the Markov-switching process that determined the variance of the error term in the previous model is linked to corresponding changes in the mean, with economic agents responding to new information about the true underlying regime as in Turner, Startz, and Nelson (1989). Again, this model allows us to directly test whether changes in the market risk premium are responsible for any evidence of mean reversion in stock prices.

An aspect of this model that warrants some further discussion is the learning process for economic agents. Turner, Startz, and Nelson (1989) report that allowing for learning is necessary to avoid obscuring the empirical evidence of a positive risk premium. To see why, suppose that there is a transition into a persistent high variance regime. Elementary financial theory implies that the high variance regime should correspond to higher expected returns. However, if the transition is unanticipated, then a one-time negative return will occur as economic agents learn about the new regime and, consequently, bid prices downwards in order to make future expected returns high enough to compensate for the increase in non-diversifiable risk. Therefore, to properly identify the risk premium, we must separate out this negative return associated with learning about the high variance regime from the subsequent higher returns in a fully anticipated regime. To do so, we follow Turner, Startz, and Nelson by modeling the prior and posterior probabilities of the unobserved state variable that governs the variance regime as follows. First, at the beginning of each trading period, the prior probability of a high variance regime is equal to the expected value of the state variable, conditional on prior information. But, by the end of the period, the posterior probability is equal to the actual value of the state. Thus, the risk premium should correspond to the prior probability, whereas learning occurs when there is a large change in the posterior probability.

The Markov-switching risk premium with learning model of mean reversion is given as follows:

$$r_t - m_t = b(k)\sum_{j=1}^{k}(r_{t-j} - m_{t-j}) + e_t \ , \tag{1.4}$$

where

$$m_t = m_0 + m_1 S_t + g \cdot \Pr[S_t = 1 | y_{t-1}], \tag{1.5}$$

with $\Pr[S_t = 1 | y_{t-1}]$ denoting the probability that $S_t = 1$, given information up to time *t-1*. As in the previous model, the variance of the error term in equation (1.4) is dependent upon the unobserved 2-state first-order Markov-switching state variable S_t, with $S_t = 1$ corresponding to the high variance regime. Note that this model is the same as the first model, except m is allowed to change over time according to equation (1.5). The inclusion of the $\Pr[S_t = 1 | y_{t-1}]$ term in equation (1.5), then, allows for a learning effect. Specifically, when $\Pr[S_t = 1 | y_{t-1}]$ and S_t are different, learning about an unexpected new regime has occurred. On the other hand, when they are the same, S_t is fully anticipated. Thus, $m_1 + g$ reflects the increase in the risk premium associated with a fully anticipated high variance regime.

We use Hamilton’s (1989) filter to obtain maximum likelihood estimates of $b(k)$ and the other parameters of the model. Details can be found in Appendix A.3.

1.3 Empirical Results

1.3.1 Data

For all empirical exercises, we employ data from the CRSP file. Specifically, the basic data are total monthly returns for value-weighted and equal-weighted portfolios of all NYSE listed stocks, where "total" denotes capital gains plus dividend yields as calculated by CRSP. Following Fama and French (1988), we deflate nominal returns by the monthly CPI (not seasonally adjusted) for all urban consumers from Ibbotson Associates in order to get a measure of real returns. The data are converted into continuously compounded real returns by taking natural logarithms of the simple gross real returns.

The data are available for the period of January 1926 to December 1996. But, for the purpose of comparing our findings with those in Fama and French (1988), Poterba and Summers (1988), and Kim, Nelson, and Startz (1991), we employ data through December 1986 in our re-examination of the existing mean reversion evidence. This ensures that our findings are a direct consequence of our approach, rather than the addition of ten years of new data. However, the behaviour of stock prices in those ten years, which include the October 1987 crash, is also of considerable interest. Therefore, we conclude this section by examining the more recent empirical evidence.

1.3.2 A Re-examination of the Mean Reversion Evidence: 1926-86

First, we re-examine the existing evidence of mean reversion as presented in studies by Fama and French (1988), Kim, Nelson, and Startz (1991), and Poterba and Summers (1988). Specifically, we apply the four models described in the previous section to CRSP data for the sample period of January 1926 to December 1986 used in Kim, Nelson, and Startz. The results follow.

1.3.2.1 The Basic Jegadeesh Model

Table 1.1 reports results for the basic Jegadeesh model. The results are qualitatively similar to what has been previously reported in the literature on mean reversion, and are reported here to provide a benchmark case and motivate our extensions of the basic model. The results can be summarized as follows. First, when the entire 1926-86 sample is used, there is some evidence for economically significant mean reversion, especially for equal-weighted real returns. Furthermore, the evidence is strongest for holding periods of 72 and 96 months. These findings correspond to those in Fama and French (1988) and Poterba and Summers (1988). Second, the evidence is stronger for both types of returns when only pre-1947 data are used. Conversely, when only post-War data are used, the evidence for mean reversion is much weaker, especially for value-weighted real returns, which appear to be more consistent with mean aversion than mean reversion. These findings correspond to those in Kim, Nelson, and Startz (1991). Finally, we report a Chow statistic for a breakpoint in January 1947 to formalize what the subsample results seem to suggest. Notably, for both types of returns and holding periods of 72 and 96 months, the Chow statistic implies a change in stock price behaviour in the post-War era.

It may, however, be dangerous to draw strong conclusions from these results. First of all, the January 1947 breakpoint is assumed known *a priori*. A large body of research, exemplified by Andrews (1993) and Zivot and Andrews (1992), has revealed the potential pitfalls of assuming a known breakpoint. Furthermore, as Poterba and Summers (1988) point out, it may also be dangerous to ignore the pre-War data. They warn that "the 1930s, by virtue of the large movements in [stock] prices, contain a great deal of information about the persistence of [stock] price shocks." It is these considerations which motivate our use of the time-varying coefficient version of the regression model. In contrast to the basic Jegadeesh model, the time-varying coefficient model allows for changes in price behaviour without pre-specifying when and how many breakpoints occur, or even if they occur at all, while at the same time incorporating all available sample information.[5]

---

[5]However, the time-varying parameter approach is suboptimal for known types of changes (Engle and Watson, 1987).

1.3.2.2 The Time-Varying Coefficient Model

Figure 1.1 displays the results for the time-varying coefficient model. Both filtered and smoothed inferences about $b_t(k)$ are presented. The filtered inferences use information up to time *t*, thus simulating 'real-time' estimation suitable for forecasting, although the inferences are actually conditional on hyper-parameters estimated from the entire 1926-86 sample. The smoothed inferences use information up to the end of the sample, and are also conditional on the hyper-parameters. There are two things to notice about the time-varying inferences. First, allowing for time variation appears unnecessary for holding periods of 48 and 120 months (rows 1 and 4, respectively). In fact, the smoothed inferences are constant over time and almost identical to OLS estimates for the basic Jegadeesh model.[6] As with the OLS estimates, the statistical evidence of mean reversion is not particularly convincing at these horizons, although point estimates for the 48-month holding period certainly imply economically significant mean reversion. Second, there is considerable evidence of time variation in the slope coefficient for holding periods of 72 and 96 months (rows 2 and 3, respectively). The filtered and smoothed inferences typically start out negative and significant, but are dramatically revised upwards in the post-War period to essentially equal zero for the remaining sample period. The one exception to this behaviour is for value-weighted real returns with a 96-month holding period, for which inferences are practically equal to zero throughout the entire sample.

Table 1.2 reports the maximum likelihood estimates of the hyper-parameters.[7] The estimates for the variance of the time-varying coefficient confirm what is seen in Figure 1.2 and are generally consistent with the Chow test results reported above. Specifically, for both types of returns and holding periods of 48 and 120 months, the estimated variance of the time-varying coefficient is essentially equal to zero, suggesting no time

---

[6] Differences arise from the need for starting values in the Kalman filter, as discussed in Appendix A.1.

[7] All maximum likelihood estimation was conducted using the OPTMUM procedure for the GAUSS programming language. Numerical derivatives were used in estimation, as well as for calculation of asymptotic standard errors. Parameters were appropriately constrained (e.g., variances were constrained to be non-negative). Inferences appear robust to a variety of starting values.

variation in $b_t(k)$.[8] On the other hand, for holding periods of 72 and 96 months, there is evidence of time variation in $b_t(k)$. For value-weighted real returns, the likelihood ratio statistic for the null hypothesis of no time variation $H_0: s_v^2 = 0$ is 5.10 for the 72-month holding period with a corresponding *p*-value of 0.024. For equal-weighted real returns, the likelihood ratio statistic is 9.68 for the 72-month holding period and 6.98 for the 96-month holding period, with corresponding *p*-values of 0.002 and 0.008, respectively.[9]

As a diagnostic check, we test for serial correlation in both standardized forecast errors and squares of standardized forecast errors. Tests reveal some evidence for serial correlation (e.g., for value-weighted real returns and a holding period of 72-months, *Q*-statistics are *Q*(12)=20, *Q*(24)=37, and *Q*(36)=46 for the standardized forecast errors, and *Q*(12)=75, *Q*(24)=117, and *Q*(36)=171 for the squares of the standardized forecast errors). For the standardized forecast errors, this serial correlation appears to be a consequence of positive correlation at very short horizons and is generally significant at the 10% level, but not at the 5% level. For the squares of the standardized forecast errors, the serial correlation reflects heteroskedasticity in the form of persistent volatility and is significant at the 1% level. This heteroskedasticity motivates our use of the time-varying coefficient with Markov-switching variance model.

Another motivation for using the time-varying coefficient with Markov-switching variance model is the likely failure of the Normality assumption for the error term in equation (1.2). As discussed in Hamilton (1994a,b), this assumption is important for a general interpretation of inferences for time-varying parameter models, not just for estimation efficiency. Unfortunately, testing this assumption is complicated by the fact

---

[8] However, Stock and Watson (1998) point out that the maximum likelihood estimate of the coefficient variance has a point mass at zero when the true variance is small.

[9] Although we report standard errors for all the parameters in the tables, we emphasize the likelihood ratio statistic for testing this hypothesis since Garbade (1977) shows that it has good finite sample properties in detecting a variety of forms of parameter instability. To be clear, the likelihood ratio test does not have the highest local asymptotic power against specific forms of parameter instability such as a random walk coefficient (Nyblom, 1989). However, our interest is in more general forms of instability, as well as in the actual time path of the coefficient through time, which the time-varying parameter approach conveniently provides. In addition, the asymptotic distribution of the likelihood ratio statistic is concentrated towards the origin under the null hypothesis, making the test of parameter constancy conservative in the sense that

that we do not directly observe the sample errors. However, since our estimates in Table 1.2 suggest that the variance of the time-varying coefficient is relatively small, an informal argument can be made for using the conditional forecast error as a proxy for the sample errors. Furthermore, since the variance for the smoothed inferences is always smaller than for the filtered inferences, this approximation is most reasonable when we use the smoothed forecast errors. Therefore, we apply the test of Normality developed by Jarque and Bera (1980), which has a $\chi^2(2)$ distribution under the null hypothesis of Normality, to the smoothed standardized forecast errors of the time-varying coefficient model.[10] For value-weighted real returns, the test statistics range from 379.97 for the 72-month holding period to 462.41 for the 96-month holding period. For equal-weighted real returns, the test statistics range from 508.75 for the 96-month holding period to 812.40 for the 120-month holding period. In every case, Normality is strongly rejected, with *p*-values essentially equal to zero. Thus, we turn to the time-varying coefficient with Markov-switching variance model.

1.3.2.3 The Time-Varying Coefficient with Markov-Switching Variance Model

Figure 1.2 displays the results for the time-varying coefficient with Markov-switching variance model. The main result is that the general character of the inferences is relatively unchanged from the previous model. Again, smoothed inferences are constant for holding periods of 48 and 120 months. Also, for holding periods of 72 and 96 months, the filtered inferences start out negative and mostly significant, but are revised upwards during the post-War period to essentially equal zero for the remaining sample. The similarity of these inferences with those in Figure 1.1 provides some evidence that the consequences of a failure of Normality for the results presented in the previous subsection are not too severe.

---

reported *p*-values understate the true level of significance (Garbade; also see Kendall and Stuart, 1973). Again, this stacks the evidence against our main findings.

[10] See Jarque and Bera (1980) for details of this test. Briefly, the test statistic compares the third and fourth sample moments with their population counterparts under the null hypothesis of Normality.

There are, however, a few notable changes in our results that warrant discussion. First, while most of the inferences are qualitatively the same, the confidence bands are a somewhat tighter. This result is not too surprising since accounting for the heteroskedasticity in the data makes inferences more like generalized least square (GLS) and, therefore, more efficient. But, this result is encouraging given the approximate nature of estimation for this model due to the necessity of collapsing across states at each iteration of the Kalman filter.[11] The increased efficiency of the inferences suggests that the gains in efficiency from accounting for the heteroskedasticity outweigh the effects of collapsing. Second, the only point estimate evidence of mean reversion to survive the initial time-varying parameter treatment is considerably weaker when we allow for a Markov-switching variance. Specifically, filtered and smoothed inferences for the 48-month holding period and both types of returns are practically equal to zero throughout the sample. Again, this result is a likely consequence of accounting for the heteroskedasticity in the data. Intuitively, the information contained in the highly volatile Depression years implies mean reversion, but is given relatively less weight in inferences. This result is consistent with McQueen (1992), who finds that GLS estimation of an overlapping autoregression of stock returns produces much weaker evidence for mean reversion.

Table 1.3 reports the maximum likelihood estimates of the hyper-parameters for this model. Again, the estimates for the variance of the time-varying coefficient imply that there is no time variation in $b_t(k)$ for holding periods of 48 and 120 months. Also, for holding periods of 72 and 96 months, there is still some evidence of time variation, although it is weaker than for the previous model. For value-weighted real returns, the likelihood ratio statistic for the null hypothesis of no time variation $H_0: s_v^2 = 0$ is 2.00 for the 72-month holding period and 1.29 for the 96-month holding period, with corresponding *p*-values of 0.157 and 0.256, respectively. For equal-weighted real returns, the likelihood ratio statistic is 2.10 for the 72-month holding period and 3.80 for the 96-month holding period, with corresponding *p*-values of 0.147 and 0.045, respectively.

[11] See Kim (1994) for a discussion of the effects of collapsing on inferences.

In terms of the Markov-switching parameters and specification, making inferences is difficult since several of the classical assumptions of asymptotic distribution theory fail to hold. Notably, as discussed in Garcia (1995) and Hansen (1992), the transition probabilities p and q are not identified under the null hypothesis $H_0 : \sigma^2_{\varepsilon 0} = \sigma^2_{\varepsilon 1}$ of no Markov switching. Since this means that the distribution of test statistics are model and data dependent, Hansen argues for the use of computationally intensive simulations to determine the small sample distributions. Garcia, however, provides asymptotic distributions of a likelihood ratio test for a simple two-state Markov-switching model. The largest asymptotic critical value Garcia reports is 17.52, corresponding to a 1% significance level for a test of a 2-state Markov-switching mean and variance model with an uncorrelated and heteroskedastic noise function. All of our likelihood ratio statistics far exceed this critical value. For value-weighted real returns, the likelihood ratio statistics range from 81.90 for the 120-month holding period to 84.72 for the 96-month holding period. For equal-weighted real returns, the likelihood ratio statistics range from 127.04 for the 120-month holding period to 134.88 for the 96-month holding period. Therefore, the 2-state Markov-switching specification appears quite significant, with a high variance state that is somewhere between 5 and 10 times as volatile as the low-variance state. Furthermore, the strong persistence of both states implied by the estimates of the transition probabilities suggests that the Markovian specification for regime switching is appropriate (Engel and Hamilton, 1990).

Figure 1.3 displays the filtered and smoothed probabilities of a high variance state. The probabilities appear robust across different holding periods, providing further support for the Markov-switching specification. Also, both filtered and smoothed probabilities imply that most of the high volatility occurred in the 1930s, although there is some evidence of a reoccurrence in the 1970s. This result is consistent with previous findings on stock return volatility (e.g., French, Schwert, and Stambaugh, 1987, and Schwert, 1989a,b).

Finally, our diagnostic tests produce much weaker evidence of serial correlation for this model, especially in terms of the squares of the standardized forecast errors (e.g., for value-weighted real returns and a holding period of 72-months, *Q*-statistics are *Q*(12)=18, *Q*(24)=29, and *Q*(36)=36 for the standardized forecast errors, and *Q*(12)=9, *Q*(24)=30, and *Q*(36)=78 for the squares of the standardized forecast errors). Likewise, the evidence against Normality is weaker for this model. For value-weighted real returns, the Jarque and Bera (1980) test statistics range from 3.91 for the 120-month holding period to 5.40 for the 96-month holding period, with corresponding *p*-values of 0.14 and 0.07, respectively. For equal-weighted real returns, the test statistics range from 1.11 for the 120-month holding period to 4.35 for the 48-month holding period, with corresponding *p*-values of 0.57 and 0.11, respectively. These results suggest that 2-state Markov-switching specification is sufficient to model the heteroskedasticity and non-Normality in the error term.

While the diagnostics for this model represent a significant improvement over the previous specification, both models only make inferences about how price behaviour has changed over time. We are, of course, also interested in why price behaviour has changed. The timing of the volatility displayed in Figure 1.3 suggests that time-varying expected returns due to a trade-off between risk and return may be responsible. This prospect, discussed in greater detail in Section 1.4, motivates our use of the Markov-switching risk premium with learning model, the results for which we discuss next. Specifically, we look at risk-adjusted returns to see if the historical evidence of mean reversion can be fully accounted for by time-varying expected returns.

1.3.2.4 The Markov-Switching Risk Premium with Learning Model

Table 1.4 reports the estimates for the Markov-switching risk premium with learning model. The main result here is that the point estimates of $b(k)$ are much smaller in absolute value terms and less significant than for the basic Jegadeesh model. In particular, for value-weighted real returns and holding periods of 48 and 72 months, point estimates of $b(k)$ drop in absolute value terms from -0.0078 to -0.0011 and from -0.0060 to -0.0012, respectively. For equal-weighted real returns and holding periods of 48 and 96 months, point estimates drop in absolute value terms from -0.0091 to -0.0021 and from

-0.0088 to -0.0023, respectively. While the statistical significance of point estimates of mean reversion was never clear, these results clearly challenge the previously unassailable claim that, significant or not, point estimates imply relatively large predictable price movements over 3–5-year horizons. Specifically, once predictable price movements due to relatively infrequent changes in expected returns are accounted for, remaining movements in prices are largely unpredictable—not just statistically, but even according to the point estimates.

The point estimates for m_0, m_1, and g, corresponding to the risk premium and the learning process, are also of considerable interest. For value-weighted real returns, the results for these parameters are very similar to those reported in Turner, Startz, and Nelson (1989).[12] Specifically, the signs of the individual parameters are all as predicted by theory. A surprise transition into a high variance state has a negative effect on returns ($m_1 < 0$), corresponding to the learning process discussed in the previous section. Anticipation of a high variance state has a positive effect on returns ($g > 0$). In addition, for holding periods of 72, 96, and 120 months, expected returns are positive in both states of the world ($m_0 > 0$ and $m_0 + m_1 + g > 0$). Contrary to theory, but consistent with the findings in Turner, Startz, and Nelson, the estimated risk premium is negative ($m_1 + g < 0$) for all but the 120-month holding period, although it is not significant. However, for equal-weighted real returns all the signs are as predicted by theory, including a positive, but insignificant, risk premium ($m_1 + g > 0$) for all but the 48-month holding period.

As with the previous model, making inferences about the Markov-switching specification is difficult. However, the likelihood ratio statistics for the null hypothesis $H_0\!:m_1 = g = 0, s^2_{e0} = s^2_{e1}$ of no Markov-switching are very large. For value-weighted real returns, the test statistics range from 246.06 for the 48-month holding period to 90.36 for the 120-month holding period. For equal-weighted real returns, the test statistics range from 314.79 for the 48-month holding period to 138.07 for the 120-month holding period.

Figure 1.4 displays the filtered and smoothed probabilities of the high variance state for this model. The timing of the high variance state is qualitatively the same as for the previous model. Importantly, this timing provides a partial explanation of why we sometimes find a positive risk premium, whereas Turner, Startz, and Nelson (1989) do not. In their paper, they look at data from the post-War period only, with just one episode of high volatility in the 1970s associated with below average returns. By expanding the sample, we are able to look at two additional episodes of high volatility in the pre-War period. Furthermore, these episodes are more clearly distinct in character from 'normal' times than the 1970s episode and seem to be associated with above average returns. Of course, given that there are only three distinct episodes for the entire sample, it is not surprising that the risk premium fails to be estimated with any precision and is always insignificant.

Finally, our diagnostic tests for this model produce similar evidence of serial correlation as the time-varying coefficient with Markov-switching variance model (e.g., for value-weighted real returns and a holding period of 72-months, Q-statistics are $Q(12)=24$, $Q(24)=42$, and $Q(36)=52$ for the standardized forecast errors, and $Q(12)=36$, $Q(24)=58$, and $Q(36)=73$ for the squares of the standardized forecast errors). The evidence against Normality is somewhere in between what it was for the previous two models. For value-weighted real returns, the Jarque and Bera (1980) test statistics range from 36.81 for the 48-month holding period to 102.51 for the 120-month holding period. For equal-weighted real returns, the test statistics range from 92.10 for the 48-month holding period to 118.91 for the 120-month holding period. The p-values are all essentially equal to zero. These results suggest that a more complicated model specification may be needed to completely address the non-Normality of stock returns. However, there is no reason to think that addressing this lingering heteroskedasticity will qualitatively alter our main findings. Specifically, in previous studies, accounting for heteroskedasticity has only led to weaker evidence for mean reversion (e.g., Kim, Nelson, and Startz, 1991, and McQueen, 1992).

---

[12] Turner, Startz, and Nelson (1989) use excess returns from the Standard and Poor's composite index for

1.3.3 Recent Evidence: 1987-96

In this subsection, we examine the more recent empirical evidence and determine if it is consistent with our main findings above. First, we employ a post-sample forecast comparison to see how well the different models perform when confronted with ten years of data not used in estimation. Then, we look at the time-varying inferences for these ten years to see what they imply about the October 1987 crash and the subsequent behaviour of prices. Finally, we look at the estimated probabilities of a high variance regime in this period for the Markov-switching risk premium with learning model to see if any recurring evidence of mean reversion is a consequence of a tradeoff between risk and return.

1.3.3.1 Forecast Comparison

In order to evaluate competing forecasts, we use the traditional root mean squared error (RMSE) criterion. The RMSE statistics are calculated for the sample period of January 1987 + j to December 1996, where j is the forecast horizon in months. We use the parameter estimates given in Tables 1.1-1.4. The forecasting horizon j is the horizon at which the evidence of mean reversion should be strongest for a given holding period. Specifically, j is equal to half of k, where k is the holding period. Details for the construction of the j-period forecast for each model can be found in Appendix A.4.

Table 1.5 reports the post-sample (1987-96) RMSE for each model. Interestingly, the basic Jegadeesh model does quite well, although the Markov-switching risk premium with learning model is usually best for value-weighted real returns. The time-varying coefficient specifications, on the other hand, do relatively poorly in this sample period, especially for equal-weighted real returns. Thus, the remainder of this subsection is devoted to determining why the basic model did better than the time-varying coefficient models, and what the implications of these results are for our main findings in the previous section.

---

the sample period of January 1946 to December 1987.

1.3.3.2 The Time-Varying Coefficient Model

Figure 1.5 displays the post-sample inferences about $b_t(k)$ for the time-varying coefficient model. These inferences are found using estimates of the hyper-parameters reported in Table 1.2 and data for the period of January 1987 to December 1996. However, inferences are qualitatively unchanged if we use hyper-parameter estimates based on the entire 1926-96 sample. The main result here is that the October 1987 crash corresponds to a downward revision in the filtered inferences in all cases, but the magnitude of this revision is relatively small compared with the upward revisions displayed in Figure 1.1. Furthermore, the filtered and smoothed inferences are not even close to being significant following the crash. This result suggests that, while the behaviour of stock prices in the post-sample appears more consistent with mean reversion than before, thus explaining the better forecasting performance of the basic model, the behaviour is also statistically consistent with random walk behaviour.

1.3.3.3. The Markov-Switching Risk Premium with Learning Model

Figure 1.6 displays the post-sample probabilities of a high variance state for the Markov-switching risk premium with learning model. The purpose of this figure is to determine if the downward revision of the filtered inferences evident in Figure 1.5 is a consequence of time-variation in the risk premium. Not surprisingly, there is a spike in volatility surrounding the October 1987 crash, potentially explaining why the Markov-switching risk premium with learning model generally performs as well as the basic model and almost always better than the time-varying coefficient models in the post-sample forecasting exercise. However, the smoothed probabilities of this crash never get close to one and die out very quickly, thus explaining why the time-varying evidence of mean reversion in the post-sample is not particularly strong for the longer holding periods typically considered in studies of mean reversion. Therefore, we conclude that the post-sample experience is consistent with our main findings in the previous subsection. Specifically, the post-War predictability of stock price movements is not economically significant, and any evidence of mean reversion that does exist can be accounted for by changes in the risk premium driven by periods of high volatility.

1.4 Discussion

In this section, we provide a brief discussion of our main findings and their implications for market efficiency. First, we discuss why our results for the time-varying coefficient models support the Kim, Nelson, and Startz (1991) interpretation that the subsample evidence reflects a change in the underlying behaviour of stock prices following the War, with market efficiency holding in the post-War period. Then, we discuss in detail why our results for the Markov-switching risk premium with learning model support the time-varying expected returns explanation for the evidence of mean reversion, suggesting that market efficiency also holds in the pre-War period.

Figure 1.7 displays key examples of our time-varying inferences in Figure 1.1 and compares them with their OLS subsample counterparts. The purpose of this comparison is to help demonstrate how our results support the view that there was a change in the behaviour of stock prices in the post-War period. First, as noted in the previous section, the time paths of the inferences reflect the changes implied by the OLS subsample estimates. Superficially, these time paths alone support the Kim, Nelson, and Startz (1991) interpretation. However, the reason why the inferences lend more than superficial support has to do with the nature of the Kalman filter and smoothing. Notably, filtered and smoothed inferences would not have changed a lot if the OLS subsample evidence were merely a consequence of an unchanged underlying behaviour of prices being harder to detect in the post-War period. That is, as long as the underlying behaviour remains the same, inferences that reflect this behaviour will not consistently produce forecast errors which imply the presence of a different underlying behaviour. Thus, it is the general approach combined with the time paths depicted in Figure 1.7 that supports the view that price behaviour has changed. It is, however, the specific time paths that support the traditional Efficient Market Hypothesis in the post-War period. Specifically, the post-War inferences imply a disappearance of any tendency for price movements to be predictable using past returns.

Figure 1.8, then, summarizes how the Markov-switching risk premium with learning model supports the time-varying expected returns due to changes in the market risk premium interpretation of the mean reversion evidence. For demonstration, the first panel depicts two hypothetical episodes, each consisting of a long period of high expected returns followed by a long period of low expected returns. Over shorter periods of time, observed returns fluctuate randomly about these rationally determined expected values, reflecting market efficiency. In this scenario, a researcher who assumes constant expected returns will likely find evidence of mean reversion and mistakenly conclude that the stock market is inefficient. The second panel then displays the smoothed probability of the high mean and variance state for equal-weighted real returns and a 72-month holding period to show why this scenario is relevant. Notably, there are two full episodes of high followed by low expected returns in the early part of the sample. As is evident from the above scenario, these episodes will produce evidence for mean reversion in observed stock prices, even if market efficiency holds. Furthermore, the timing of these episodes implies that the evidence of mean reversion should be strongest at the 3–5-year horizon found in previous studies. The timing also implies that mean reversion should be largely a pre-War phenomenon, as we have found with our time-varying inferences.

The estimates for equal-weighted real returns with a 72-month holding period presented in Table 1.4 confirm that time-varying expected returns are likely responsible for evidence of mean reversion, as implied by Figure 1.8. First, the estimated increase in the risk premium in a fully anticipated high variance state is relatively large ($m_1 + g = 0.0115$), although its magnitude is economically reasonable.[13] This suggests

[13] As a check on the economic reasonableness of the risk premium, we calculate the implied coefficient of relative risk aversion for the representative investor, as in Merton (1980). Specifically, the coefficient of relative risk aversion is estimated by the ratio of the market risk premium to the expected variance of market returns. This ratio is about 3.92 in the low variance state, but is only 0.82 in the high variance state, although it is probably not statistically different than 3.92 in both states given the uncertainty surrounding the estimates of the risk premium in the high variance state. These results compare favourably to those in French, Schwert, and Stambaugh (1987), who estimate a ratio of about 2.41 for an ARCH-in-mean model of the risk premium using the Standard and Poor's composite index for the sample period of January 1928

that the assumption of constant expected returns is inappropriate, even as an approximation. Second, the point estimate of $b(k)$, which corresponded to the strongest evidence of mean reversion in Table 1.1, drops in absolute value from -0.0166 to -0.0069 and is statistically insignificant when we account for time-varying expected returns due to changes in the risk premium. These results, like the findings in a related study by Cecchetti, Lam, and Mark (1990), argue against what is referred to in that paper as "a tendency to conclude that evidence of mean reversion in stock prices constitutes a rejection of equilibrium models of rational asset pricing."[14] Contrary to that tendency, we conclude that time-varying expected returns due to changes in the market risk premium are largely responsible for evidence of mean reversion, thus supporting market efficiency.

1.5 Conclusion

This chapter challenges and explains the mean reversion evidence presented in Fama and French (1988) and Poterba and Summers (1988). Like Kim, Nelson, and Startz (1991), McQueen (1992), Richardson (1993), and Richardson and Stock (1989), we question the strength of the evidence for long-horizon mean reversion. In particular, by employing time-varying coefficient versions of a mean reversion regression model due to Jegadeesh (1991), we find that post-War stock prices show little or no tendency for mean reversion, even when we look at the most recent period that includes the October 1987 crash. However, for both time-varying specifications, with and without a Markov-switching variance, we find some evidence of mean reversion in the pre-War era. We then try to explain this evidence and why the behaviour of prices might have changed since that time. Theoretical explanations of mean reversion have typically been based on the existence of either fads or equilibrium based time-varying expected returns.[15] In this

---

to December 1984. In fact, our general results about the size and timing of changes in the market risk premium correspond well to the results for the ARCH-in-mean model reported in that paper.

[14] Briefly, Cecchetti, Lam, and Mark (1990) show that the levels of mean reversion implied by point estimates in Fama and French (1988) and Poterba and Summers (1988) are consistent with a relatively simple equilibrium model of asset prices that incorporates a Markov-switching endowment process estimated from consumption, output, and dividends series.

[15] For an introduction to this literature, see Shleifer and Summers (1990) for fads and Cecchetti, Lam, and Mark (1990) for equilibrium based time-varying expected returns.

chapter, we take the rationally determined time-varying expected returns approach and examine whether an empirical measure of the risk premium due to Turner, Startz, and Nelson (1989) can account for the evidence. Notably, we find that the evidence for mean reversion is much weaker for risk-adjusted returns. Furthermore, the timing of changes in the estimated risk premium corresponds directly to the changes in price behaviour implied by our time-varying coefficient estimates. Thus, we conclude that the behaviour of stock prices in the past 70 years provides no convincing evidence for systematic departures from market efficiency.

Possible future directions for research include applying the techniques presented in this chapter to regressions of stock returns on lagged values of financial fundamentals or macroeconomic indicators. Also, we could look at data sets covering longer time spans, such as the data set described in Schwert (1990); although the CRSP data are the best available for the recent U.S. experience.

TABLE 1.1: The Basic Jegadeesh Model – OLS Estimates and Chow Breakpoint Test, 1926-86

Value-Weighted Real Returns

k	***b*** *(k)* 1926-86	***b*** *(k)* 1926-46	***b*** *(k)* 1947-86	Chow Breakpoint *F* -Statistic	*p* -value
48	-0.0078	-0.0150	-0.0002	1.2640	0.2832
	(0.0096)	(0.0152)	(0.0075)		
72	-0.0060	-0.0338*	0.0052	5.4376	0.0045
	(0.0066)	(0.0175)	(0.0059)		
96	0.0007	-0.0269	0.0045	3.1061	0.0455
	(0.0048)	(0.0221)	(0.0044)		
120	0.0029	-0.0013	0.0029	0.2120	0.8091
	(0.0039)	(0.0199)	(0.0038)		

Equal-Weighted Real Returns

k	***b*** *(k)* 1926-86	***b*** *(k)* 1926-46	***b*** *(k)* 1947-86	Chow Breakpoint *F* -Statistic	*p* -value
48	-0.0091	-0.0102	-0.0070	0.0940	0.9103
	(0.0101)	(0.0142)	(0.0098)		
72	-0.0166*	-0.0324*	-0.0021	4.4264	0.0123
	(0.0092)	(0.0168)	(0.0082)		
96	-0.0088	-0.0373**	-0.0010	4.3163	0.0137
	(0.0064)	(0.0181)	(0.0062)		
120	-0.0009	-0.0021	-0.0009	0.0197	0.9805
	(0.0058)	(0.0250)	(0.0057)		

White's (1980) heteroskedasticity-consistent standard errors are reported in parentheses.
Samples are adjusted for lagged variables.
Chow statistic is calculated for breakpoint in January 1947.
* t -statistic for H_0 : $b(k)$ =0 is significant at 10% level.
** t -statistic for H_0 : $b(k)$ =0 is significant at 5% level.

TABLE 1.2: The Time-Varying Coefficient Model – Maximum Likelihood Estimates, 1926-86

Value-Weighted Real Returns

k	Adjusted Sample	s_v	s_e	m	log-likelihood value
48	1935-86	0.0001	0.0460	0.0061	1035.68
		(0.0015)	(0.0013)	(0.0013)	
72	1936-86	0.0012**	0.0459	0.0055	1014.89
		(0.0008)	(0.0013)	(0.0015)	
96	1937-86	0.0003	0.0462	0.0071	992.29
		(0.0005)	(0.0013)	(0.0020)	
120	1938-86	0.0000	0.0453	0.0029	982.77
		(0.0001)	(0.0013)	(0.0018)	

Equal-Weighted Real Returns

k	Adjusted Sample	s_v	s_e	m	log-likelihood value
48	1935-86	0.0000	0.0600	0.0094	869.67
		(0.0001)	(0.0017)	(0.0017)	
72	1936-86	0.0015***	0.0592	0.0100	858.05
		(0.0009)	(0.0017)	(0.0012)	
96	1937-86	0.0016***	0.0592	0.0097	841.14
		(0.0010)	(0.0017)	(0.0009)	
120	1938-86	0.0000	0.0582	0.0074	836.39
		(0.0009)	(0.0017)	(0.0018)	

Asymptotic standard errors based upon second derivatives are reported in parentheses.
Samples are adjusted for lagged variables and starting values of Kalman filter.
** Likelihood ratio statistic for H_0: s_v =0 is significant at 5% level.
*** Likelihood ratio statistic for H_0: s_v =0 is significant at 1% level.

TABLE 1.3: The Time-Varying Coefficient Model with Markov-Switching Variance Model – Maximum Likelihood Estimates, 1926-86

Value-Weighted Real Returns

k	Adjusted Sample	s_v	s_{e0}	s_{e1}	q	p	m	log-likelihood value
48	1935-86	0.0000	0.0373	0.0869	0.9826	0.8676	0.0079	1077.27
		(0.0002)	(0.0016)	(0.0111)	(0.0117)	(0.0831)	(0.0017)	
72	1936-86	0.0007	0.0368	0.0877	0.9789	0.8461	0.0077	1057.21
		(0.0006)	(0.0018)	(0.0119)	(0.0136)	(0.0960)	(0.0017)	
96	1937-86	0.0005	0.0368	0.0872	0.9795	0.8562	0.0086	1034.65
		(0.0004)	(0.0018)	(0.0118)	(0.0133)	(0.0927)	(0.0018)	
120	1938-86	0.0000	0.0365	0.0857	0.9826	0.8656	0.0114	1023.72
		(0.0001)	(0.0017)	(0.0115)	(0.0129)	(0.0913)	(0.0055)	

Equal-Weighted Real Returns

k	Adjusted Sample	s_v	s_{e0}	s_{e1}	q	p	m	log-likelihood value
48	1935-86	0.0000	0.0404	0.0963	0.9730	0.9190	0.0112	937.07
		(0.0001)	(0.0020)	(0.0083)	(0.0119)	(0.0386)	(0.0021)	
72	1936-86	0.0011	0.0403	0.0950	0.9750	0.9258	0.0104	923.48
		(0.0009)	(0.0020)	(0.0078)	(0.0116)	(0.0361)	(0.0011)	
96	1937-86	0.0012**	0.0400	0.0974	0.9727	0.9151	0.0104	908.58
		(0.0009)	(0.0019)	(0.0081)	(0.0115)	(0.0388)	(0.0009)	
120	1938-86	0.0000	0.0410	0.1032	0.9728	0.8768	0.0103	899.91
		(0.0003)	(0.0021)	(0.0108)	(0.0121)	(0.0530)	(0.0005)	

Asymptotic standard errors based upon second derivatives are reported in parentheses.
Samples are adjusted for lagged variables and starting values of Kalman filter.
** Likelihood ratio statistic for $H_0: s_v=0$ is significant at 5% level.

TABLE 1.4: The Markov-Switching Risk Premium with Learning Model – Maximum Likelihood Estimates, 1926-86

Value-Weighted Real Returns

k	$b(k)$	s_{e0}	s_{e1}	q	p	m_0	m_1	g	log-likelihood value
48	-0.0011	0.0385	0.1126	0.9947	0.9606	0.0068	-0.0314	0.0207	1116.07
	(0.0030)	(0.0014)	(0.0087)	(0.0030)	(0.0163)	(0.0024)	(0.0143)	(0.0131)	
72	-0.0012	0.0383	0.1131	0.9938	0.9438	0.0068	-0.0234	0.0223	1099.08
	(0.0026)	(0.0014)	(0.0106)	(0.0035)	(0.0231)	(0.0022)	(0.0154)	(0.0141)	
96	0.0002	0.0353	0.0757	0.9820	0.9073	0.0057	-0.0314	0.0291	1097.59
	(0.0016)	(0.0018)	(0.0078)	(0.0087)	(0.0302)	(0.0028)	(0.0101)	(0.0120)	
120	0.0010	0.0348	0.0756	0.9812	0.9031	0.0033	-0.0301	0.0335	1062.28
	(0.0017)	(0.0018)	(0.0078)	(0.0091)	(0.0329)	(0.0042)	(0.0105)	(0.0121)	

Equal-Weighted Real Returns

k	$b(k)$	s_{e0}	s_{e1}	q	p	m_0	m_1	g	log-likelihood value
48	-0.0021	0.0474	0.1521	0.9954	0.9679	0.0094	-0.0221	0.0170	954.55
	(0.0042)	(0.0019)	(0.0117)	(0.0032)	(0.0162)	(0.0025)	(0.0192)	(0.0153)	
72	-0.0069	0.0474	0.1576	0.9949	0.9542	0.0088	-0.0068	0.0183	944.48
	(0.0046)	(0.0018)	(0.0147)	(0.0031)	(0.0208)	(0.0024)	(0.0124)	(0.0129)	
96	-0.0023	0.0441	0.1119	0.9877	0.9282	0.0093	-0.0239	0.0264	948.61
	(0.0032)	(0.0021)	(0.0117)	(0.0058)	(0.0281)	(0.0030)	(0.0145)	(0.0139)	
120	-0.0002	0.0432	0.1089	0.9870	0.9203	0.0075	-0.0297	0.0313	926.65
	(0.0017)	(0.0021)	(0.0117)	(0.0062)	(0.0333)	(0.0033)	(0.0140)	(0.0144)	

Asymptotic standard errors based upon second derivatives are reported in parentheses.
Samples are adjusted for lagged variables.

TABLE 1.5: Forecast Comparison – Post-Sample Root Mean Squared Error, 1987-96

Value-Weighted Real Returns

Model	*Forecast Horizon (Months)* 24	36	48	60
1	1.21$^{(4)}$	1.07$^{(3)}$	0.99$^{(2)}$	1.11$^{(1)}$
2	0.99$^{(1)}$	1.30$^{(4)}$	1.04$^{(3)}$	1.44$^{(3)}$
3	1.00$^{(2)}$	1.07$^{(2)}$	1.24$^{(4)}$	1.92$^{(4)}$
4	1.05$^{(3)}$	0.96$^{(1)}$	0.95$^{(1)}$	1.29$^{(2)}$

Equal-Weighted Real Returns

Model	*Forecast Horizon (Months)* 24	36	48	60
1	1.32$^{(1)}$	1.60$^{(1)}$	1.05$^{(1)}$	1.25$^{(2)}$
2	1.61$^{(3)}$	2.33$^{(3)}$	2.31$^{(3)}$	1.22$^{(1)}$
3	1.89$^{(4)}$	2.40$^{(4)}$	2.46$^{(4)}$	2.20$^{(4)}$
4	1.46$^{(2)}$	1.61$^{(2)}$	1.37$^{(2)}$	1.30$^{(3)}$

Model 1 is the basic Jegadeesh (1991) model.
Model 2 is the time-varying coefficient model.
Model 3 is the time-varying coefficient with Markov-switching variance model.
Model 4 is the Markov-switching risk premium with learning model.
Root mean squared error is calculated for the period of January, 1987 + j to December, 1996, where j is the forecast horizon, using parameter estimates from Tables 1-4.
Comparison rankings for the models are given in parentheses.

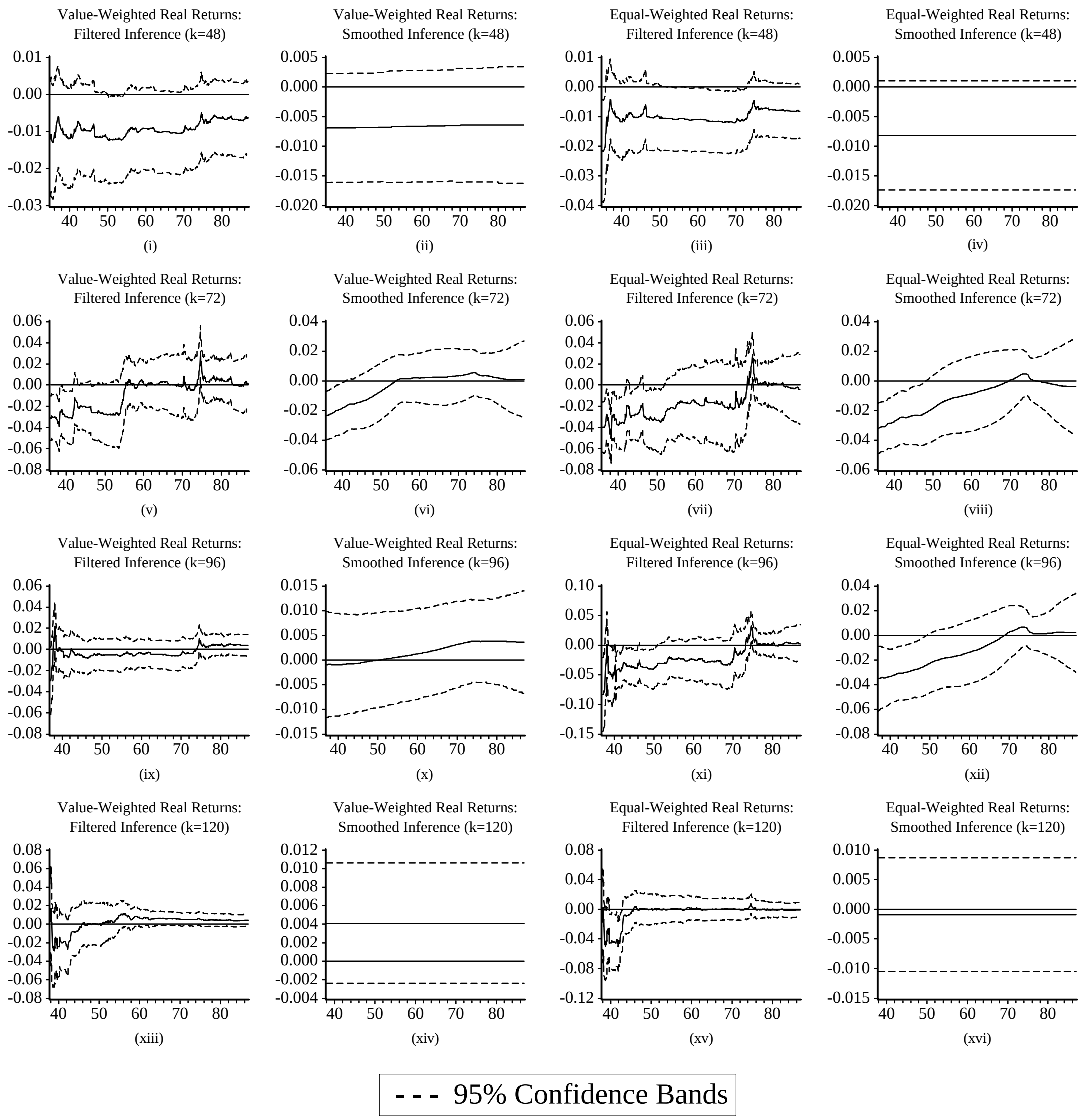


FIGURE 1.1: The Time-Varying Coefficient Model – Filtered and Smoothed Inferences about $b(k)$, 1926-86.

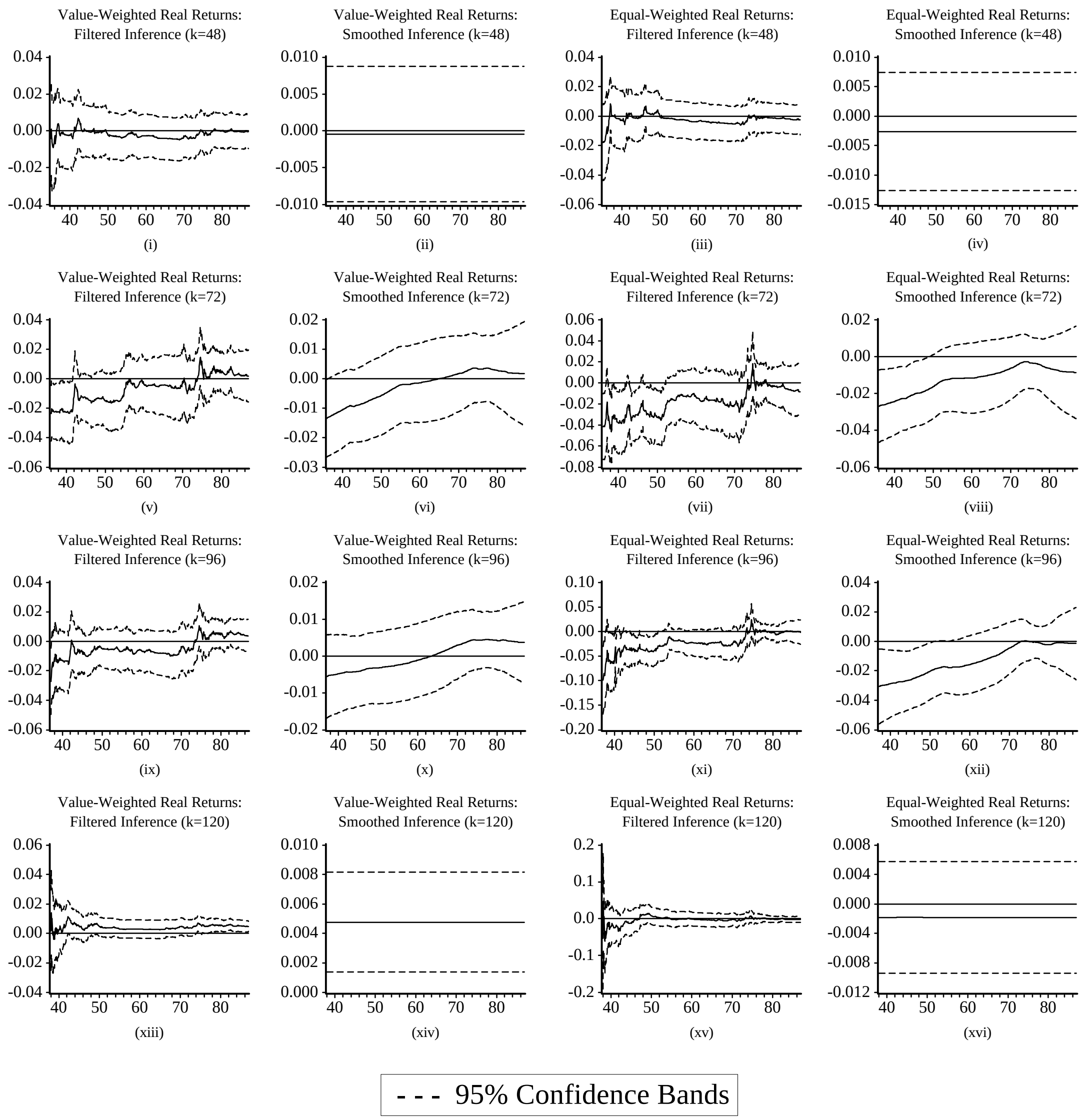


FIGURE 1.2: The Time-Varying Coefficient and Markov-Switching Variance Model – Filtered and Smoothed Inferences about $b(k)$, 1926-86.

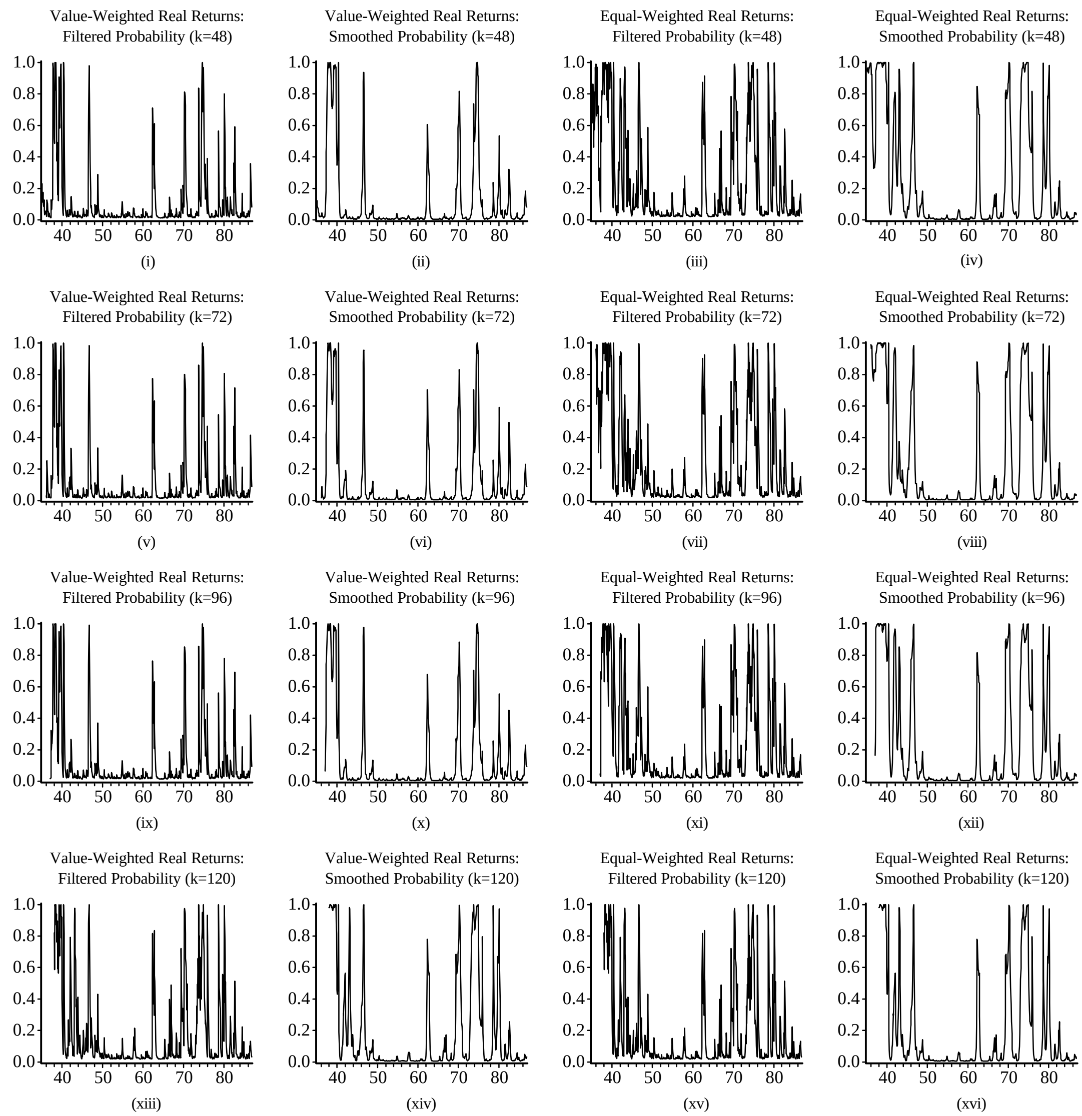


FIGURE 1.3: The Time-Varying Coefficient and Markov-Switching Variance Model – Filtered and Smoothed Probabilities of the High Variance Regime, 1926-86.

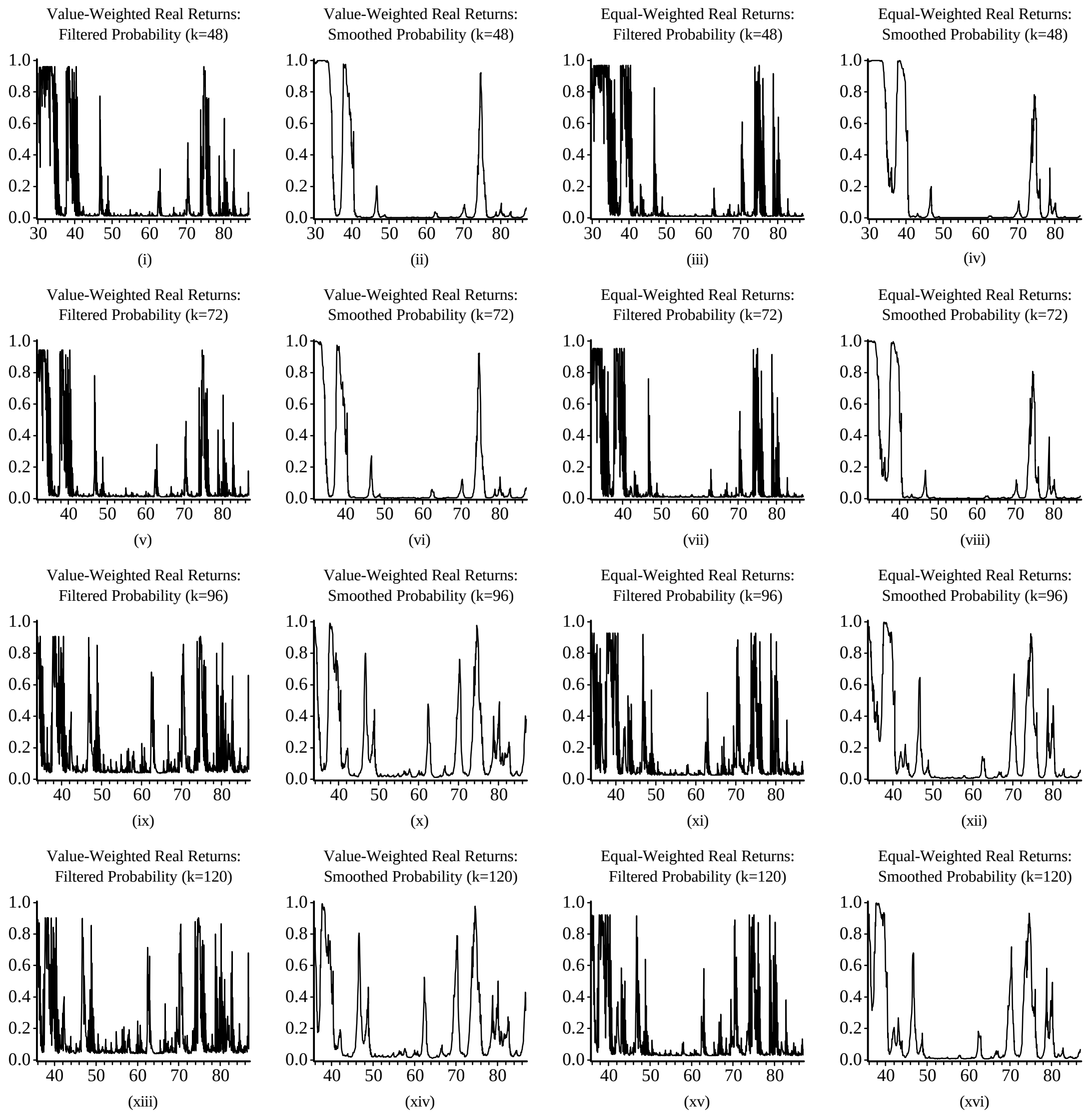


FIGURE 1.4: The Markov-Switching Risk Premium with Learning Model – Filtered and Smoothed Probabilities of the High Variance Regime, 1926-86.

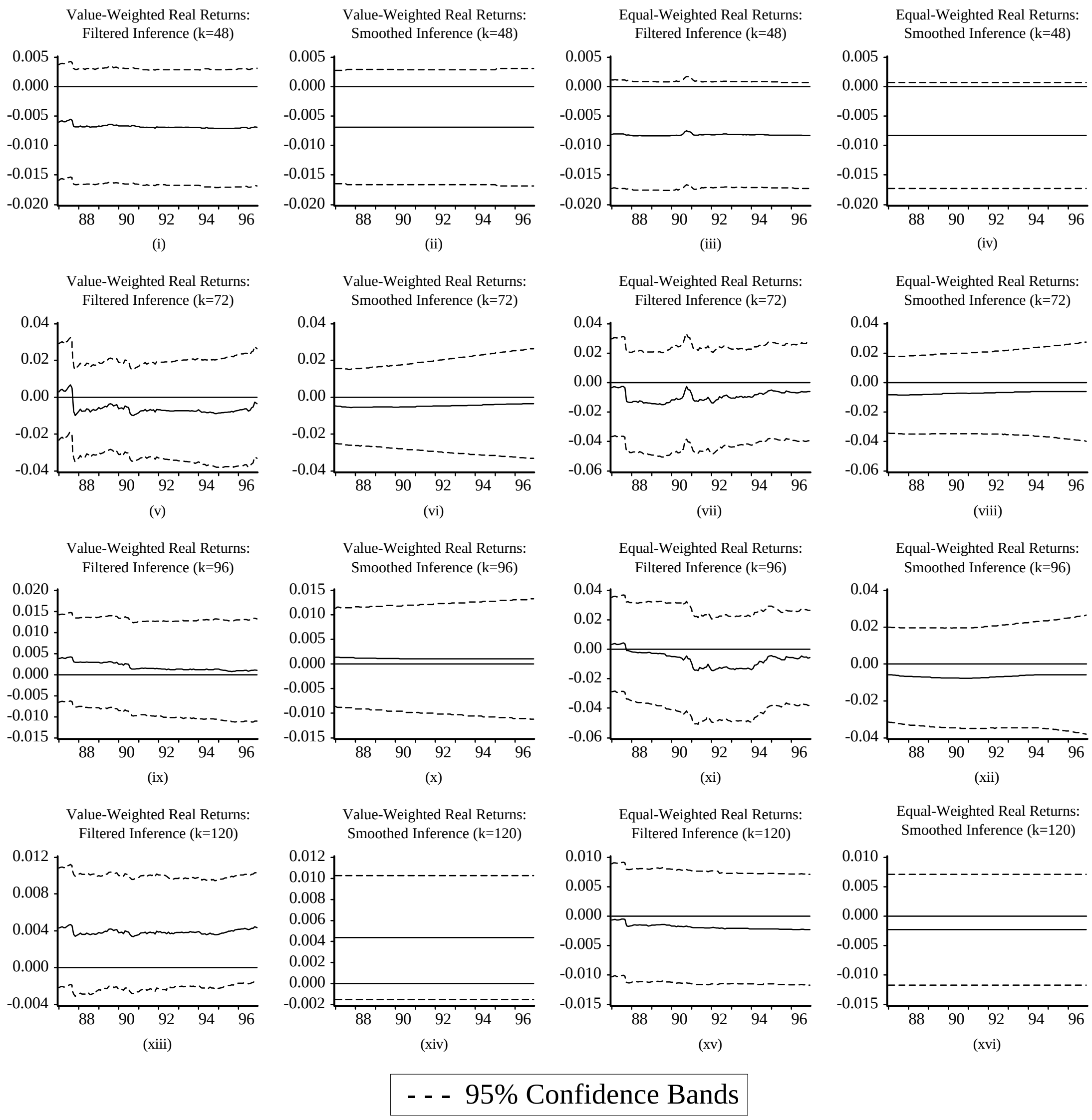


FIGURE 1.5: The Time-Varying Coefficient Model – Filtered and Smoothed Inferences about $b(k)$, 1987-96.

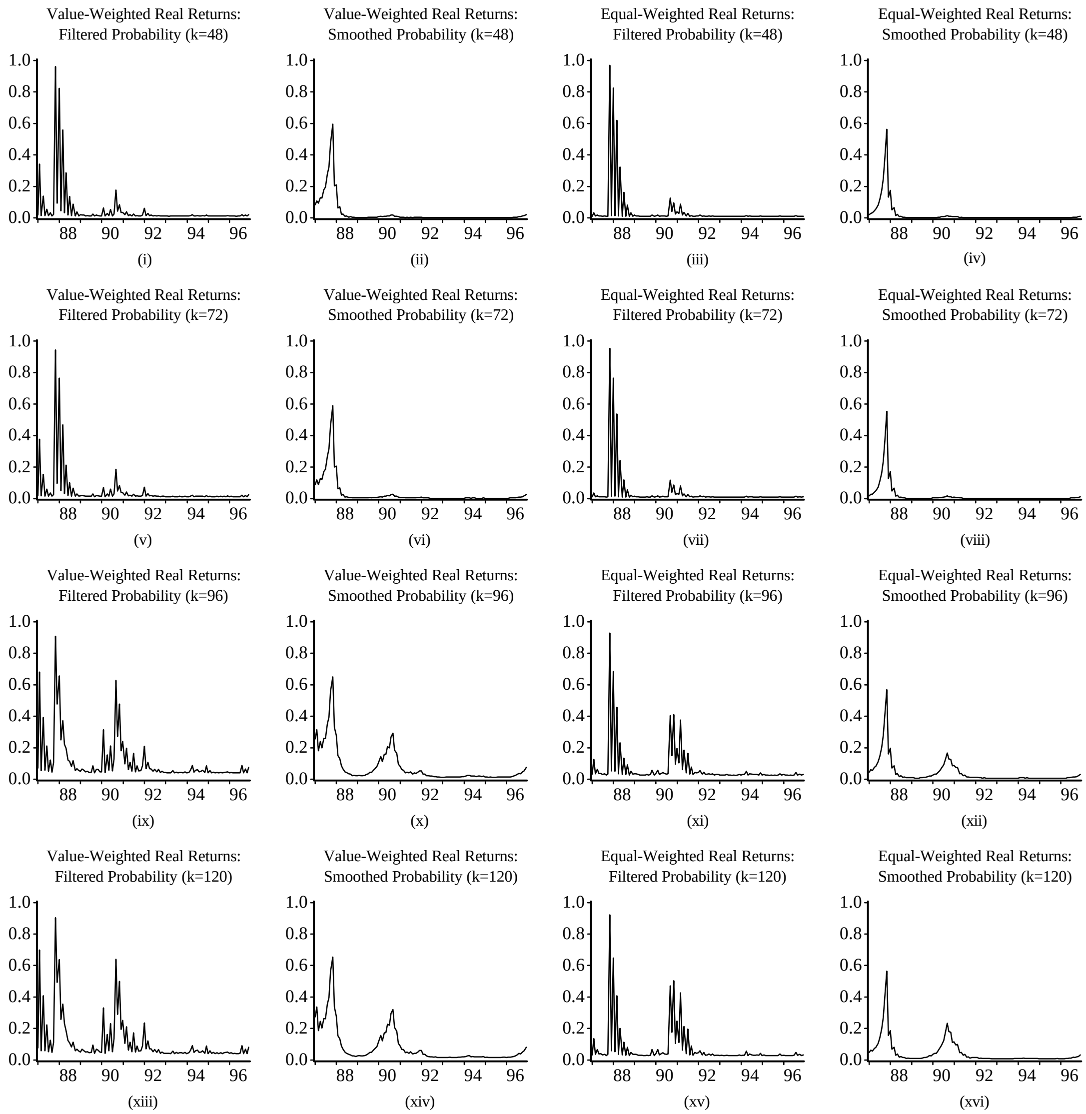
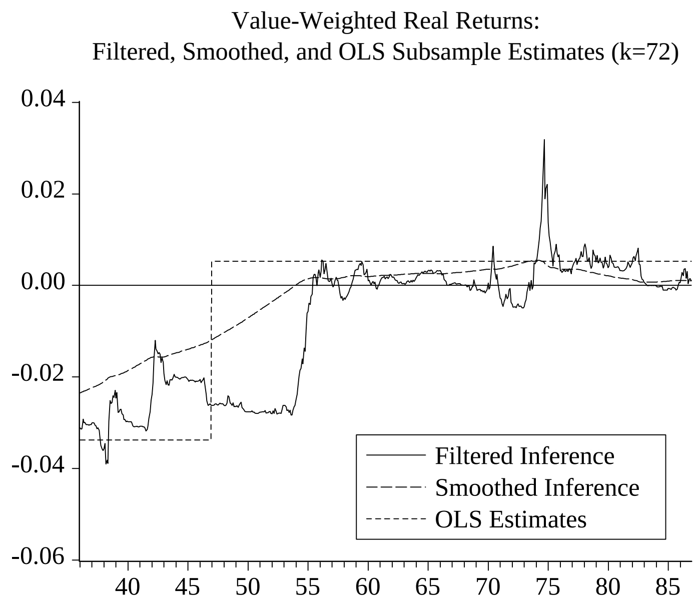
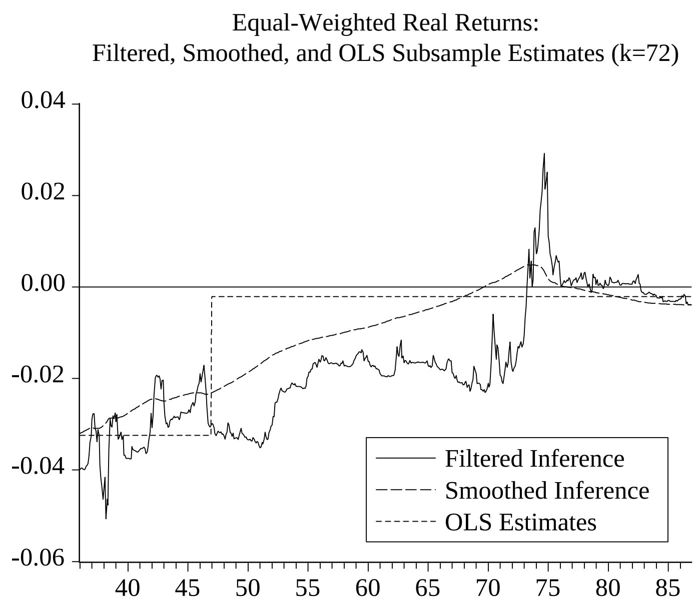


FIGURE 1.6: Markov-Switching Risk Premium with Learning Model – Filtered and Smoothed Probabilities of the High Variance Regime, 1987-96.



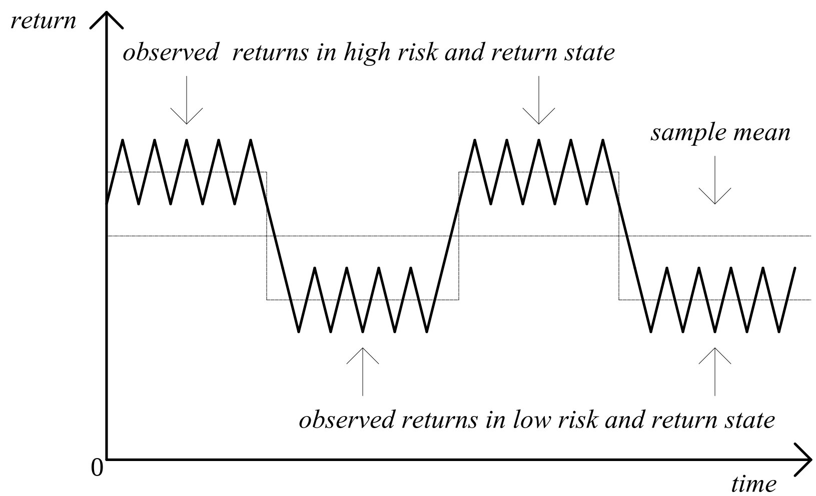
(i)



(ii)

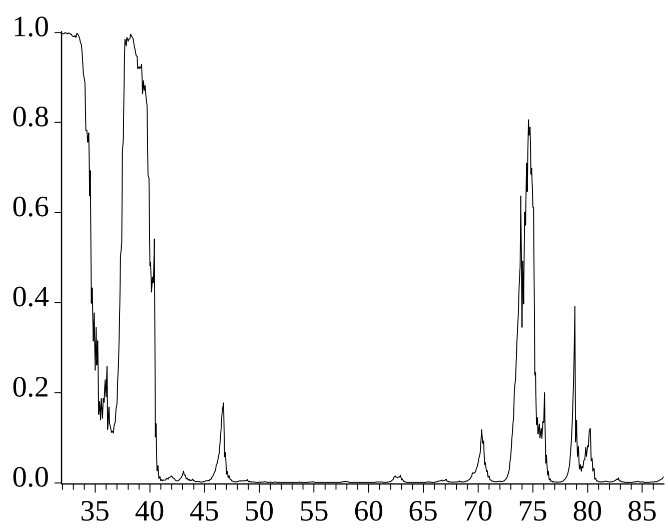
FIGURE 1.7: Comparison of Results – Time-Varying Coefficient Inferences and OLS Subsample Estimates of $b(k)$, 1926-86.

Long-Horizon Mean Reversion



(i)

Equal-Weighted Real Returns:
Smoothed Probability of High Mean and Variance State (k=72)



(ii)

FIGURE 1.8: Time-Varying Expected Returns and Mean Reversion – Theoretical Effects and Smoothed Probability, 1926-86.

CHAPTER 2: A Markov-Switching Model of Stock Returns with Learning and Volatility Feedback*

2.1 Introduction

A central issue in the modern empirical finance literature is the intertemporal relationship between market volatility and stock returns. Interest in this relationship stems from the idea that market volatility is the most easily motivated measure of aggregate risk. To the extent that it is a good measure of risk, and investors are rational and risk-averse, volatility should have a positive effect on the market risk premium, defined as expected excess returns on a market portfolio. Likewise, a positive price of risk should imply a negative effect of unexpected volatility on realized excess returns, due to 'volatility feedback.' That is, as investors learn about higher future volatility, the discount rate rises, putting immediate downward pressure on prices. In this chapter, we develop a Markov-switching model of stock returns to investigate the empirical evidence for these effects. In particular, we examine whether a model with learning and endogenous volatility feedback implies a positive intertemporal relationship between volatility and the market risk premium.

Previous studies have produced conflicting evidence on the intertemporal relationship between volatility and the market risk premium. In a seminal paper, French, Schwert, and Stambaugh (1987) find a positive relationship using basic time-series models, including a generalized autoregressive conditional heteroskedasticity in mean (GARCH-M) model inspired by Engle, Lilien, and Robins (1987). Likewise, Bollerslev, Engle, and Wooldrige (1988), Harvey (1989), and Campbell and Hentschel (1992) find a positive relationship. However, other empirical studies using similar data and methodologies have produced less favourable results. For example, Glosten, Jagannathan, and Runkle (1993) find a negative relationship between conditional expected excess returns and conditional variance. They employ a modified version of the GARCH-M model that incorporates the nominal interest rates in the calculation of the conditional

* This chapter contains materials first presented in Kim, Morley, and Nelson (1999).

variance. Similarly, Whitelaw (1994) finds a negative relationship between conditional expected returns and conditional variance calculated by generalized method of moments (GMM) estimation with multiple financial variable instruments. Other studies that find a negative relationship include Fama and Schwert (1977), Campbell (1987), Breen, Glosten, and Jagannathan (1989), and Nelson (1991). Furthermore, theoretical studies by Abel (1988), Backus and Gregory (1993), and Gennotte and Marsh (1993) have convincingly demonstrated that modern general equilibrium models of stock prices can imply a negative relationship between volatility and the market risk premium for some range of parameter values.

The empirical results have been more uniform for studies incorporating volatility feedback. Volatility feedback—originally proposed by Pindyck (1984) as an explanation for the lackluster performance of the stock market in the 1970s—is the idea that higher expected future volatility puts immediate negative pressure on stock prices. This effect should be easier to detect empirically than the effect of volatility on a contemporaneous expected excess return since it reflects the cumulative effect of all future volatility on discounted changes of future expected excess returns. Thus, when French, Schwert, and Stambaugh (1987) employ a two-step approach that regresses excess returns on errors from an integrated autoregressive moving average (ARIMA) model of volatility, they are able to find strong evidence for a negative volatility feedback effect. Similarly, Turner, Startz, and Nelson (1989) find a negative feedback effect using a special case of the Markov-switching model employed in this chapter. Campbell and Hentschel (1992) find a negative feedback effect using a quadratic generalized autoregressive conditional heteroskedasticity (QGARCH) model of volatility. But their paper is more notable for developing a theoretical model of volatility feedback. They show that, if dividend news is subject to a QGARCH process and there is a linear relationship between news volatility and expected excess returns, Campbell and Shiller's (1988a,b) log-linear present value model of stock prices endogenously generates volatility feedback. Furthermore, the explicit solution for the volatility feedback term implies that any empirical evidence of a negative volatility feedback effect is equivalent to evidence of a positive relationship between volatility and the market risk premium. More recently, Bakaert and Wu (1999) also find empirical support for a negative volatility feedback. They employ a model that compares the volatility feedback effect with the leverage effect proposed by Black (1976) and Christie (1982).

In this chapter, we extend Turner, Startz, and Nelson's (1989) model of stock returns to provide an alternative version of Campbell and Hentschel's (1992) volatility feedback model, with dividend news subject to a two-state Markov-switching variance process instead of a QGARCH process. Like Campbell and Hentschel, we find an endogenously generated volatility feedback term, with an explicit form that directly links negative feedback to a positive relationship between volatility and the market risk premium. Our approach has the advantage that the Markov-switching specification avoids the estimation complications, discussed in Campbell and Hentschel, arising from the quadratic relation between excess returns and dividend news for the QGARCH specification. Also, the Markov-switching specification is, arguably, a better model of stock market volatility at the monthly frequency looked at in this chapter. In any event, the Markov-switching specification provides us with flexibility in modeling the learning process for economic agents. We are able to consider a variety of specifications, including the original 'partial learning' specification given in Turner, Startz, and Nelson (1989) and an alternative 'full learning' specification that resolves the time inconsistency of the partial learning specification. In these cases, agents learn about a given volatility regime within a month. But we also consider the possibility that learning is more protracted to determine if there has been change in its speed at which agents acquire information since 1926.

Using CRSP data on excess returns for a value-weighted market portfolio, we find strong support for a positive relationship between volatility and the market risk premium, especially in the post-War period. We also find some evidence that the learning process is slower than that considered by Turner, Startz, and Nelson (1989). But, surprisingly, there is little evidence for a change in the speed of this process over the course of the 1926-96 sample period.

The rest of this chapter is organized as follows. Section 2.2 motivates and develops the Markov-switching model of stock returns with learning and volatility feedback. Section 2.3 reports estimates for the model using a variety of specifications for the learning process and monthly returns from the CRSP data file for the period of 1926-96. Section 2.4 concludes. Tables and figures follow the conclusion.

2.2 The Model

2.2.1 Background

Stock returns are related to prices by the following identity:

$$R_{t+1} \equiv \frac{P_{t+1} + D_{t+1}}{P_t} - 1, \tag{2.1}$$

where R_{t+1} denotes the return on a stock or portfolio held from time t to $t+1$, P_{t+1} is the (ex-dividend) price of a the stock or portfolio at the end of period $t+1$, and D_{t+1} is dividend at time $t+1$, claimed by the owner at the beginning of time $t+1$. Then, we can solve recursively for P_t and apply the expectations operator to see how prices reflect future dividends as follows:

$$P_t = \mathrm{E}\left[\sum_{i=1}^{\infty} \left\{ D_{t+i} \Big/ \prod_{j=1}^{i} (1 + R_{t+j}) \right\} \Big| y_t \right], \tag{2.1'}$$

where y_t is information available at time t and a transversality condition imposed to rule out the existence of rational bubbles in asset prices. Clearly, good news about future dividends has a direct positive effect on prices and returns, while bad news has a negative effect.

Following French, Schwert, and Stambaugh (1987) and Turner, Startz, and Nelson (1989), Campbell and Hentschel (1992) propose an additional ‘volatility feedback’ effect. To understand how this additional effect operates, it is useful to look at the log-linear approximate present value model developed by Campbell and Shiller (1988a,b) and Campbell (1991). In the log-linear framework, the present value relation equivalent to (1') is given as follows:

$$p_t = \frac{k}{1-r} + \mathrm{E}\left[\sum_{j=0}^{\infty} r^j [(1-r) d_{t+1+j} - r_{t+1+j}] \middle| y_t \right], \qquad (2.1'')$$

where lower-case letters denote log values and r and k are parameters of linearization (see Campbell and Shiller, 1988a). This model conveniently allows us to simultaneously examine the effects of changes in expected dividends and changes in future expected returns. Again, news about dividends would have an obvious direct effect on prices. But, as Campbell and Hentschel hypothesize, news about dividends could also have an indirect effect on prices through implied revisions in future expected returns. Specifically, they argue that news about dividends contains information about future volatility of news, with future expected returns assumed to be a positive function of that volatility. Thus, a persistent increase in volatility signaled by extreme news, whether good or bad, is hypothesized to increase future expected returns. This should dampen what would otherwise be a large positive return or accentuate what would otherwise be a large negative return due to a downward pressure on prices. Conversely, a period of calm signaled by little or no news is hypothesized to lower future expected returns. This should increase this period's return due to a positive pressure on prices. Hence, Campbell and Hentschel's claim that "no news is good news."

The immediate appeal of the volatility feedback model is that it potentially explains the presence of any negative skewness and excess kurtosis in stock returns, even if the underlying news about dividends is symmetric and conditionally Normal. It also provides a theoretical basis for the empirically supported assumption in Pindyck (1984), French, Schwert, and Stambaugh (1987) and Turner, Startz, and Nelson (1989) that unanticipated, but persistent, changes in volatility are negatively correlated with excess returns.

2.2.2 Volatility Feedback

Campbell and Hentschel's (1992) volatility feedback model is a partial equilibrium model of stock returns that relies on two simple assumptions. First, news about dividends is assumed to follow a QGARCH process. Second, expected returns are assumed to be a linear function of the conditional variance of news about future dividends. As mentioned in Campbell and Hentschel, the first assumption can be amended to allow news to follow a Markov-switching process. This is the approach we take in this chapter. In particular, we make the following two assumptions:

(i) news about dividends is subject to the following zero-mean, two-state Markov-switching variance process:

$$e_t \sim \mathrm{N}(0, s^2_{S_t}), \tag{2.2}$$

$$s^2_{S_t} = s^2_0(1 - S_t) + s^2_1 S_t,$$

$$s^2_0 < s^2_1,$$

$$\Pr[S_t = 0 | S_{t-1} = 0] = q,$$

$$\Pr[S_t = 1 | S_{t-1} = 1] = p,$$

where e_t denotes new information about dividends that arrives during trading period t, $s^2_{S_t}$ is the variance of e_t, S_t is a Markov-switching state variable that takes on discrete values of zero or one according to the prevailing volatility regime, and q and p are the transition probabilities governing the evolution of S_t;

(ii) expected returns are a linear function of market perceptions—formed rationally in the sense of Muth (1960)—about the volatility of news about dividends:

$$\mathrm{E}[r_t | y_t] = m_0 + m_1 \mathrm{P}r[S_t = 1 | y_t], \tag{2.3}$$

where m_0 and m_1 are assumed to be positive, reflecting a positive relationship between volatility and expected returns.

Numerous studies have used a Markov-switching variance assumption to model stock returns (Schwert, 1989b, Turner, Startz, and Nelson, 1989, Hamilton and Susmel, 1994, Schaller and van Norden, 1997, Kim, Nelson, and Startz, 1998, and Kim and Nelson, 1998). In terms of motivating its use here, Hamilton and Susmel's findings are of particular interest. They apply a Markov-switching autoregressive conditional heteroskedasticity (SWARCH) model to weekly stock returns and find that, once ARCH parameters are allowed to switch between regimes according to an unobserved Markov-switching state variable, estimated ARCH effects are much less persistent than for a standard ARCH model. Indeed, their results imply that monthly stock returns can be reasonably modeled as having only a Markov-switching conditional variance. Furthermore, it turns out that Markov switching avoids estimation complications associated with a nonlinear relation between excess returns and dividend news for a QGARCH specification. Instead, estimation is always a straightforward application of the procedure discussed in Hamilton (1989).

Following Campbell and Hentschel (1992), we can manipulate the log-linear present value model given in (1'') to show that returns have three components reflecting expected returns, news, and a volatility feedback effect:

$$r_t = \mathrm{E}[r_t | y_t^b] + e_t - f_t , \tag{2.4}$$

where

$$e_t \equiv \mathrm{E}\left[\sum_{j=0}^{\infty} r^j \Delta d_{t+j} \middle| y_t^e\right] - \mathrm{E}\left[\sum_{j=0}^{\infty} r^j \Delta d_{t+j} \middle| y_t^b\right]$$

denotes news about dividends (y_t^b is information available at the beginning of the trading period, while y_t^e is information available by the end of the trading period) and

$$f_t \equiv \mathrm{E}\left[\sum_{j=1}^{\infty} r^j r_{t+j} \middle| y_t^e\right] - \mathrm{E}\left[\sum_{j=1}^{\infty} r^j r_{t+j} \middle| y_t^b\right]$$

denotes revisions in future expected returns, and turns out to be the volatility feedback term.

The assumptions, given in (2) and (3), directly provide us with expressions for the first two components of (4). However, we need to solve for the third component, f_t. Note that expected returns at some arbitrary point in the future are given by

$$\mathrm{E}[r_{t+j} | y_t] = m_0 + m_1 \Pr[S_t = 1] + m_1 l^j (\Pr[S_t = 1 | y_t] - \Pr[S_t = 1]) , \tag{2.5}$$

where $l \equiv p + q - 1$ (see Hamilton, 1989). This expression can be used to show that the discounted sum of future expected returns is

$$\mathrm{E}\left[\sum_{j=1}^{\infty} r^j r_{t+j} \middle| y_t\right] = \frac{m_0}{1-r} + \frac{m_1}{1-r} \Pr[S_t = 1] + \frac{m_1}{1-r\,l} (\Pr[S_t = 1 | y_t] - \Pr[S_t = 1]) , \tag{2.6}$$

which, in turn, allows us to solve for f_t:

$$f_t = \frac{m_1}{1-r\,l} (\Pr[S_t = 1 | y_t^e] - \Pr[S_t = 1 | y_t^b]) . \tag{2.7}$$

Given these specific expressions for the three components of equation (2.4), we can write the model of the effects of Markov-switching volatility on stock returns as follows:

$$r_t = m_0 + m_1 \Pr[S_t = 1 | y_t^b] + d\{\Pr[S_t = 1 | y_t^e] - \Pr[S_t = 1 | y_t^b]\} + e_t, \tag{2.4'}$$

where

$$d \equiv -\frac{m_1}{1 - r l}. \tag{2.8}$$

Thus, a positive relationship between volatility and expected returns directly implies that, as long as volatility is persistent (i.e., $p + q > 1$), the coefficient on the volatility feedback term, d, is negative and vice versa. Note that the parameter of linearization, r, which is the average ratio of the stock price to the sum of the stock price and the dividend, is slightly less than one (0.997) in practice.

2.2.3 Learning

Before we proceed to estimation of the model, we need to specify exactly how economic agents acquire information over time. First, the assumption of rational expectations implies that both y_t^b and y_t^e should contain information available to econometricians, including past returns ($r_{t-1}, r_{t-2,} \ldots$). Second, a volatility feedback effect only exists if the information available to agents changes within trading periods—that is, y_t^e contains information not in y_t^b. Thus, we can only generate a volatility feedback effect by assuming that agents acquire some information about current and past states ($S_t, S_{t-1}, \ldots$) not directly available to econometricians.

We consider the following information specifications. First, we assume that there is no learning within a trading period—i.e., $y_t^b = y_t^e = y_t$. Initially, we also assume that expected excess returns are constant over time ($m_1 = 0$) to demonstrate that evidence of Markov-switching volatility is not a consequence of a switching mean. Then, we allow expected excess returns to switch according to the volatility regime, with agents observing past returns ($y_t = r_{t-1}, r_{t-2}, \ldots$) and the true state ($y_t = S_t$), respectively. These two specifications provide benchmarks polar cases for information available to agents given rational expectations. Second, we assume that learning occurs—i.e., $y_t^b \neq y_t^e$. Initially, we assume that there is only partial learning, as in Turner, Startz, and Nelson (1989). That is, $y_t^b = r_{t-1}, r_{t-2}, \ldots$ and $y_t^e = S_t$, where the true regime S_t is a proxy for what is actually known by agents at the end of time *t*. We consider both the case where d is freely estimated and where it is restricted according to equation (2.8). Then, we assume there is full learning. That is, $y_t^b = S_{t-1}$ and $y_t^e = S_t$. Again, we consider both cases for d. Finally, we assume that there is generalized partial learning. That is, we let estimation determine what information $(r_{t-1}, r_{t-2}, \ldots, S_t, S_{t-1}, \ldots)$ economic agents act upon. We estimate all model specifications using the filter for Markov switching models discussed in Hamilton (1989).

For generalized partial learning a practical complication arises that requires further discussion. To understand this complication, note that, if agents fully observed the state S_t at time *t*, knowledge about any prior state would be redundant due to the Markovian property of the state variable. That is,

$$\Pr[S_t = 1 | S_t, S_{t-1}, \ldots, r_{t-1}, r_{t-2}, \ldots] = S_t. \tag{2.9}$$

But, if available information is only approximated by a set of states (that is, learning occurs only partially over *k* periods), then the implied redundancy may not be true. The problem, then, is determining what alternative expression to use for the conditional probability in this case. We propose the following:

$$\begin{aligned} \Pr[S_t = 1 | S_t, S_{t-1}, \ldots, S_{t-K}, r_{t-1}, r_{t-2}, \ldots] &\approx a_r \Pr[S_t = 1 | r_{t-1}, r_{t-2}, \ldots] \\ &+ \sum\nolimits_k a_k \Pr[S_t = 1 | S_{t-k}, r_{t-1}, r_{t-2}, \ldots], \\ k = 0,1,\ldots,K, \end{aligned} \tag{2.10}$$

where the a 's are coefficients of linear combination. While this expression is clearly an approximation, it should be an improvement over using equation (2.9).

Then, the resulting regression equation for the case where information about more than one state is acted upon is given by

$$r_t = p_c + p_r \Pr[S_t = 1 | r_{t-1}, r_{t-2}, \ldots] + \sum_k p_k \Pr[S_t = 1 | y_t^k] + e_t\,, \qquad (2.4'')$$
$$k = 0,1,\ldots,K,,$$

where the p 's are functions of m_0, m_1, d, and the a_k 's. Note that, while this model can tell us what information is relevant to economic agents, it is a reduced-form model in the sense that it cannot provide us with economically interpretable parameters.

2.3 Empirical Results

2.3.1 Data

In our empirical exercises, we work with excess stock returns for a market portfolio. In particular, we employ continuously compounded monthly total returns for a value-weighted portfolio of all NYSE-listed stocks in excess of continuously compounded one-month U.S. Treasury bill yields. The data are drawn from the CRSP data files for the sample period of January 1926 to December 1996. Total returns represent capital gains plus dividend yields. Continuously compounded returns are calculated by taking natural logarithms of simple gross returns.

The use of excess returns means that ‘news’ refers to future dividends relative to future interest rates. A relative measure of this kind makes sense since the theoretical effects of volatility on real returns alone are ambiguous, even if we assume a positive relationship between volatility and the market risk premium. In particular, an increase in risk would cause investors to substitute away from riskier assets, putting downward pressure on interest rates.

For the empirical exercises, we split the sample into two subperiods: 1926-51 and 1952-96. The breakpoint corresponds to the Fed-Treasury Accord and is also used in

Campbell and Hentschel (1992). The reasons for looking at these periods separately are twofold. First, we are most interested in the more recent experience. Using the full data sample may produce misleading implications about the effects of volatility on stock returns if the behaviour of volatility has changed dramatically since the Great Depression, and there is considerable reason to believe that this behaviour has changed (see Pagan and Schwert, 1990). As we shall see, pre-War excess returns underwent episodes in which volatility was considerably higher and more persistent than ever occurred in the post-War period, suggesting the need for a more complicated model of volatility than is employed in this chapter. Kim, Nelson, and Startz (1998) and Kim and Nelson (1998) provide such a model, which allows for three volatility regimes. Nevertheless, their findings suggest that the two-regime model employed here and in Turner, Startz, and Nelson (1989) is sufficient to describe post-War returns. Properly accounting for pre-War volatility behaviour in a model of returns for the entire sample, then, would unnecessarily complicate analysis, without producing any additional insights into the more recent experience. Second, we are also interested in determining how those effects have changed over time. By comparing the results for the two subsamples, we hope to provide an empirical answer to the question of whether the speed at which agents learn about volatility regimes has changed over time.

2.3.2 Markov-Switching Volatility

Table 2.1 reports maximum likelihood estimates for constant and switching variance models of stock returns for the sample periods of 1926-51 and 1952-96.[1] For both samples, there is a huge improvement in the log likelihood values for the Markov-switching specification. For the 1926-51 sample, the likelihood ratio statistic is 145.94254. For the 1952-96 sample, the likelihood ratio statistic is 47.04434. While the distribution of these test statistics is non-standard since the transition probabilities q and p

[1] All maximum likelihood estimation was conducted using the OPTMUM procedure for the GAUSS programming language. Numerical derivatives were used in estimation, as well as for calculation of asymptotic standard errors. Parameters were appropriately constrained (e.g., variances were constrained to be non-negative). Inferences appear robust to a variety of starting values.

are not identified under the null hypothesis $H_0 : s_0 = s_1$ of no Markov switching (see Hansen, 1992, and Garcia, 1995), their values far exceed asymptotic critical values reported in Garcia (1995). Therefore, we can be confident about the presence of Markov-switching volatility in stock returns.

The estimates also confirm our claim that pre-War returns underwent episodes of higher and more persistent volatility than in the post-War period. In particular, note that the standard deviation of returns in the high volatility regime is 0.12086 in the 1926-51 sample period versus 0.05999 in the 1952-96 sample period. Likewise, the probability of staying in the high volatility regime is 0.96050 versus 0.90807, corresponding to an expected duration of about 25 months versus 11 months. The numbers are also implicit in Figure 2.1, which displays excess returns and smoothed probabilities of a high volatility regime for the sample periods of 1926-51 and 1952-96. First, note the scales for returns across the two sample periods. Pre-War returns at their most volatile are about twice as volatile as post-War returns. Second, note the greater persistence of pre-War episodes of high volatility. These findings all suggest that the separation of the entire 1926-96 sample period into the two subsamples is appropriate, although they do not pinpoint with any great precision the exact date at which the behaviour changed. The Fed-Treasury Accord merely provides a convenient dividing line.

Figure 2.1 suggests a relationship between periods of high volatility and NBER-dated recessions. The onset of a recession often corresponds to the onset of a period of high volatility, although episodes of high volatility generally appear to last longer than recessions. The shading makes it clear that the onset of recessions is typically associated with large negative returns and the ends of recessions, when output is still contracting, are often associated with large positive returns. Given the association of recessions with episodes of market volatility, this timing fits well with the story behind volatility feedback. In particular, an unexpected onset of a recession and high market volatility drives stock prices down, while knowledge that a recession and the associated high market volatility are near an end drives stock prices up.

But, at this point, we have not explicitly modeled or estimated any interaction between volatility and returns implied by volatility feedback. Instead, we have assumed that mean stock returns remain constant within the subsamples. Even so, the results in Table 2.1 do have an important implication for the different bahaviour of returns in the two regimes. In particular, accounting for Markov-switching heteroskedasticity effectively weighs the high volatility returns less than the low volatility returns in the estimation of the mean stock return within the subsamples. Since the estimates of the mean return increase substantially when Markov-switching heteroskedasticity is accounted for—from 0.00520 to 0.01201 for the 1926-51 subsample and from 0.00508 to 0.00695 for the 1952-96 subsample—the implication is that the high volatility returns are, on average, less than the low volatility returns. The most immediate explanation for this result is that there is actually a negative relationship between volatility and the market risk premium. However, we spend the rest of this chapter examining and challenging the robustness of this explanation.

2.3.3 Volatility and the Market Risk Premium

Table 2.2 reports maximum likelihood estimates for the model of the effects of Markov-switching volatility on stock returns with benchmark specifications for information about volatility available to economic agents. While the estimated relationship between volatility and the market risk premium appears to be consistently negative for pre-War returns, it depends crucially on the information specification for post-War returns. In particular, the two specifications, representing polar cases of information availability in the presence of rational expectations, produce conflicting estimates of the sign for post-War excess returns. For the first specification, in which agents are uncertain about the volatility regime and are assumed to observe only what econometricians observe, higher expected volatility is estimated to increase the market risk premium—i.e., $m_1 > 0$—although this effect is not statistically significant at traditional levels. Conversely, for the second specification, in which agents are certain about the true volatility regime, there is a statistically significant negative relationship—

i.e., $m_1 < 0$. Such conflicting findings make it clear that we will have to take a stand on the information available to agents before we can make any strong conclusions about the effects of volatility on returns.

A question then arises as how best to determine what information is available to economic agents. A reasonable criterion is statistical likelihood. On this count, the second specification does better for both sample periods—the comparable log likelihood values are 448.41838 and 450.40496 for the 1926-51 subsample, while they are 987.17438 and 989.57509 for the 1952-96 subsample—although it should be noted that the two specifications are not nested. Also, Turner, Startz, and Nelson (1989) find the exact opposite result for post-War excess returns given the same specifications and only slightly different data and sample period.[2]

A specification that does nest the two cases is the benchmark volatility feedback specification first presented in Turner, Startz, and Nelson (1989). This specification recognizes that information available to economic agents about a given volatility regime is likely to change throughout a monthly trading period, generating a volatility feedback effect. However, the volatility feedback coefficient is not restricted according to equation (2.8) in this specification, but is freely estimated. The learning process for this specification is given as follows: at the beginning of the trading period agents observe only what econometricians observe, as in the first specification above; but by the end of the trading period they partially observe the true volatility regime, as in the second specification above.

Table 2.3 reports maximum likelihood estimates for the model of the effects of Markov-switching volatility on stock returns with feedback in the presence of partial learning. For the 1926-51, allowing for a volatility feedback effect as in Turner, Startz, and Nelson (1989) has little impact on the log likelihood value. But, for the 1952-96 subsample, allowing for a volatility feedback effect produces a huge improvement in the log likelihood value over both of the previous specifications. For this sample, the

---

[2] Turner, Startz, and Nelson (1989) use excess returns from the Standard and Poor's composite index for the sample period of January 1946 to December 1987.

likelihood ratio statistic for the null hypothesis $H_0\!:\!d = 0$ that agents observe only what econometricians observe is 21.36604, while the likelihood ratio statistic for the null hypothesis $H_0\!:\!m_1 = 0$ that agents always observe the true regime is 16.56462. Both statistics should have conventional asymptotic distributions and are, therefore, highly significant, suggesting the presence of a volatility feedback effect in post-War excess returns.

The actual estimated direct effect of volatility on market risk premium for the benchmark volatility feedback model is negative for the 1926-51 sample, while it is positive, but insignificant, for the 1952-96 sample. This latter result contrasts with Turner, Startz, and Nelson (1989) finding of a negative, but insignificant, direct effect. However, the more notable finding, also found by French, Schwert, and Stambaugh (1987), Turner, Startz, and Nelson, and Campbell and Hentschel (1992), is a significantly negative volatility feedback effect for both sample periods. This finding provides strong evidence for a positive relationship between volatility and the market risk premium. For instance, it is most obviously consistent with the story that an upward revision in future expected volatility leads to an upward revision in future expected excess returns and, consequently, a downward revision in this period's return due to a negative effect on prices. It also explains how it is possible for the effect of volatility on the market risk premium to be positive, as it is for the first specification above, even though the average excess return is lower in periods of high volatility, as we found in Table 2.1.

The results in Table 2.3 for the partial learning model with the feedback parameter restricted according to theory are more notable. In particular, for both sample periods, we find evidence of a positive relationship between volatility and the market risk premium, although the evidence is not statistically significant for the 1926-51 sample. These results support our conjecture about the implications of our findings for the model in which the feedback effect is freely estimated. However, the result for the pre-War sample should be treated with some caution since the likelihood ratio statistic for testing the restriction given in equation (2.8) is 3.22134, with a corresponding p-value of 0.07286. On the other hand, the comparable likelihood ratio statistic for the post-War period is only 0.27068, with a corresponding p-value of 0.60288. Therefore, we can be reasonably confident in the restricted volatility feedback model, as well as the finding of a positive relationship between volatility and the market risk premium, for post-War returns.

There remains an unresolved issue arising from the apparent inconsistency in the evolution of information available to economic agents. In particular, if agents have full knowledge about the true volatility regime by the end of a trading period, then they implicitly have information about the true volatility regime at the beginning of the next period beyond what is observed by econometricians. A plausible explanation for such a time-inconsistent learning process is that agents do not literally observe the true volatility regime, but observe information for which the true regime can serve as an imperfect, but reasonable, proxy. This is the explanation that underlies our partial learning specification. However, another solution is to this inconsistency is to allow for full learning by assuming that agents observe the previous regime at the beginning of a trading period.

Table 2.4 reports maximum likelihood estimates for the Markov-switching model of stock returns with volatility feedback in the presence of full learning. The results are broadly similar to those for partial learning. When the feedback parameter is freely estimated, the estimated direct effect of volatility on the market risk premium is negative for both samples, although it is only significant for the 1926-51 sample. However, the feedback parameter is negative and significant for both samples. When the feedback parameter is restricted by theory, the estimated relationship between volatility and the market risk premium is positive for both samples. It is statistically significant for the 1952-96 sample, although it appears that the restriction can be rejected for both samples. The likelihood ratio statistics for the 1926-51 and 1052-96 samples are 3.96186, with a corresponding p-value of 0.04654, and 6.86248, with a corresponding p-value of 0.00880, respectively.

Finally, we consider a generalized partial learning specification of the Markov-switching model that allows us to determine what information agents act upon. The generalized partial learning specification incorporates both past returns and states in the model of returns. In addition to allowing us to determine what information is most important, the generalized partial learning specification allows us to determine how fast learning has occurred in both subsamples.

Table 2.5 reports maximum likelihood estimates for the generalized partial learning specifications of the Markov-switching model of stock returns. For both sample periods, the significance of information contained in the one-month lagged state is unclear. The likelihood ratio statistics are 1.58560 and 1.71786, with corresponding p-values of 0.20796 and 0.18997, respectively, but the t-statistics are 2.26681 and 2.53248, with corresponding p-values of 0.01204 and 0.00580, respectively. Meanwhile, the information contained in the two-month lagged state is not significant. The likelihood ratio statistics are 0.01590 and 1.13206, with corresponding p-values of 0.89966 and 0.28734, respectively, and the t-statistics are 0.15585 and 1.24280, with corresponding p-values of 0.43813 and 0.10724, respectively. But, these findings are more notable because they suggest that the speed at which agents learn about volatility regimes has changed over time. Specifically, it appears that, throughout the entire 1926-96 sample period, it has taken possibly more than one month, but almost certainly less than two months for agents to acquire all relevant information about a given volatility regime.

Note that the generalized partial learning model with learning occurring within two months nests the full learning model reported in Table 2.4. For the 1926-51 sample, we cannot reject the full learning model. The likelihood ratio statistic is 0.02170, with corresponding p-value of 0.88289, and the t-statistic is 0.12710, with corresponding p-value of 0.44947. But, for the 1952-96 sample, we can reject the full learning model. The likelihood ratio statistic is 3.85396, with corresponding p-value of 0.04963, and the t-statistic is 1.76563, with corresponding p-value of 0.03901.

Figure 2.2 displays excess returns and smoothed probabilities of a high volatility regime from the second generalized learning specifications in Table 2.5. This figure provides an informal check on the notion that volatility is the exogenous force driving changes in the market risk premium. Indeed, the similarities between this figure and the previous one supports the idea that changes in the volatility regime are exogenous, with the market risk premium and returns reacting to such changes. The differences between this figure and the previous one are subtle, but there appears to be an even stronger relationship between periods of high volatility and NBER-dated recessions. In particular, the occurrence of high volatility outside of recessions is rare, and when it does occur, it is typically very short-lived.

2.4 Conclusions

In summary, our findings for the volatility feedback specifications provide considerable support for a positive relationship between volatility and the market risk premium, but little support for a change in the speed at which economic agents acquire information about volatility. The positive relationship between volatility and the market risk premium is particularly evident in the post-War period, but even holds in the pre-War period when we estimate the volatility feedback model with parameters restricted by theory.

These results are similar to those reported in Campbell and Hentschel (1992), but estimation of our model is easier to implement. Campbell and Hentschel's QGARCH assumption implies that returns are a complicated nonlinear function of the contemporaneous error term since its square shows up in the volatility feedback term. This nonlinearity makes maximum likelihood estimation difficult. But, for the Markov-switching assumption, returns depend linearly on the error term since the volatility feeback term incorporates the underlying state variable, not the realized error term. Thus, estimation is a straightforward application of maximum likelihood estimation using the filter discussed in Hamilton (1989).

Finally, the apparent relationship between market volatility and the business cycle is notable, especially since our model only employs stock return data. Future research may include further examination of this relationship, as in the classic studies by Schwert (1989a,b).

TABLE 2.1 – Part 1: Stock Market Volatility – Constant and Markov-Switching Variance Specifications, 1926-51

	Model Specification	
Parameters	Constant Variance $\{s_0 = s_1\}$	Switching Variance $\{s_0 < s_1\}$
m	0.00520	0.01201
	(0.00412)	(0.00268)
s_0	0.07273	0.03927
	(0.00291)	(0.00210)
s_1	-	0.12086
	-	(0.01058)
q	-	0.98597
	-	(0.01058)
p	-	0.96050
	-	(0.02926)
Log Likelihood	375.05524	448.02651

Maximum likelihood estimates are calculated using continuously compounded total monthly value-weighted excess returns of all NYSE-listed stocks for the sample period of January 1926 to December 1951. Asymptotic standard errors are reported in parentheses. The model has the following form:

$$r_t = m + e_t,$$

where

$$e_t | S_t \sim \mathrm{N}(0, s^2_t),$$
$$s^2_t = (1-S_t)s^2_0 + S_t s^2_1,$$
$$\Pr[S_t = 0 | S_{t-1} = 0] = q \text{, and } \Pr[S_t = 1 | S_{t-1} = 1] = p.$$

TABLE 2.1 – Part 2: Stock Market Volatility – Constant and Markov-Switching Variance Specifications, 1952-96

	Model Specification	
Parameters	Constant Variance $\{s_0 = s_1\}$	Switching Variance $\{s_0 < s_1\}$
m	0.00508	0.00695
	(0.00176)	(0.00161)
s_0	0.04070	0.03155
	(0.00124)	(0.00175)
s_1	-	0.05999
	-	(0.00665)
q	-	0.96922
	-	(0.01760)
p	-	0.90807
	-	(0.06763)
Log Likelihood	962.57177	986.09394

Maximum likelihood estimates are calculated using continuously compounded total monthly value-weighted excess returns of all NYSE-listed stocks for the sample period of January 1952 to December 1996. Asymptotic standard errors are reported in parentheses. The model has the following form:

$$r_t = m + e_t,$$

where

$$e_t | S_t \sim \mathrm{N}(0, s^2_t),$$
$$s^2_t = (1-S_t) s^2_0 + S_t s^2_1,$$
$$\Pr[S_t = 0 | S_{t-1} = 0] = q, \text{ and } \Pr[S_t = 1 | S_{t-1} = 1] = p.$$

TABLE 2.2 – Part 1: Volatility and the Market Risk Premium – Benchmark Specifications, 1926-51

	Model Specification	
Parameters	Agents observe past returns $\{y_t = r_{t-1}, r_{t-2}, \ldots\}$	Agents observe true state $\{y_t = S_t\}$
m_0	0.01318	0.01353
	(0.00300)	(0.00278)
m_1	-0.01164	-0.02826
	(0.01289)	(0.01296)
s_0	0.03928	0.03900
	(0.00206)	(0.00212)
s_1	0.11886	0.11721
	(0.01007)	(0.01005)
q	0.98704	0.98533
	(0.00954)	(0.01152)
p	0.96714	0.95977
	(0.02587)	(0.03085)
Log Likelihood	448.41838	450.40496

Maximum likelihood estimates are calculated using continuously compounded total monthly value-weighted excess returns of all NYSE-listed stocks for the sample period of January 1926 to December 1951. Asymptotic standard errors are reported in parentheses. The model has the following general form:

$$r_t = m_0 + m_1 \Pr[S_t = 1 | y_t] + e_t,$$

where

$$e_t | S_t \sim \mathrm{N}(0, s^2_t),$$
$$s^2_t = (1 - S_t) s^2_0 + S_t s^2_1, \; s^2_0 < s^2_1,$$
$$\Pr[S_t = 0 | S_{t-1} = 0] = q, \text{ and } \Pr[S_t = 1 | S_{t-1} = 1] = p.$$

TABLE 2.2 – Part 2: Volatility and the Market Risk Premium – Benchmark Specifications, 1952-96

	Model Specification	
Parameters	Agents observe past returns $\{y_t = r_{t-1}, r_{t-2}, \ldots\}$	Agents observe true state $\{y_t = S_t\}$
m_0	0.00456	0.00956
	(0.00242)	(0.00194)
m_1	0.01550	-0.01955
	(0.02138)	(0.00918)
s_0	0.03253	0.03143
	(0.00352)	(0.00191)
s_1	0.06445	0.06007
	(0.01853)	(0.00636)
q	0.97172	0.95681
	(0.01583)	(0.02384)
p	0.88038	0.85317
	(0.15691)	(0.09904)
Log Likelihood	987.17438	989.57509

Maximum likelihood estimates are calculated using continuously compounded total monthly value-weighted excess returns of all NYSE-listed stocks for the sample period of January 1952 to December 1996. Asymptotic standard errors are reported in parentheses. The model has the following general form:

$$r_t = m_0 + m_1 \Pr[S_t = 1 | y_t] + e_t,$$

where

$$e_t | S_t \sim N(0, s^2_t),$$
$$s^2_t = (1 - S_t) s^2_0 + S_t s^2_1, \; s^2_0 < s^2_1,$$
$$\Pr[S_t = 0 | S_{t-1} = 0] = q, \text{ and } \Pr[S_t = 1 | S_{t-1} = 1] = p.$$

TABLE 2.3 – Part 1: Volatility and the Market Risk Premium – Feedback in the Presence of Partial Learning, 1926-51

	Model Specification	
	d is freely estimated. Agents observe past returns and partially learn about state during period t $\{y^b_t = r_{t-1}, r_{t-2}, \ldots\}$,	d is restricted by theory.* Agents observe past returns and partially learn about state during period t $\{y^b_t = r_{t-1}, r_{t-2}, \ldots\}$,
Parameters	$\{y^e_t = S_t\}$	$\{y^e_t = S_t\}$
m_0	0.01242	0.01057
	(0.00300)	(0.00285)
m_1	-0.02314	0.00162
	(0.01401)	(0.00147)
d	-0.04016	-0.02652
	(0.01804)	(0.01710)
s_0	0.03897	0.03920
	(0.00214)	(0.00210)
s_1	0.11591	0.11706
	(0.01024)	(0.01018)
q	0.98601	0.98745
	(0.01281)	(0.01151)
p	0.95699	0.95444
	(0.03410)	(0.03209)
Log Likelihood	450.84133	449.23066

Maximum likelihood estimates are calculated using continuously compounded total monthly value-weighted excess returns of all NYSE-listed stocks for the sample period of January 1926 to December 1951. Asymptotic standard errors are reported in parentheses. The model has the following general form:

$$r_t = m_0 + m_1 \Pr[S_t = 1 | y^b_t] + d\{\Pr[S_t = 1 | y^e_t] - \Pr[S_t = 1 | y^b_t]\} + e_t,$$

where

$$e_t | S_t \sim N(0, s^2_t),$$
$$s^2_t = (1 - S_t) s^2_0 + S_t s^2_1,\ s^2_0 < s^2_1,$$
$$\Pr[S_t = 0 | S_{t-1} = 0] = q, \text{ and } \Pr[S_t = 1 | S_{t-1} = 1] = p.$$

* The following parameter restriction applies: $d = -m_1/(1 - r\,l)$, where $r = 0.997$ and $l = p + q - 1$.

TABLE 2.3 – Part 2: Volatility and the Market Risk Premium – Feedback in the Presence of Partial Learning, 1952-96

	Model Specification	
	d is freely estimated. Agents observe past returns and partially learn about state during period t $\{y^b_t = r_{t-1}, r_{t-2}, \ldots\}$, $\{y^e_t = S_t\}$	d is restricted by theory.* Agents observe past returns and partially learn about state during period t $\{y^b_t = r_{t-1}, r_{t-2}, \ldots\}$, $\{y^e_t = S_t\}$
Parameters		
m_0	0.00495	0.00445
	(0.00192)	(0.00168)
m_1	0.00223	0.00577
	(0.00707)	(0.00182)
d	-0.05327	-0.05099
	(0.00808)	(0.00649)
s_0	0.03130	0.03141
	(0.00102)	(0.00100)
s_1	0.05360	0.05359
	(0.00290)	(0.00288)
q	0.97444	0.97413
	(0.00653)	(0.00656)
p	0.91977	0.91544
	(0.02394)	(0.02330)
Log Likelihood	997.85740	997.72206

Maximum likelihood estimates are calculated using continuously compounded total monthly value-weighted excess returns of all NYSE-listed stocks for the sample period of January 1952 to December 1996. Asymptotic standard errors are reported in parentheses. The model has the following general form:

$$r_t = m_0 + m_1 \Pr[S_t = 1 | y^b_t] + d\{\Pr[S_t = 1 | y^e_t] - \Pr[S_t = 1 | y^b_t]\} + e_t,$$

where

$$e_t | S_t \sim N(0, s^2_t),$$
$$s^2_t = (1 - S_t) s^2_0 + S_t s^2_1, \; s^2_0 < s^2_1,$$
$$\Pr[S_t = 0 | S_{t-1} = 0] = q, \text{ and } \Pr[S_t = 1 | S_{t-1} = 1] = p.$$

* The following parameter restriction applies: $d = -m_1/(1 - r\,l)$, where $r = 0.997$ and $l = p + q - 1$.

TABLE 2.4 – Part 1: Volatility and the Market Risk Premium – Feedback in the Presence of Full Learning, 1926-51

	Model Specification	
Parameters	d is freely estimated. Agents observe past returns and fully learn about state during period t $\{y^{b}_{t} = S_{t-1}\}$, $\{y^{e}_{t} = S_{t}\}$	d is restricted by theory.* Agents observe past returns and fully learn about state during period t $\{y^{b}_{t} = S_{t-1}\}$, $\{y^{e}_{t} = S_{t}\}$
m_0	0.01118 (0.00301)	0.00980 (0.00289)
m_1	-0.02271 (0.01446)	0.00613 (0.00367)
d	-0.10618 (0.03453)	-0.08538 (0.03022)
s_0	0.03955 (0.00209)	0.03973 (0.00205)
s_1	0.12032 (0.01067)	0.12400 (0.01103)
q	0.98437 (0.00984)	0.98416 (0.00977)
p	0.94947 (0.03219)	0.94678 (0.03330)
Log Likelihood	451.62328	449.64235

Maximum likelihood estimates are calculated using continuously compounded total monthly value-weighted excess returns of all NYSE-listed stocks for the sample period of January 1926 to December 1951. Asymptotic standard errors are reported in parentheses. The model has the following general form:

$$r_t = m_0 + m_1 \Pr[S_t = 1 | y^{b}_{t}] + d\{\Pr[S_t = 1 | y^{e}_{t}] - \Pr[S_t = 1 | y^{b}_{t}]\} + e_t,$$

where

$$e_t | S_t \sim N(0, s^2_t),$$
$$s^2_t = (1 - S_t) s^2_0 + S_t s^2_1,\ s^2_0 < s^2_1,$$
$$\Pr[S_t = 0 | S_{t-1} = 0] = q, \text{ and } \Pr[S_t = 1 | S_{t-1} = 1] = p.$$

* The following parameter restriction applies: $d = -m_1/(1 - r\,l)$, where $r = 0.997$ and $l = p + q - 1$.

TABLE 2.4 – Part 2: Volatility and the Market Risk Premium – Feedback in the Presence of Full Learning, 1952-96

	Model Specification	
Parameters	d is freely estimated. Agents observe past returns and fully learn about state during period t $\{y^{b}_{t}=S_{t-1}\}$, $\{y^{e}_{t}=S_{t}\}$	d is restricted by theory.* Agents observe past returns and fully learn about state during period t $\{y^{b}_{t}=S_{t-1}\}$, $\{y^{e}_{t}=S_{t}\}$
m_0	0.00667	0.00456
	(0.00190)	(0.00182)
m_1	-0.00751	0.01007
	(0.00803)	(0.00471)
d	-0.08590	-0.07401
	(0.01414)	(0.01611)
s_0	0.03082	0.03082
	(0.00155)	(0.00154)
s_1	0.05322	0.05499
	(0.00461)	(0.00506)
q	0.97159	0.97246
	(0.01325)	(0.01381)
p	0.88739	0.89416
	(0.05972)	(0.06600)
Log Likelihood	996.78935	993.35811

Maximum likelihood estimates are calculated using continuously compounded total monthly value-weighted excess returns of all NYSE-listed stocks for the sample period of January 1952 to December 1996. Asymptotic standard errors are reported in parentheses. The model has the following general form:

$$r_t = m_0 + m_1 \Pr[S_t=1|y^{b}_{t}] + d\{\Pr[S_t=1|y^{e}_{t}] - \Pr[S_t=1|y^{b}_{t}]\} + e_t,$$

where

$$e_t|S_t \sim N(0, s^2_t),$$
$$s^2_t = (1-S_t)s^2_0 + S_t s^2_1,\ s^2_0 < s^2_1,$$
$$\Pr[S_t=0|S_{t-1}=0]=q, \text{ and } \Pr[S_t=1|S_{t-1}=1]=p.$$

* The following parameter restriction applies: $d=-m_1/(1-r\,l)$, where $r=0.997$ and $l=p+q-1$.

TABLE 2.5 – Part 1: Volatility and the Market Risk Premium – Generalized Learning, 1926-51

	Model Specification		
Parameters	Partial learning occurs within one month $\{y^r_t = r_{t-1}, r_{t-2}, \ldots\}$, $\{y^k_t = S_{t-k}, k=0\}$	Partial learning occurs within two months $\{y^r_t = r_{t-1}, r_{t-2}, \ldots\}$, $\{y^k_t = S_{t-k}, k=0,1\}$	Partial learning occurs within three months $\{y^r_t = r_{t-1}, r_{t-2}, \ldots\}$, $\{y^k_t = S_{t-k}, k=0,1,2\}$
p_c	0.01242	0.01133	0.01134
	(0.00300)	(0.00324)	(0.00319)
p_r	0.01702	-0.00272	-0.00502
	(0.01864)	(0.02140)	(0.02577)
p_0	-0.04016	-0.10627	-0.10704
	(0.01796)	(0.03374)	(0.03426)
p_1	-	0.08564	0.08343
	-	(0.03778)	(0.03964)
p_2	-	-	0.00514
	-	-	(0.03298)
s_0	0.03897	0.03952	0.03954
	(0.00214)	(0.00209)	(0.00208)
s_1	0.11591	0.12038	0.12039
	(0.01024)	(0.01067)	(0.01065)
q	0.98601	0.98434	0.98439
	(0.01281)	(0.00978)	(0.00971)
p	0.95699	0.95020	0.95052
	(0.03410)	(0.03233)	(0.03236)
Log Likelihood	450.84133	451.63413	451.64208

Maximum likelihood estimates are calculated using continuously compounded total monthly value-weighted excess returns of all NYSE-listed stocks for the sample period of January 1926 to December 1951. Asymptotic standard errors are reported in parentheses. The model has the following general form:

$$r_t = p_c + p_r \Pr[S_t = 1 | y^r_t] + \Sigma_k p_k \Pr[S_t = 1 | y^k_t] + e_t,$$

where

$$e_t | S_t \sim N(0, s^2_t),$$
$$s^2_t = (1 - S_t) s^2_0 + S_t s^2_1, \; s^2_0 < s^2_1,$$
$$\Pr[S_t = 0 | S_{t-1} = 0] = q, \text{ and } \Pr[S_t = 1 | S_{t-1} = 1] = p.$$

TABLE 2.5 – Part 2: Volatility and the Market Risk Premium – Generalized Learning, 1952-96

	Model Specification		
Parameters	Partial learning occurs within one month $\{y^r_t = r_{t-1}, r_{t-2}, \ldots\}$, $\{y^k_t = S_{t-k}, k=0\}$	Partial learning occurs within two months $\{y^r_t = r_{t-1}, r_{t-2}, \ldots\}$, $\{y^k_t = S_{t-k}, k=0,1\}$	Partial learning occurs within three months $\{y^r_t = r_{t-1}, r_{t-2}, \ldots\}$, $\{y^k_t = S_{t-k}, k=0,1,2\}$
p_c	0.00495	0.00523	0.00501
	(0.00191)	(0.00201)	(0.00197)
p_r	0.05551	0.02373	0.05400
	(0.00847)	(0.01344)	(0.02969)
p_0	-0.05327	-0.08250	-0.07228
	(0.00805)	(0.01641)	(0.01624)
p_1	-	0.05926	0.05155
	-	(0.02340)	(0.03397)
p_2	-	-	-0.03107
	-	-	(0.02500)
s_0	0.03130	0.03100	0.03072
	(0.00101)	(0.00181)	(0.00146)
s_1	0.05360	0.05341	0.05195
	(0.00290)	(0.00500)	(0.00371)
q	0.97444	0.97394	0.97098
	(0.00654)	(0.01339)	(0.01160)
p	0.91977	0.91540	0.91506
	(0.02395)	(0.04255)	(0.02704)
Log Likelihood	997.85740	998.71633	999.28236

Maximum likelihood estimates are calculated using continuously compounded total monthly value-weighted excess returns of all NYSE-listed stocks for the sample period of January 1952 to December 1996. Asymptotic standard errors are reported in parentheses. The model has the following general form:

$$r_t = p_c + p_r \Pr[S_t=1|y^r_t] + \Sigma_k p_k \Pr[S_t=1|y^k_t] + e_t,$$

where

$$e_t|S_t \sim N(0, s^2_t),$$
$$s^2_t = (1-S_t)s^2_0 + S_t s^2_1,\ s^2_0 < s^2_1,$$
$$\Pr[S_t=0|S_{t-1}=0]=q,\ \text{and}\ \Pr[S_t=1|S_{t-1}=1]=p.$$

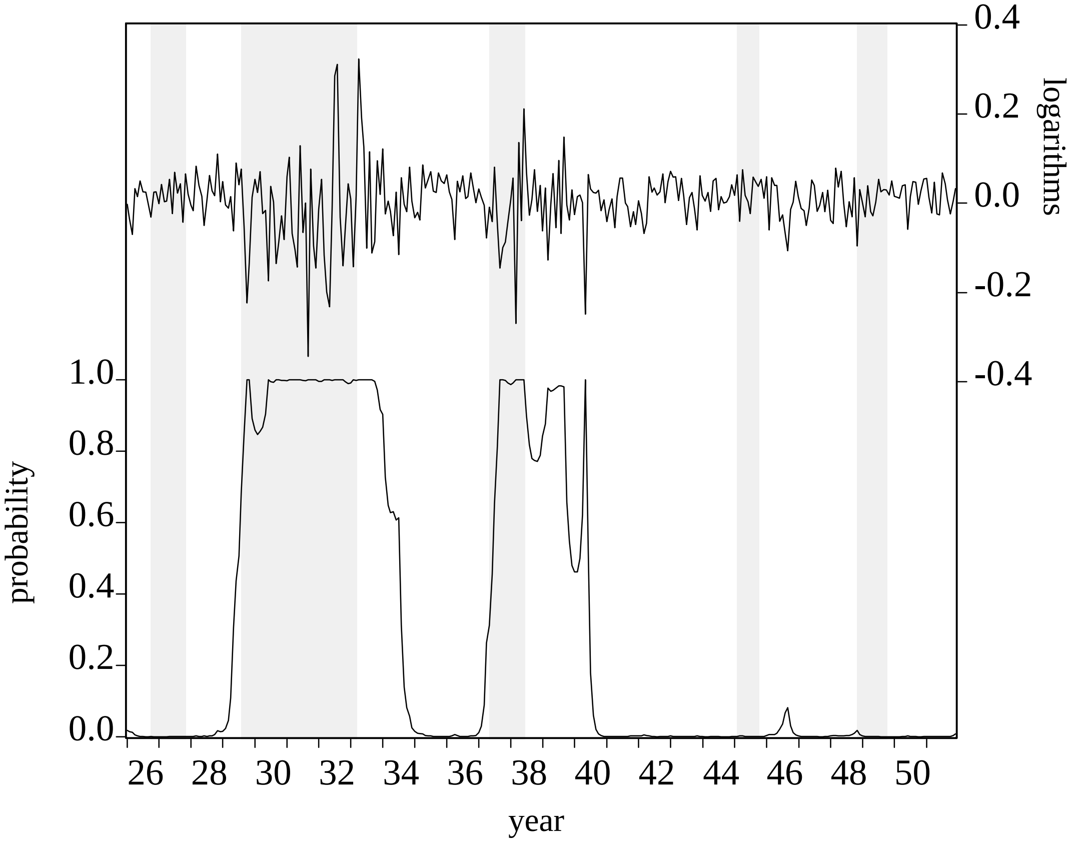


FIGURE 2.1 – Part 1: Excess Stock Returns and Smoothed Probabilities of a High Volatility Regime for the Markov-Switching Variance Specification of Stock Market Volatility, 1926-51.

Source: Excess stock returns are continuously compounded total monthly value-weighted returns of all NYSE-listed stocks from CRSP in excess of one-month Treasury bill yields from CRSP for the sample period of January 1926 to December 1951. Smoothed probabilities are calculated using Kim's (1994) smoothing algorithm and maximum likelihood estimates for the Markov-switching specification presented in Table 1. NBER-dated recessions are shaded.

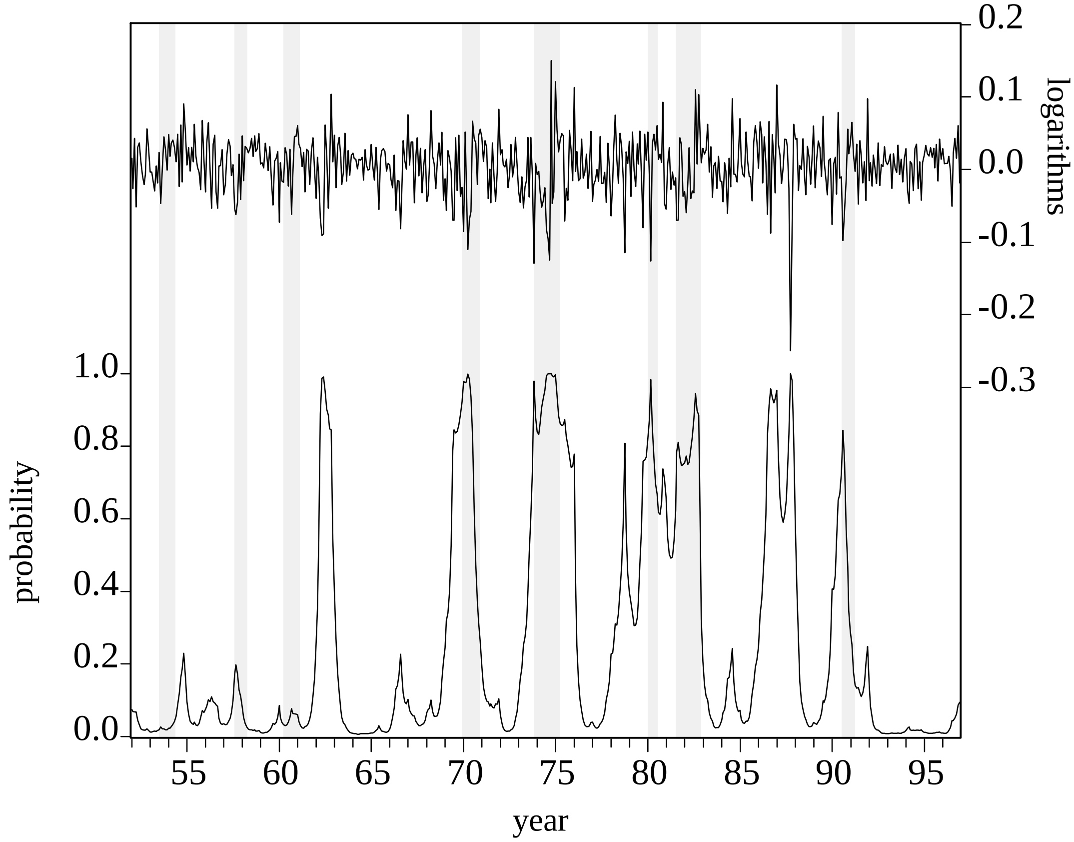


FIGURE 2.1 – Part 2: Excess Stock Returns and Smoothed Probabilities of a High Volatility Regime for the Markov-Switching Variance Specification of Stock Market Volatility, 1952-96.

Source: Excess stock returns are continuously compounded total monthly value-weighted returns of all NYSE-listed stocks from CRSP in excess of one-month Treasury bill yields from CRSP for the sample period of January 1952 to December 1996. Smoothed probabilities are calculated using Kim's (1994) smoothing algorithm and maximum likelihood estimates for the Markov-switching specification presented in Table 1. NBER-dated recessions are shaded.

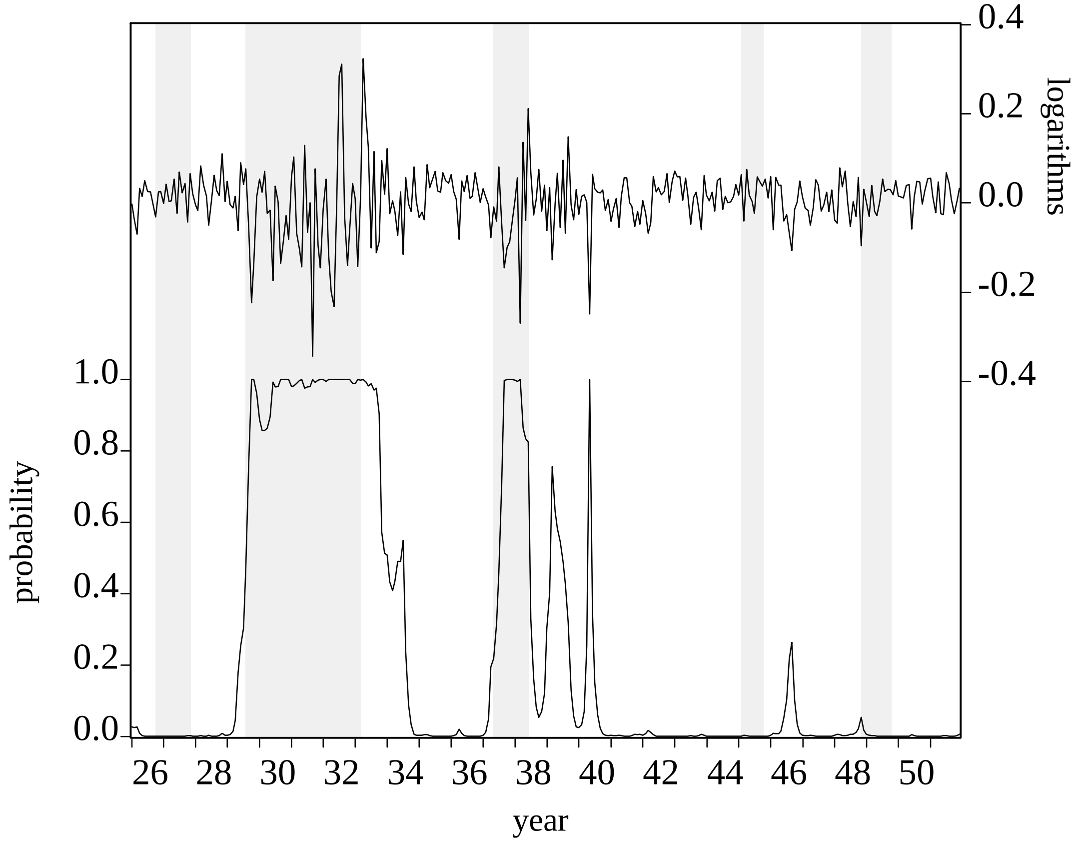


FIGURE 2.2 – Part 1: Excess Stock Returns and Smoothed Probabilities of a High Volatility Regime for a Generalized Learning Specification of Volatility and the Market Risk Premium, 1926-51.

Source: Excess stock returns are continuously compounded total monthly value-weighted returns of all NYSE-listed stocks from CRSP in excess of one-month Treasury bill yields from CRSP for the sample period of January 1926 to December 1951. Smoothed probabilities are calculated using Kim's (1994) smoothing algorithm and maximum likelihood estimates for the second generalized learning specification presented in Table 5. NBER-dated recessions are shaded.

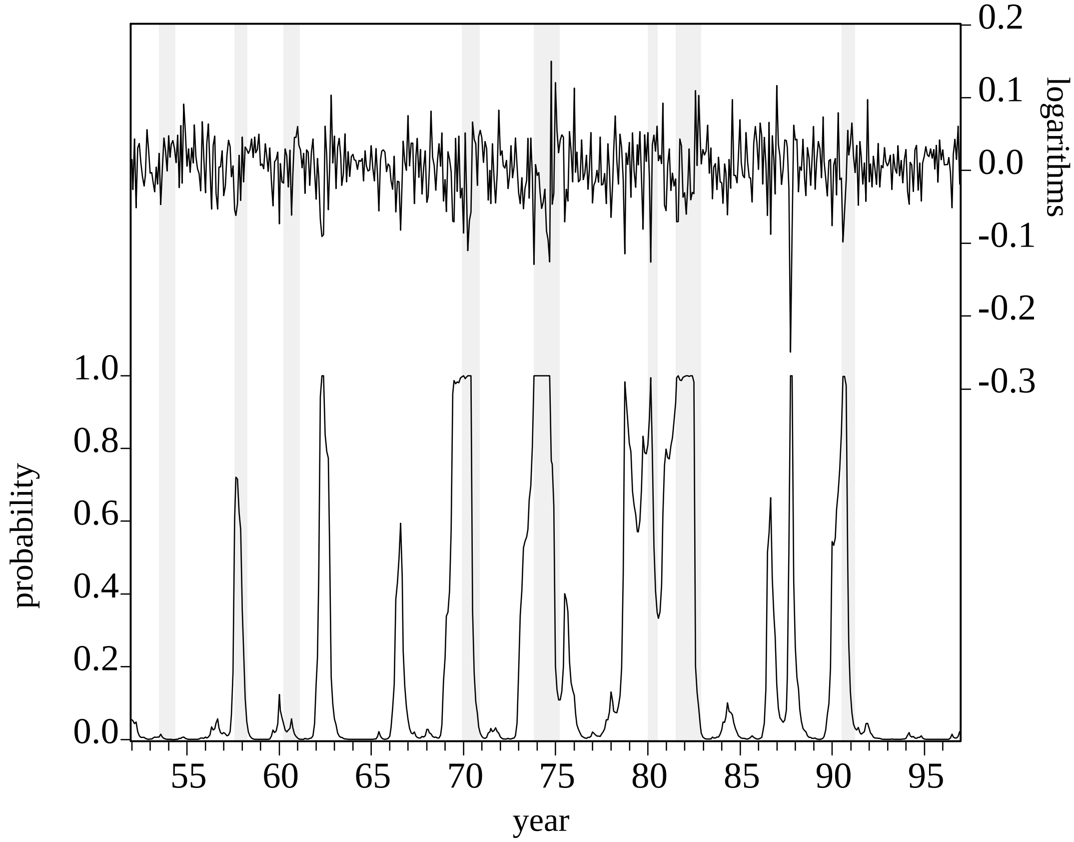


FIGURE 2.2 – Part 2: Excess Stock Returns and Smoothed Probabilities of a High Volatility Regime for a Generalized Learning Specification of Volatility and the Market Risk Premium, 1952-96.

Source: Excess stock returns are continuously compounded total monthly value-weighted returns of all NYSE-listed stocks from CRSP in excess of one-month Treasury bill yields from CRSP for the sample period of January 1952 to December 1996. Smoothed probabilities are calculated using Kim's (1994) smoothing algorithm and maximum likelihood estimates for the second generalized learning specification presented in Table 5. NBER-dated recessions are shaded.

CHAPTER 3: Testing the Expectations Hypothesis with a State-Space Model of the Term Structure*

3.1 Introduction

Modern empirical studies of the term structure of interest rates, beginning with Shiller (1979), have focused on a formulation of the Expectations Hypothesis that assumes rational expectations and constant term premia. A testable implication of this formulation is that forward rates are unbiased (up to a constant term premium) predictors of future spot rates. Most studies in the literature test and statistically reject this implication using postwar U.S. data, although the evidence for other periods and countries is somewhat mixed.[1] Cook and Hahn (1990) and Shiller (1990) provide excellent surveys of this vast literature.

An alternative testable implication of the modern Expectations Hypothesis is that the variances of term premia, extracted from interest rates under the assumption of rational expectations, are equal to zero. This approach has received less attention than forward rate unbiasedness since it is not immediately obvious how to obtain estimates of the variances of unobserved term premia. Startz (1982) proposes a method based on the decomposition implied by rational expectations of the variance of an excess forward returns—that is, forward rates minus the realized spot rates—into a term premium variance and the market forecast error variance. Given an upper bound estimate of the forecast error variance, which he calculates from basic interest rate forecasting equations, he is able to back out a lower bound estimate of the term premium variance. Contrary to the Expectations Hypothesis, he finds statistically significant estimates of positive term premia variances using U.S. data for the period of 1953 to 1971. DeGennaro and Moser (1989) confirm Startz's findings with more recent data from 1970 to 1982.

---

* This chapter contains materials first presented in Gravelle and Morley (1999).

[1] See Campbell and Shiller (1991), Fama (1984), Fama and Bliss (1987), and Shiller, Campbell, and Schoenholtz (1983) for examples of studies that reject the Expectations Hypothesis using postwar U.S. data; Mankiw and Miron (1986) for a study that fails to reject the hypothesis using pre-World War I data; and Hardouvelis (1994) and Mankiw (1986) for studies that fail to reject using postwar data from other countries, including Canada, France, Germany, Italy, Japan, and the U.K.

In this chapter, we take a slightly different approach to testing this alternative implication of the Expectations Hypothesis. Specifically, we apply the Kalman filter to estimate a state-space model of excess forward returns and the term structure.[2] The intuition behind our approach is straightforward. Under rational expectations, market forecast errors are serially uncorrelated. Therefore, any systematic components of excess forward returns over different horizons not related to data overlap should reflect term premia. The state-space framework, then, allows us to test the Expectations Hypothesis by extracting these systematic components from the observed data and examining whether they vary significantly over time.

Using Canadian spot and forward interest rates, we find strong evidence against the Expectations Hypothesis. Specifically, we reject the hypothesis that the variances of the term premia are equal to zero for all horizons. This finding is notable because it is contrary to findings in previous studies of the Expectations Hypothesis using Canadian data, including Hardouvelis (1994) and Stréliski (1998). Indeed, we use the same data as Stréliski, suggesting that our approach is able to detect evidence obscured by more traditional tests. We also find some evidence of a positive relationship between interest rate volatility and the levels of the term premia. Finally, we find evidence for negative term premia in the late 1980s, providing support for Modigliani and Sutch's (1966) "preferred habitat" view of market preferences. Alternatively, the evidence may reflect a failure of strict rational expectations or the presence of a "peso problem" arising from a change in Canadian monetary policy.

The rest of the chapter is organized as follows. Section 3.2 presents the state-space model of the term structure, including the theoretical framework underlying the approach. Section 3.3 describes the data and presents the empirical results. Section 3.4 concludes. Tables and figures follow the conclusion.

---

[2] Iyer (1997) also applies the Kalman filter to estimate time-varying term premia. He posits an ARMA(1,2) with ARCH variance specification for the term premium on one-month-ahead forward rates and finds evidence against the Expectations Hypothesis using postwar U.S. data for the period of 1951 to 1991. By contrast, we employ different term premia specifications and examine forward rates over multiple horizons.

3.2 The Model

3.1.1 Theoretical Framework

The approach we take to testing the Expectations Hypothesis involves developing a state-space model of the term structure that allows us to identify the behaviour of term premia over time. In particular, we employ the Kalman filter to extract term premia from excess forward returns over various *j*-period horizons ($j = 1,2,\ldots,J$ months).

To motivate this approach, we decompose an observed *j*-period-ahead forward interest rate into an unobserved term premium and unobserved market expectations about the future underlying spot rate. That is,

$$f_t^{j,k} \equiv q_t^{j,k} + E_t[r_{t+j}^k], \tag{3.1}$$

where $f_t^{j,k}$ is the interest rate on a future *k*-period security (e.g., a 90-day bankers' acceptance) set this period in a forward contract with a settlement date *j* periods in the future; $q_t^{j,k}$ is the term premium on the forward contract and is associated with the maturity date of the underlying security; $E_t[r_{t+j}^k]$ is this period's market expectation about the future spot rate on the same *k*-period security; and $t = 1,2,\ldots,T$. As is, equation (3.1) represents an identity since we define the term premium as the residual between the forward rate and market expectations.

Subtracting the realized spot rate from both sides of equation (3.1) produces the following expression for an excess forward return on a *k*-period security over a *j*-period horizon:

$$f_t^{j,k} - r_{t+j}^{j,k} = q_t^{j,k} + E_t[r_{t+j}^{j,k}] - r_{t+j}^{j,k} \equiv q_t^{j,k} + \sum_{i=0}^{j-1} e_{t+i,t+i+1}^{j-i,k}, \tag{3.2}$$

where $e_{t+i,t+i+1}^{j-i,k} \equiv E_{t+i}[r_{t+j}^k] - E_{t+i+1}[r_{t+j}^k]$ denotes the change in market expectations about a given future spot interest rate between periods i and $i+1$.

The Expectations Hypothesis, then, can be thought of in terms of the specific restrictions it imposes on the unobserved components of excess forward returns given in equation (3.2). Broadly speaking, the Expectations Hypothesis is the idea that market expectations are the primary determinant of the shape of the term structure. While the modern formulation of the hypothesis allows other factors reflected in term premia to play a role in determining the shape, it restricts their role to be fixed over time. Meanwhile, the modern formulation assumes that expectations are rational in the sense of Muth (1960). Thus, the restrictions on equation (3.2) are given as follows:

$$q_t^{j,k} = \bar{q}^{j,k}, \tag{3.3}$$

$$E_{t+i}[e_{t+i,t+i+1}^{j,k}] = 0. \tag{3.4}$$

That is, term premia are constant over time and non-overlapping changes in market expectations have no serial correlation.

We can test the Expectations Hypothesis by determining if either restriction fails in practice. Our approach is to use the restriction of no serial correlation to identify term premia in a state-space model of excess forward returns. Then, we examine how the term premia have evolved over time and test the restriction that they are constant.

3.1.2 A State-Space Model of the Term Structure

For the state-space model of the term structure, we need to specify state equations describing the evolution of the unobserved components of excess forward returns given in equation (3.2). Starting with the sum of changes in expectations, we consider a moving average (MA) specification, which follows from the assumption that, while changes in expectations about different future interest rates are serially uncorrelated, they are contemporaneously correlated. In particular, the contemporaneous correlation implies that the sum of changes in expectations will have the same unconditional moments as an MA($j-1$) process.[3] That is,

$$\sum_{i=0}^{j-1} e_{t+i,t+1+i}^{j-i,k} = y\,(L)e_{t+j}\,, \tag{3.5}$$

where

$$y\,(L) \equiv 1 + y_{1}L + y_{2}L^{2} + \cdots + y_{\,j-1}L^{j-1} \tag{3.6}$$

and

$$e_{t+j} \sim i.i.d.\,N(0, s_{e}^{2}(S_{t})), \tag{3.7}$$

$$s_{e}(S_{t}) = (1-S_{t})s_{e}(0) + S_{t}s_{e}(1), \tag{3.8}$$

$$S_{t+i} = \{0,1\},\;\; s_{e}^{2}(0) < s_{e}^{2}(1), \tag{3.9}$$

$$\Pr[S_{t+i} = 0] = 1-p, \text{ and } \Pr[S_{t+i} = 1] = p. \tag{3.10}$$

Note that we assume that the errors for the MA process switch independently between two Normal distributions. This mixture distribution is designed to capture the presence of large transient outliers in excess forward returns.

In terms of the state equation describing the evolution of the term premium, we consider four different specifications. In the first specification, the term premium is assumed to be constant, as in equation (3.3). In the second specification, the term premium follows a driftless random walk. That is,

$$q_{t}^{j,k} = q_{t-1}^{j,k} + v_{t}^{j,k},\;\; v_{t}^{j,k} \sim i.i.d.\,N(0, s_{vj,k}^{\,2}), \tag{3.11}$$

[3] See Appendix B.1 for details.

where $s^2_{vj,k} = 0$ corresponds to the first specification. In the third specification, the term premium follows a stationary first-order autoregressive (AR(1)) process. That is,

$$q_t^{j,k} = c + f q_{t-1}^{j,k} + v_t^{j,k}, \;\; v_t^{j,k} \sim i.i.d.N(0, s^2_{vj,k}), \tag{3.12}$$

where $|f| < 1$ implies stationarity. In the fourth specification, the term premium follows a modified AR(1) process, with a switching intercept term that is conditional on the volatility regime for e_t. Specifically, the constant term c in equation (3.12) is replaced with the following:

$$c(S_t) = (1 - S_t)c(0) + S_t c(1). \tag{3.13}$$

The first three specifications are used to test the Expectations Hypothesis. The fourth specification is used to examine the interaction between interest rate volatility and the levels of the term premia.

The random walk and AR(1) specifications warrant some further discussion. On the one hand, the advantage of the random walk specification is that it provides a parsimonious way of allowing for permanent changes in term premia, while being fairly robust to misspecification, as long as the true dynamics are highly autocorrelated.[4] Allowing for permanent changes is consistent with findings on cointegration between spot and forward interest rates that imply a non-stationary component in excess forward returns.[5] Also, the random walk specification is more conservative than the AR(1) specification—or, for that matter, any higher order AR(p) specification, with or without a unit root—in attributing higher frequency movements in excess forward returns to term premia. Thus, it provides a more convincing test of the Expectations Hypothesis. One the

[4] See Garbade (1977).
[5] See Evans and Lewis (1994). Gravelle, Muller, and Stréliski (1998) confirm their findings for the Canadian data used in this paper.

other hand, the advantage of the AR(1) specification is that it provides a more accurate measure of term premia if most of the movements in the term premia are not permanent, but mean reverting. Also, the AR(1) specification allows us to directly examine the interaction between interest rate volatility and the levels of the term premia, as done in the fourth specification, while information about levels should not be relevant in a random walk specification.

Given state equations for the components of equation (3.2), we can proceed to estimate the state-space model of the term structure using a modified version of Kim's (1994) filter and approximate maximum likelihood estimation. In particular, Kim's filter allows for regime switching in a state-space model framework by combining the Kalman filter with Hamilton's (1989) filter for Markov-switching. We modify the filter to the more restrictive case of independent switching.[6] Note that the approximations involved in estimation via Kim's (1994) filter arise from the switching variance of the MA errors and imply that we will never find the most efficient estimates.[7] However, to the extent that accounting for heteroskedasticity improves the efficiency of our estimates over the homoskedastic case, such approximations are justified.

3.3 Empirical Results

3.3.1 Data

In our empirical applications, we use daily closing yields on Canadian 90-day bankers' acceptances and *j*-month-ahead forward rate agreements (*j*=1,...,9 months) on bankers' acceptances from the Bank of Canada's data files. In order to avoid problems associated with overlapping observations, we sample the data at a monthly frequency, looking only at end-of-the-month closing yields. This gives us 119 observations for each type of forward contract over the period of August 1988 to June 1998. The excess forward return series are constructed by taking difference between the continuously compounded yield on a forward rate agreement and the subsequent realized continuously

[6] See Appendix B.2 for details.
[7] Similarly, if the MA errors had an ARCH distribution, estimation would involve approximations (Harvey, Ruiz, and Sentana, 1992).

compounded yield on the underlying bankers' acceptance and multiplying by 100 to obtain percentages.

Figure 3.1 displays the excess forward return series. Consistent with the MA($j-1$) implication discussed above, the returns get more persistent as the horizon increases. The only returns that should have no serial correlation under the Expectations Hypothesis are the one-month-ahead returns, which appear to have large transient outliers. The Jarque and Bera (1980) test statistic of Normality based on the third and fourth sample moments of the returns is 725.673, which given a $c^2(2)$ distribution under the null of Normality is strongly significant.

Testing the null hypothesis of homoskedasticity against the alternative of independent switching for one-month-ahead excess forward returns produces a likelihood ratio statistic of 59.153. While the distribution of this test statistic is non-standard since the transition probability p is not identified under the null hypothesis (see Garcia, 1995, and Hansen, 1992), it is large and far exceeds any of the asymptotic critical values reported in Garcia (1995). Therefore, we can be reasonably confident in allowing for independent switching. Indeed, the Jarque and Bera (1980) statistic for the standardized returns in the presence of independent switching is only 3.513, with a corresponding p-value of 0.173. Meanwhile, the likelihood ratio test for the null of independent switching against the alternative of Markov switching, which does have a standard distribution, is only 0.240, with a corresponding p-value of 0.624. Therefore, we use only independent switching to capture the heteroskedasticity in excess forward returns.

3.3.2 Testing the Expectations Hypothesis

Table 3.1 reports log-likelihood values for the four term premium specifications and the nine forward rate horizons.[8] Broadly speaking, the likelihood values provide consistent evidence against the Expectations Hypothesis. For instance, if we use the

[8] Maximum likelihood estimation was conducted using the OPTMUM procedure for the GAUSS programming language. Numerical derivatives were used in estimation, as well as for calculation of asymptotic standard errors. The variance and probability parameters were appropriately constrained. Inferences were robust to a variety of starting values.

random walk specification to test the Expectations Hypothesis, we obtain likelihood ratio statistics for the null hypothesis $H_0 : s^{2}_{vj,k} = 0$ ranging from 8.382 for the eight-month-ahead case to as much as 24.080 for the one-month-ahead case. Given a $c^{2}(1)$ distribution, these statistics are all significant at better than a 1 percent level. Furthermore, Kendell and Stuart (1973) and Garbade (1977) argue that using a $c^{2}(1)$ distribution for this test actually understates the true significance due to a bunching of the probability mass toward the origin under the null hypothesis. Therefore, the evidence against the Expectations Hypothesis is even stronger.

Figure 3.2 displays the smoothed inferences about the term premia for the random walk specification and the nine forward rate horizons. The smoothed (two-sided) inferences represent point estimates and 95 percent confidence bands conditional on all available sample information. It turns out that the smoothed inferences are very similar to the filtered (one-sided) inferences, which are conditional on information up to time t only. We report the smoothed inferences because they tend to better reflect the estimated degree of variability, including remaining constant if the estimated variance is zero. However, it is evident from the figure that, consistent with the results in Table 3.1, none of the estimated term premia is constant. We discuss these inferences in more detail below.

Figure 3.3 displays the smoothed inferences about the term premia for the stationary AR(1) specification and the nine forward rate horizons. What is noticeable about these inferences is how similar they are to the inferences for the random walk specification. This is, perhaps, not too surprising given the results in Table 3.1. In particular, while the implied likelihood ratio statistics for the null of the random walk against the alternative of the AR(1) are large by conventional standards (they range from 4.500 for the four-month-ahead case to 7.250 for the eight-month-ahead case), they are probably affected by the upward bias described in Dickey and Fuller (1981). In any event, inferences for both specifications strongly contradict the Expectations Hypothesis. Again, we discuss these in more detail below.

As a diagnostic check, we examine the smoothed inferences about the standardized innovations of the MA($j-1$) processes for serial correlation and Normality. To be clear, these inferences are an imperfect proxy for the true unobserved innovations and might well display serial correlation and non-Normality even if the true innovations do not. However, this approximation should be reasonable given that the variances of the term premia are small relative to the MA innovations. Indeed, tests reveal only weak evidence for serial correlation (e.g., for the one-month-ahead random walk specification, the Box and Pierce (1970) Q-statistics are $Q(12)$=16.567, $Q(24)$=25.316 and $Q(36)$=47.839, while for the nine-month-ahead random walk specification, they are $Q(12)$=10.965, $Q(24)$=16.696 and $Q(36)$=21.459). Meanwhile, the Jarque-Bera test of the third and fourth sample moments generally provide support for Normality (e.g., for the one-month-ahead random walk specification, the statistic is 2.866, with a p-value of 0.239, although for the nine-month-ahead random walk specification, the statistic is 91.991, with a p-value <0.001).

3.3.3 Volatility and Term Premia

An interesting feature of Figures 3.2 and 3.3 is that estimated term premia appear to peak around the same time that excess forward returns appear most volatile in Figure 3.1. In this subsection, we formally investigate the interaction between volatility and the levels of the term premia. Notably, our approach can be thought of as an alternative to Engle, Lilien, and Robins' (1987) autoregressive conditional heteroskedasticity in mean (ARCH-M) model that relates term premia to the conditional variance of excess holding yields. An advantage of our approach is that independent switching allows for less persistent shifts in volatility than an ARCH model. This is appropriate since the Canadian data are characterized by little or no persistence; although with little persistence comes little reason to expect a relationship between volatility and the levels of the term premia.

The results in Table 3.1 suggest that the switching intercepts are only significant for the shorter forward rate horizons. In particular, the implied likelihood ratio statistics are 3.250 for one-month-ahead case, with a p-value of 0.071, 3.764 for the two-month-ahead case, with a p-value of 0.052, and 2.880 for the four-month-ahead case, with a p-value of 0.090. Note that these statistics should have a standard distribution since the MA errors are switching under both competing hypotheses. In all three cases, volatility has a positive effect on the level of the term premia, as in Engle, Lilien, and Robins (1987). Specifically, the transition from a low volatility regime to a high volatility regime increases the intercepts for these term premia by an estimated 25, 55, and 49 basis points, respectively.

Figure 3.4 displays smoothed inferences about the term premium for the AR(1) with switching intercept specification and the one-period-ahead forward horizon against the smoothed probability of being in a high volatility regime. The comparable figures for the two-month-ahead and the four-month-ahead cases are very similar and are not reported here to conserve space. The purpose of this figure, then, is to demonstrate the timing of the relationship between volatility and the term premium. For instance, we can see the clustering of high volatility episodes in the middle of the sample coinciding with the period in which the estimated term premium was highest. In particular, the two spikes in 1992, which correspond to Standard and Poor's downgrading of Canadian debt and the failure of a national referendum on a constitution amendment, produced large increases in the level of the term premium.[9] Meanwhile, the absence of such volatile events and Canada's recently improved fiscal situation corresponds to the decline in the term premium at the end of the sample.

The lack of stronger evidence of a link between volatility and the level of the term premium for longer forward rate horizons might seem surprising in light of the evidence for shorter horizons reported above, but it makes some sense given a lack of persistence for the high volatility regime. In particular, to the extent that term premia are risk premia, they should reflect beliefs about future volatility. But, if volatility is not very persistent, then observation of this period's volatility provides little information about future volatility.

3.3.4. Negative Term Premia

Perhaps, more surprising is the evidence of negative term premia in the early part of the sample. This evidence, while not clearly statistically significant according to the 95 percent confidence bands in Figures 3.2 and 3.3, still provides a serious challenge to the traditional view dating back to Lutz (1940) that term premia are always positive and increasing with maturity. Instead, it supports the alternative view of Modigliani and Sutch (1966) that individuals can have different preferred investment horizons, or “habitats,” leading to the possibility of negative term premia in the event that market participants have a stronger preference for lending at longer horizons and borrowing at shorter horizons.

An alternative explanation for our findings, though, is the possible failure of the serially uncorrelated expectation errors assumption that underlies our identification of the term premia. This possibility is particularly relevant for Canada given the circumstances surrounding the Bank of Canada’s monetary policy in the late-1980s and early 1990s. Notably, the Bank of Canada shifted to a more explicit anti-inflation stance following the then-Governor John Crow’s Hanson Lecture in 1988 and culminating in the formal introduction of inflation reduction targets in February 1991. A plausible scenario, then, is that prior to the formal announcement of targets the Bank of Canada lacked credibility in its anti-inflation stance. Thus, when the Bank of Canada initially raised interest rates before the targets were introduced, financial markets may have expected a quick reversal of the Bank’s position. The fact that the Bank did not reverse its position may then have resulted in systematic expectation errors as market participants repeatedly expected interest rates to fall when they subsequently did not.

This situation can be thought of as an example of the so-called “peso problem.”[10] Specifically, suppose market participants believed that there was a significant probability of a future return to a less anti-inflationary interest rate policy regime. Then, given that there was no return to this regime within the observed sample, expectations would appear

[9] See Clinton and Zelmer (1997) and Zelmer (1995) for a full chronology of volatile economic and political events that have likely affected Canadian term premia.
[10] See, for example, Bekaert, Hodrick, and Marshall (1997).

irrational even though they were not.[11] But our methodology would incorrectly attribute any systematic expectation errors to the term premia. Thus, the true term premia might have been positive during this early period even though our inferences suggest otherwise. It is only for the later period, when the Bank of Canada's inflation fighting credibility was more firmly established, that we can be more confident in assuming that our inferences reflect the true term premia.[12]

Ultimately, whether the estimated negative term premia are a consequence of changes in "preferred habitats" or a "peso problem" cannot be resolved by our methodology. But, our finding of negative term premia merely implies that a combination of a simple liquidity premium story with serially uncorrelated expectation errors is inappropriate.

3.4 Conclusions

In this chapter, we test the implication of the Expectations Hypothesis that term premia are constant over time. Specifically, we use the Kalman filter under the assumption of rational expectations to extract estimates of term premia from excess forward returns and, contrary to the Expectations Hypothesis, we find evidence of statistically significant time variation in term premia for Canada. Then, we investigate the interaction between interest rate volatility and term premia. At the shorter forward rate horizons, we find some evidence for a positive tradeoff. Finally, we address the issue of the possible existence of negative term premia. We conclude that our finding of an estimated negative term premium during the early part of the sample may reflect changes in "preferred habitats" in the Canadian securities market, although it may reflect a "peso problem" arising from a change in Canadian monetary policy.

---

[11] Of course, market expectations may have actually been irrational. Froot (1989) uses U.S. survey data to show that market expectations as measured by the surveys are biased, especially at short maturities.

[12] It is important to note, however, that our rejection of the Expectations Hypothesis is not sensitive to the exclusion of the first two years of the sample in estimation.

TABLE 3.1: Log Likelihood Values for a State-Space Model of the Term Structure, August 1988 to June 1998

	Forward Rate Horizon		
Term Premium Specification	*j=1*	*j=2*	*j=3*
Constant	-68.156	-97.549	-107.340
Random Walk	-56.116	-89.094	-99.618
Stationary AR(1)	-53.543	-86.428	-96.442
AR(1) with Switching Intercept	-51.918	-84.546	-95.796
	j=4	*j=5*	*j=6*
Constant	-107.866	-123.838	-111.473
Random Walk	-101.328	-116.729	-105.713
Stationary AR(1)	-99.078	-113.910	-102.978
AR(1) with Switching Intercept	-97.638	-113.405	-102.753
	j=7	*j=8*	*j=9*
Constant	-106.641	-108.900	-114.191
Random Walk	-101.813	-104.709	-103.522
Stationary AR(1)	-98.280	-101.084	-100.618
AR(1) with Switching Intercept	-98.247	-101.056	-100.246

Log-likelihood values are for a state-space model of the term structure using 100 times the continuously compounded excess returns on Canadian forward rate agreements over their underlying 90-day bankers' acceptances. The spot and forward interest rate data are from the Bank of Canada. The model allows for independent switching in the variance of changes in expectations and is estimated using Kim's (1994) filter, modified for independent switching.

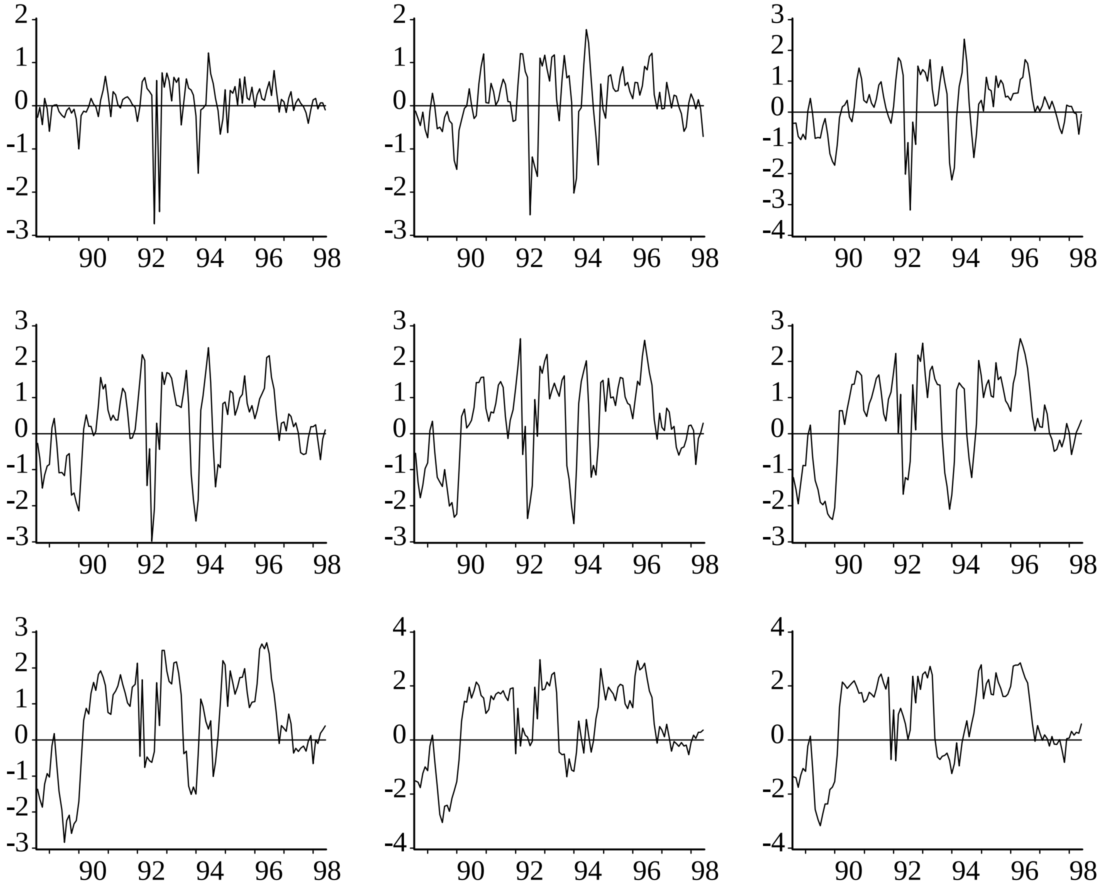


FIGURE 3.1 – Part 1: Excess Forward Returns on Canadian 90-day Bankers' Acceptances, August 1988 to June 1998.

Source: Bank of Canada. Data are continuously compounded and multiplied by 100. Row 1 displays 1-, 2-, and 3-month-ahead returns, respectively. Row 2 displays 4-, 5-, and 6-month-ahead returns, respectively. Row 3 displays 7-, 8-, and 9-month-ahead returns, respectively.

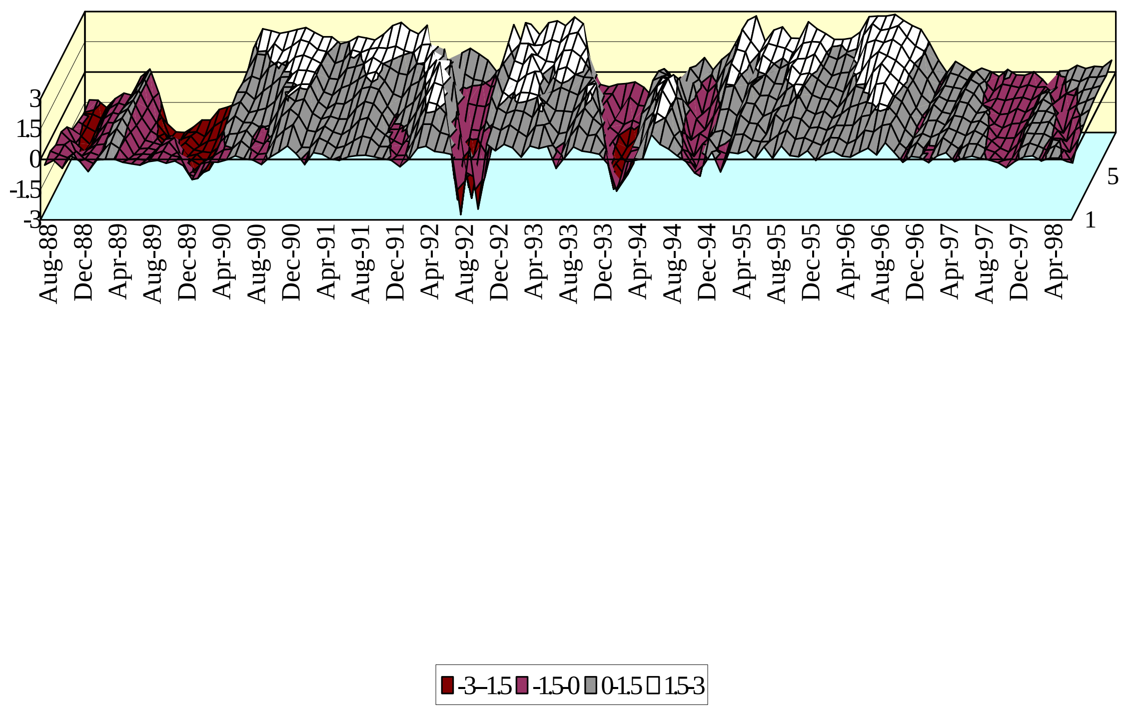


FIGURE 3.1 – Part 2: Excess Forward Returns on Canadian 90-day Bankers' Acceptances, August 1988 to June 1998.

Source: Bank of Canada. Data are continuously compounded and multiplied by 100. The *x*-axis represents the strike date for the forward contract, the *y*-axis represents months until settlement, and the *z*-axis represents the realized excess forward return.

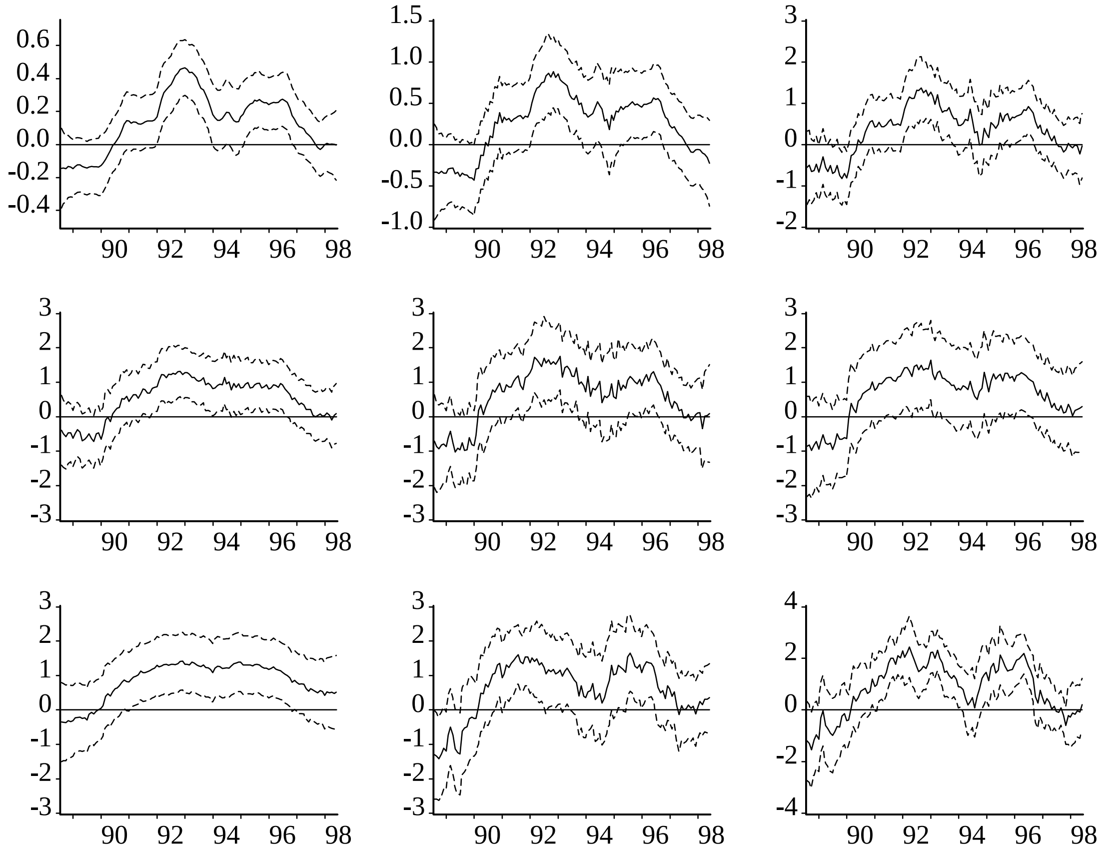


FIGURE 3.2 – Part 1: Term Premia Inferences for the Random Walk Specification, August 1988 to June 1998.

Source: Smoothed (two-sided) inferences from Kim's (1994) filter, modified for independent switching. Row 1 displays term premia on 1-, 2-, and 3-month-ahead forward rate agreements, respectively. Row 2 displays term premia on 4-, 5-, and 6-month-ahead forward rate agreements, respectively. Row 3 displays term premia on 7-, 8-, and 9-month-ahead forward rate agreements, respectively. The dashed lines denote 95 percent confidence bands.

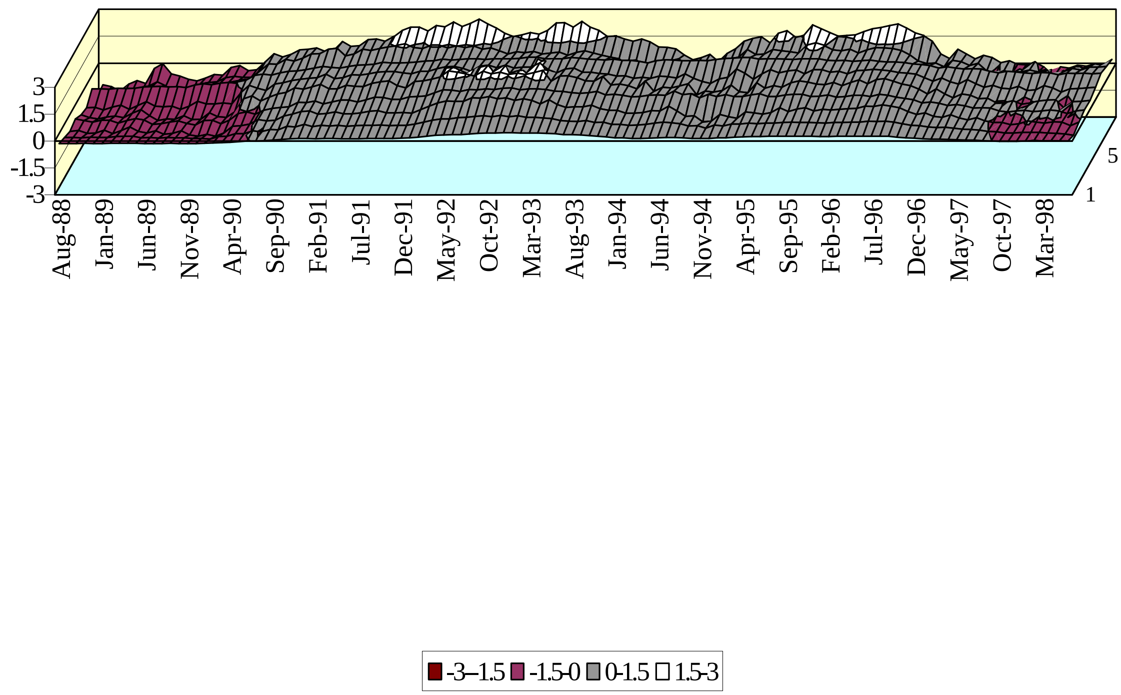


FIGURE 3.2 – Part 2: Term Premia Inferences for the Random Walk Specification, August 1988 to June 1998.

Source: Smoothed (two-sided) inferences from Kim's (1994) filter, modified for independent switching. The *x*-axis represents the strike date for the forward contract, the *y*-axis represents months until settlement, and the *z*-axis represents the estimated term premium.

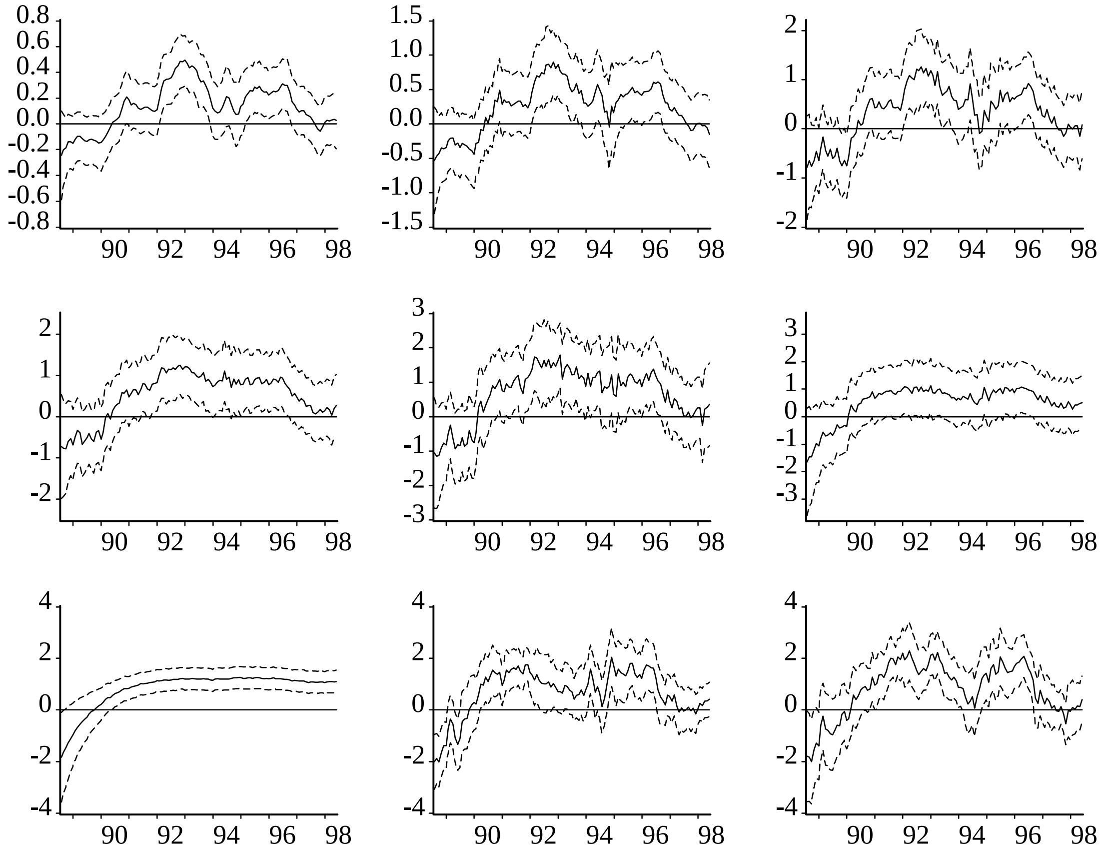


FIGURE 3.3 – Part 1: Term Premia Inferences for the Stationary AR(1) Specification, August 1988 to June 1998.

Source: Smoothed (two-sided) inferences from Kim's (1994) filter, modified for independent switching. Row 1 displays term premia on 1-, 2-, and 3-month-ahead forward rate agreements, respectively. Row 2 displays term premia on 4-, 5-, and 6-month-ahead forward rate agreements, respectively. Row 3 displays term premia on 7-, 8-, and 9-month-ahead forward rate agreements, respectively. The dashed lines denote 95 percent confidence bands.

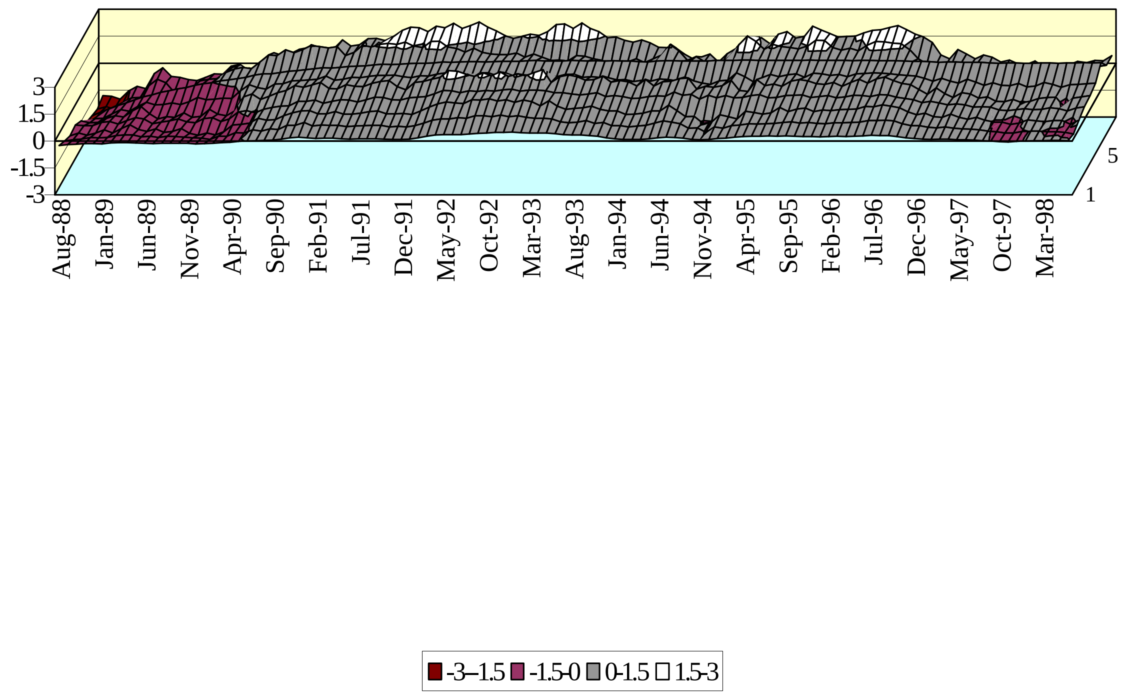


FIGURE 3.3 – Part 2: Term Premia Inferences for the Stationary AR(1) Specification, August 1988 to June 1998.

Source: Smoothed (two-sided) inferences from Kim's (1994) filter, modified for independent switching. The *x*-axis represents the strike date for the forward contract, the *y*-axis represents months until settlement, and the *z*-axis represents the estimated term premium.

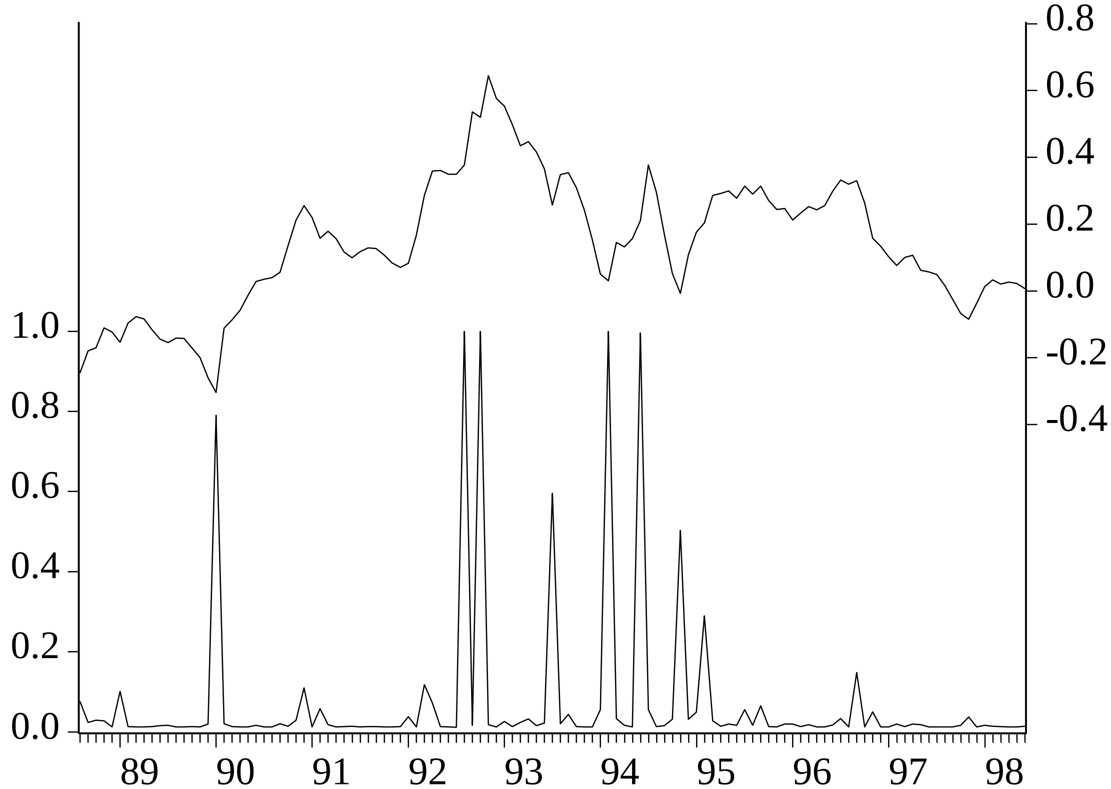


FIGURE 3.4: Term Premium Inferences for the AR(1) with Switching Intercept Specification and the Probability of a High Volatility Regime, August 1988 to June 1998.

Source: Smoothed (two-sided) inferences from Kim's (1994) filter, modified for independent switching. The term premium is on a 1-month-ahead forward rate agreement.

List of References

Abel, A.B., 1988, Stock prices under time-varying dividend risk: An exact solution in an infinite-horizon general equilibrium model, Journal of Monetary Economics 22, 375-393.

Andrews, D.W.K., 1993, Tests for parameter instability and structural change with unknown change point, Econometrica 61, 821-856.

Backus, D.K. and A.W. Gregory, 1993, Theoretical relations between risk premiums and conditional variances, Journal of Business and Economic Statistics 11, 177-185.

Bekaert, G., R. Hodrick, and D. Marshall, 1997, 'Peso problem' explanations for term structure anomalies, Federal Reserve Bank of Chicago Working Paper 97-7.

Bekaert, G. and G. Wu, 1999, Asymmetric volatility and risk in equity markets, Review of Financial Studies, forthcoming.

Black, F., 1976, Studies of stock price volatility changes, Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statisitics Section, 177-181.

Bollerslev, T., R.F. Engle, and J.M. Wooldrige, 1988, A capital asset pricing model with time-varying covariances, Journal of Political Economy 96, 116,131.

Box, G. and D. Pierce, 1970, Distribution of residual autocorrelations in autoregressive integrated moving average time series models, Journal of the American Statistical Association 65, 1509-1526.

Breen, W., L.R. Glosten, and R. Jagannathan, 1989, Economic significance of predictable variations in stock index returns, Journal of Finance 44, 1177-1189.

Campbell, J.Y., 1987, Stock returns and the term structure, Journal of Financial Economics 18, 373-399.

Campbell, J.Y., 1991, A variance decomposition for stock returns, Economic Journal 101, 157-179.

Campbell, J.Y. and L. Hentschel, 1992, No news is good news: An asymmetric model of changing volatility in stock returns, Journal of Financial Economics 31, 281-318.

Campbell, J.Y. and R. Shiller, 1988a, The dividend-price ratio and expectations of future dividends and discount factors, Review of Financial Studies 1, 195-227.

Campbell, J.Y. and R. Shiller, 1988b, Stock prices, earnings, and expected dividends, Journal of Finance 43, 661-676.

Campbell, J.Y. and R. Shiller, 1991, Yield spreads and interest rate movements: A bird's eye view, Review of Economic Studies 58, 495-514.

Cecchetti, S.G., P. Lam, and N.C. Mark, 1990, Mean reversion in equilibrium asset prices, American Economic Review 80, 398-418.

Christie, A., 1982, The stochastic behavior of common stock variances – value, leverage and interest rate effects, Journal of Financial Economics 10, 407-432.

Clinton, K. and M. Zelmer, 1997, Constraints on the conduct of Canadian monetary policy in the 1990s: Dealing with uncertainty in financial markets, Bank of Canada Technical Report #80.

Cook, T. and T. Hahn, 1990, Interest rate expectations and the slope of the money market yield curve, Federal Reserve Bank of Richmond Economic Review 76, 3-26.

Cooley, T.F. and E.C. Prescott, 1976, Estimation in the presence of stochastic parameter variation, Econometrica 44, 167-184.

DeGennaro, R. and J. Moser, 1989, Variability and stationarity of term premia, Federal Reserve Bank of Chicago Working Paper 89-16.

Dickey, D. and W. Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica 49, 1057-1072.

Doan, T., R.B. Litterman, and C.A. Sims, 1984, Forecasting and conditional projection using realistic prior distributions, Econometric Reviews 3, 1-100.

Engel, C. and J.D. Hamilton, 1990, Long swings in the dollar: Are they in the data and do markets know it? American Economic Review 80, 689-713.

Engle, R., D. Lilien, and R. Robins, 1987, Estimating time-varying premia in the term structure: The ARCH-M model, Econometrica 55, 251-77.

Engle, R.F. and M. Watson, 1987, Kalman filter: Applications to forecasting and rational-expectation models, in Advances in Econometrics, Fifth World Congress, Vol. 1, 245-281.

Evans, M. and K. Lewis, 1994, Do stationary risk premia explain it all? Journal of Monetary Economics 33, 285-318.

Fama, E.F., 1970, Efficient capital markets: A review of theory and empirical work, Journal of Finance 25, 383-416.

Fama, E.F., 1984, The information in the term structure, Journal of Financial Economics 13, 509-528.

Fama, E.F. and R. Bliss, 1987, The information in long maturity forward rates, American Economic Review 77, 680-692.

Fama, E.F. and K.R. French, 1988, Permanent and temporary components of stock prices, Journal of Political Economy 96, 246-273.

Fama, E.F. and G.W. Schwert, 1977, Asset returns and inflation, Journal of Financial Economics 5, 115-146.

French, K.R., G.W. Schwert, and R.F. Stambaugh, 1987, Expected stock returns and volatility, Journal of Financial Economics 19, 3-29.

Froot, K., 1989, New hope for the Expectations Hypothesis of the term structure of interest rates, Journal of Finance 44, 283-305.

Garbade, K., 1977, Two methods for examining the stability of regression coefficients, Journal of the American Statistical Association 72, 54-63.

Garbade, K. and P. Wachtel, 1978, Time variation in the relationship between inflation and interest rates, Journal of Monetary Economics 4, 755-765.

Garcia, R., 1995, Asymptotic null distribution of the likelihood ratio test in Markov switching models, Manuscript, University of Montreal.

Gennotte, G. and T.A. Marsh, 1993, Variations in economic uncertainty and risk premiums on capital assets, European Economic Review 37, 1021-1041.

Glosten, L.R., R. Jagannathan, and D.E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance 48, 1779-1801.

Gravelle, T. and J. Morley, 1999, Testing the Expectations Hypothesis with a state-space model of the term structure, Unpublished Manuscript, University of Washington.

Gravelle, T., P. Muller, and D. Stréliski, 1998, Towards a new measure of interest rate expectations in Canada: Estimating a time-varying term premium, Bank of Canada Working Paper.

Hamilton, J.D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica 57, 357-384.

Hamilton, J.D., 1994a, Time Series Analysis (Princeton University Press).

Hamilton, J.D., 1994b, State-space models, in: R.F. Engle and D.L. McFadden, eds., Handbook of Econometrics, Vol. 4 (North-Holland, Amsterdam) 3014-3077.

Hamilton, J.D. and R. Susmel, 1994, Autoregressive conditional heteroskedasticity and changes in regime, Journal of Econometrics 64, 307-333.

Hansen, B.E., 1992, The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP, Journal of Applied Econometrics 7, S61-S82.

Hardouvelis, G., 1994, The term structure spread and future changes in long and short rates in the G7 countries, Journal of Monetary Economics 33, 255-283.

Harvey, A.C., 1990, The Econometric Analysis of Time Series, 2nd ed. (MIT Press, Cambridge).

Harvey, A.C., E. Ruiz, and E. Sentana, 1992, Unobserved component time series models with ARCH disturbances, Journal of Econometrics 52, 129-157.

Harvey, C.R., 1989, Time-varying conditional covariances in tests of asset pricing models, Journal of Financial Economics 24, 289-317.

Iyer, S., 1996, Time-varying term premia and the behaviour of forward interest rate prediction errors, Journal of Financial Research 20, 503-507.

Jarque, C. and A. Bera, 1980, Efficient tests for Normality, heteroskedasticity, and serial independence of regression residuals, Economics Letters 6, 255-259.

Jegadeesh, N., 1991, Seasonality in stock price mean reversion: Evidence form the U.S. and the U.K., Journal of Finance 46, 1427-1444.

Jensen, M.C. (ed.), 1978, Some anomalous evidence regarding market efficiency, Journal of Financial Economics 6, 95-101.

Kendall, M., 1954, Note on bias in estimation of autocorrelation, Biometrica 41, 403-404.

Kendall, M. and A. Stuart, 1973, The Advanced Theory of Statistics, Vol. 2, 3rd ed. (Hafner Publishing Co., New York).

Kim, C.-J., 1993, Sources of monetary growth uncertainty and Economic activity: The time-varying-parameter model with heteroskedastic disturbances, Review of Economics and Statistics 75, 483-492.

Kim, C.-J., 1994, Dynamic linear models with Markov-switching, Journal of Econometrics 60, 1-22.

Kim, C.-J., J.C. Morley, and C.R. Nelson, 1998, Do changes in the market risk premium explain the empirical evidence of mean reversion in stock prices? Unpublished Manuscript, University of Washington.

Kim, C.-J., J.C. Morley, and C.R. Nelson, 1999, A Markov-switching model of stock returns with learning and volatility feedback, Unpublished Manuscript, University of Washington.

Kim, C.-J. and C.R. Nelson, 1998, Testing for mean reversion in heteroskedatic data II: Autoregression tests based on Gibbs-sampling-augmented randomization, Journal of Empirical Finance 5, 385-396.

Kim, C.-J. and C.R. Nelson, 1999, State-Space Model with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications (MIT Press, Cambridge).

Kim, C.-J., C.R. Nelson, and R. Startz, 1998, Testing for mean reversion in heteroskedastic data based on Gibbs-sampling-augmented randomization, Journal of Empirical Finance 5, 131-154.

Kim, M.-J., C.R. Nelson, and R. Startz, 1991, Mean reversion in stock prices? A reappraisal of the empirical evidence, Review of Economic Studies 48, 515-528.

Lutz, F., 1940, The structure of interest rates, Quarterly Journal of Economics 55, 36-63.

Mankiw, G., 1986, The term structure of interest rates revisited, Brookings Papers on Economic Activity 1, 61-91.

Mankiw, G. and J. Miron, 1986, The changing behaviour of the term structure of interest rates, Quarterly Journal of Economics 101, 211-242.

Marriot, F.H.C. and J.A. Pope, 1954, Bias in estimation of autocorrelations, Biometrica 41, 390-402.

McQueen, G., 1992, Long-horizon mean-reverting stock prices revisited, Journal of Financial and Quantitative Analysis 27, 1-18.

Merton, R.C., 1980, On estimating the expected return on the market: An exploratory investigation, Journal of Financial Economics 8, 323-361.

Modigliani, F. and R. Sutch, 1966, Innovations in interest rate policy, American Economic Review 56, 178-197.

Muth, J., 1960, Optimal properties of exponentially weighted forecasts, Journal of the American Statistical Association 55, 299-306.

Nelson, D., 1991, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59, 347-370.

Nyblom, J., 1989, Testing for the constancy of parameters over time, Journal of the American Statistical Association 84, 223-230.

Officer, R.R., 1973, The variability of the market factor of the New York Stock Exchange, Journal of Business 46, 434-453.

Pagan, A.R. and G.W. Schwert, 1990, Alternative models of conditional stock volatility, Journal of Econometrics 45, 267-290.

Pindyck, R.S., 1984, Risk, inflation, and the stock market, American Economic Review 45, 267-290.

Poterba, J.M. and L.H. Summers, 1988, Mean reversion in stock prices: Evidence and implications, Journal of Financial Economics 22, 27-59.

Richardson, M., 1993, Temporary components of stock prices: A skeptic's view, Journal of Business and Economic Statistics 11, 199-207.

Richardson, M. and T. Smith, 1994, A unified approach to testing for serial correlation in stock returns, Journal of Business 67, 371-399.

Richardson, M. and J.H. Stock, 1989, Drawing inferences from statistics based on multiyear asset returns, Journal of Financial Economics 25, 323-348.

Schaller, H. and S. van Norden, 1997, Regime switching in stock market returns, Applied Financial Economics 7, 177-191.

Schwert, G.W., 1989a, Why does stock market volatility change over time? Journal of Finance 44, 1115-1153.

Schwert, G.W., 1989b, Business cycles, financial crises, and stock volatility, Carnegie-Rochester Conference Series on Public Policy 31, 83-126.

Schwert, G.W., 1990, Indexes of United States stock prices from 1802 to 1987, Journal of Business 63, 399-426.

Shiller, R., 1979, The volatility of long-term interest rates and expectations models of the term structure, Journal of Political Economy 87, 1190-1219.

Shiller, R., 1990, The term structure of interest rates, in: B. Friedman and F. Hahn, eds., Handbook of Monetary Economics, Vol. 1 (North-Holland, Amsterdam) 627-723.

Shiller, R., J. Campbell, and K. Schoenholtz, 1983, Forward rates and future policy: Interpreting the term structure of interest rates, Brookings Papers on Economic Activity 1, 173-217.

Shleifer, A. and L.H. Summers, 1990, The noise trader approach to finance, Journal of Economic Perspectives 4, 19-33.

Startz, R., 1982, Do forecast errors or term premia really make the difference between long and short rates? Journal of Financial Economics 10, 323-329.

Stock, J.H. and M.W. Watson, 1996, Evidence on structural instability in macroeconomic time series relations, Journal of Business and Economic Statistics 14, 11-30.

Stock, J.H. and M.W. Watson, 1998, Median unbiased estimation of coefficient variance in a time-varying parameter model, Journal of the American Statistical Association 93, 349-358.

Stréliski, D., 1998, L'utilisation des accords de taux futur, Bank of Canada Review.

Summers, L.H., 1986, Does the stock market rationally reflect fundamental values? Journal of Finance 41, 591-601.

Turner, C.M., R. Startz, and C.R. Nelson, 1989, A Markov model of heteroskedasticity, risk, and learning in the stock market, Journal of Financial Economics 25, 3-22.

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48, 817-838.

Whitelaw, R.F., 1994, Time variations and covariations in the expectation and volatility of stock market returns, Journal of Finance 49, 515-541.

Zelmer, M., 1996, Strategies versus tactics for monetary policy operations, in: Money Markets and Central Bank Operations, Proceedings of a Conference Held by the Bank of Canada, November 1995, 211-260.

Zivot, E. and D.W.K. Andrews, 1992, Further evidence on the Great Crash, the oil price shock, and the unit root hypothesis, Journal of Business and Economic Statistics 10, 251-270.

Appendix A

A.1 Time-Varying Coefficient Model

The Kalman filter for the model specified in equations (1.2) and (1.3) is given by the following six equations (for simplicity, let $b_t \equiv b_t(k)$, $y_t \equiv r_t - m$ and $x_t \equiv \sum_{j=1}^{k} y_{t-j}$):

<u>Prediction</u>

$$b_{t|t-1} = b_{t-1|t-1} \tag{A.1}$$

$$P_{t|t-1} = P_{t-1|t-1} + s_v^2 \tag{A.2}$$

$$h_{t|t-1} \equiv y_t - y_{t|t-1} = y_t - b_{t|t-1} x_t \tag{A.3}$$

$$f_{t|t\text{-}1} = x_t P_{t|t-1} x_t' + s_e^2 \tag{A.4}$$

<u>Updating</u>

$$b_{t|t} = b_{t|t-1} + K_t h_{t|t-1} \tag{A.5}$$

$$P_{t|t} = P_{t|t-1} - K_t x_t P_{t|t-1}, \tag{A.6}$$

where $b_{t|t-1} \equiv E[b_t | y_{t-1}]$, for example, is the expectation of b_t conditional on information up to time $t-1$; $P_{t|t-1}$ is the variance of $b_{t|t-1}$; $f_{t|t-1}$ is the variance of $h_{t|t-1}$; and $K_t \equiv P_{t|t-1} x_t' f_{t|t-1}^{-1}$ is the Kalman gain.[1]

Given some initial values $b_{0|0}$ and $P_{0|0}$, we iterate through (A.1)-(A.6) for $t = 1, \ldots, T$ to obtain filtered inferences about b_t conditional on information up to time *t*. Also, as a by-product of this procedure, we obtain $h_{t|t-1}$ and $f_{t|t-1}$, which based on the prediction error decomposition (Harvey, 1990) can be used to find maximum likelihood estimates of the hyper-parameters as follows:

---

[1] For a more general discussion of the Kalman filter and time-varying parameter models, as well as details on the derivation of the Kalman gain, refer to Hamilton (1994a,b) and Kim and Nelson (1998).

$$\max_{q} l(q) = -\frac{1}{2}\sum_{t=t+1}^{T}\ln(2pf_{t|t-1}) - \frac{1}{2}\sum_{t=t+1}^{T} h_{t|t-1} f_{t|t-1}^{-1} h_{t|t-1} \,, \tag{A.7}$$

where $q = (m, s_e, s_v)$.

Note that we ignore the first t observations in calculating the likelihood function. Since we do not observe b_0, and it has no unconditional expectation under the random walk specification given in equation (3), we must make an arbitrary guess as to its value and assign our guess an extremely large variance (e.g., $b_{0|0} = 0$ and $P_{0|0} \approx \infty$). We then use the first t observations to determine $b_{t|t}$ and $P_{t|t}$, which we treat as the initial values in the Kalman filter for the purposes of MLE.[2] In practice, there is no exact rule as to what value of t to use. Roughly speaking, we choose t such that the effects of our arbitrary initial guess are minimized subject to including as much data in estimation as possible. The adjusted samples given in Tables 1.3 and 1.4 reflect our choices for t . Notably, the reported estimates are robust to larger values of t .

Finally, given $b_{T|T}$ and $P_{T|T}$ from the last iteration of the Kalman filter, we iterate backwards through the following two equations in order to obtain smoothed inferences about b_t conditional on information up to time *T*:

<u>Smoothing</u>

$$b_{t|T} = b_{t|t} + P_{t|t} P_{t+1|t}^{-1} (b_{t+1|T} - b_{t+1|t}) \tag{A.8}$$

$$P_{t|T} = P_{t|t} + P_{t|t} P_{t+1|t}^{-1} (P_{t+1|T} - P_{t+1|t}) P_{t+1|t}'^{-1} P_{t|t}' \,. \tag{A.9}$$

[2] Alternatively, we could treat the initial value as a hyper-parameter to be estimated by MLE. The filtered inferences about the time-varying coefficient are very similar in both cases. However, since the hyper-parameters are treated as known, the standard error bands surrounding the inferences in this alternative case would dramatically understate the true degree of uncertainty during the early part of the sample. This is precisely when our uncertainty should be greatest.

A.2 Time-Varying Coefficient and Markov-Switching Variance Model

The Kalman filter for the model specified in equations (1.2) and (1.3) with the Markov-switching variance assumption for the error term in equation (1.2) is given by the following six equations (conditional on $S_{t-1} = i$ and $S_t = j$, where $i = 0{,}1$ and $j = 0{,}1$):

<u>Prediction</u>

$$b_{t|t-1}^{(i)} = b_{t-1|t-1}^{(i)} \tag{A.10}$$

$$P_{t|t-1}^{(i)} = P_{t-1|t-1}^{(i)} + s_v^2 \tag{A.11}$$

$$h_{t|t-1}^{(i)} = y_t - b_{t|t-1}^{(i)} x_t \tag{A.12}$$

$$f_{t|t-1}^{(i,j)} = x_t P_{t|t-1}^{(i)} x_t' + s_{ej}^2 \tag{A.13}$$

<u>Updating</u>

$$b_{t|t}^{(i,j)} = b_{t|t-1}^{(i)} + K_t^{(i,j)} h_{t|t-1}^{(i)} \tag{A.14}$$

$$P_{t|t}^{(i,j)} = P_{t|t-1}^{(i)} - K_t^{(i,j)} x_t P_{t|t-1}^{(i)}, \tag{A.15}$$

where $K_t^{(i,j)} \equiv P_{t|t-1}^{(i)} x_t' f_{t|t-1}^{(i,j)-1}$.

For this model, Hamilton's (1989) filter is then given by the following three steps:

<u>Step 1:</u> Calculate the conditional density of y_t and the joint probability of S_t and S_{t-1}:

$$f(y_t | S_t = j, S_{t-1} = i, y_{t-1}) = \frac{1}{\sqrt{2p f_{t|t-1}^{(i,j)}}} \exp\left\{ -\frac{1}{2} h_{t|t-1}'^{(i)} f_{t|t-1}^{(i,j)-1} h_{t|t-1}^{(i)} \right\} \tag{A.16}$$

$$\Pr[S_t = j, S_{t-1} = i | y_{t-1}] = \Pr[S_t = j | S_{t-1} = i] \cdot \Pr[S_{t-1} = i | y_{t-1}], \tag{A.17}$$

where $\Pr[S_t = j | S_{t-1} = i]$ is the transition probability.

Step 2: Calculate the joint density of y_t, S_t, and S_{t-1} and collapse across all possible states to find the marginal density of y_t:

$$f(y_t, S_t = j, S_{t-1} = i | y_{t-1}) = f(y_t | S_t = j, S_{t-1} = i, y_{t-1}) \cdot \Pr[S_t = j, S_{t-1} = i | y_{t-1}] \quad (A.18)$$

$$f(y_t | y_{t-1}) = \sum_{j=0}^{1} \sum_{i=0}^{1} f(y_t, S_t = j, S_{t-1} = i | y_{t-1}) . \quad (A.19)$$

Step 3: Update the joint probability of S_t and S_{t-1} given y_t and collapse across different possible values of S_{t-1}:

$$\Pr[S_t = j, S_{t-1} = i | y_t] = \frac{f(y_t, S_t = j, S_{t-1} = i | y_{t-1})}{f(y_t | y_{t-1})} \quad (A.20)$$

$$\Pr[S_t = j | y_t] = \sum_{i=0}^{1} \Pr[S_t = j, S_{t-1} = i | y_t] . \quad (A.21)$$

Then, to complete the Kalman filter in equations (A.10)-(A.15) given the updated probabilities in equations (A.20) and (A.21), we collapse $b_{t|t}^{(i,j)}$ and $P_{t|t}^{(i,j)}$ across different possible values of S_{t-1} as follows:

$$b_{t|t}^{(j)} = \frac{\sum_{i=0}^{1} \Pr[S_t = j, S_{t-j} = i | y_t] \cdot b_{t|t}^{(i,j)}}{\Pr[S_t = j | y_t]} \quad (A.22)$$

$$P_{t|t}^{(j)} = \frac{\sum_{i=0}^{1} \Pr[S_t = j, S_{t-j} = i | y_t] \cdot \{P_{t|t}^{(i,j)} + (b_{t|t}^{(j)} - b_{t|t}^{(i,j)})(b_{t|t}^{(j)} - b_{t|t}^{(i,j)})'\}}{\Pr[S_t = j | y_t]} . \quad (A.23)$$

Our approach to estimation of this model is the same as in the previous subsection. Given some initial values $b_{0|0}^{(i)}$, $P_{0|0}^{(i)}$, and $\Pr[S_0 = i]$, we iterate through (A.10)-(A.23) for $t = 1,\ldots,T$ to obtain filtered inferences about b_t conditional on information up to time t. Specifically, given $b_{t|t}^{(j)}$ and $P_{t|t}^{(j)}$ in equations (A.22) and (A.23), we obtain $b_{t|t}$ and its variance $P_{t|t}$ by collapsing across different possible values of S_t as follows:

$$\mathbf{b}_{t|t} = \sum_{j=0}^{1} \Pr[S_t = j|y_t] \cdot \mathbf{b}_{t|t}^{(j)} \tag{A.24}$$

$$P_{t|t} = \sum_{j=0}^{1} \Pr[S_t = j|y_t] \cdot \{P_{t|t}^{(j)} + (\mathbf{b}_{t|t} - \mathbf{b}_{t|t}^{(j)})(\mathbf{b}_{t|t} - \mathbf{b}_{t|t}^{(j)})'\}\,. \tag{A.25}$$

Also, we use the marginal density of y_t given in equation (A.17) to find maximum likelihood estimates of the hyper-parameters as follows:

$$\max_{q} l(q) = \sum_{t=t+1}^{T} \ln(f(y_t|y_{t-1}))\,, \tag{A.26}$$

where $q = (s_v, s_{e0}, s_{e1}, q, p, m)$.

Note that we continue to ignore the first t observations in calculating the likelihood function. Again, due to the random walk specification in equation (1.3), we need to make an arbitrary guess as to the value of $b_0^{(i)}$ (e.g., $b_{0|0}^{(i)} = 0$ and $P_{0|0}^{(i)} \approx \infty$). Then, as before, we use the first t observations to determine $b_{t|t}^{(i)}$ and $P_{t|t}^{(i)}$, which we treat as the initial values in the Kalman filter for the purposes of MLE. The initial probability of each state, on the other hand, is simply given by its unconditional probability:

$$\Pr[S_0 = 0] = \frac{1-p}{2-p-q} \tag{A.27}$$

$$\Pr[S_0 = 1] = \frac{1-q}{2-p-q}. \tag{A.28}$$

In addition to making the above inferences, we also obtain the smoothed probability $\Pr[S_t = 1|y_T]$ by employing Kim's (1994) smoothing algorithm. Specifically, given the probability from (A.21), including $\Pr[S_T = j|y_T]$ from the last iteration of the above filter, and the filtered probability $\Pr[S_t = j|y_{t-1}]$, which can be found from (A.17) by collapsing across S_{t-1}, we iterate backwards through the following two equations (conditional on $S_t = j$ and $S_{t+1} = l$, where $j = 0,1$ and $l = 0,1$):

$$\Pr[S_{t+1} = l, S_{t+1} = j|y_T] = \frac{\Pr[S_{t+1} = l|y_T] \cdot \Pr[S_t = j|y_t] \cdot \Pr[S_{t+1} = l|S_t = j]}{\Pr[S_{t+1} = l|y_t]} \tag{A.29}$$

$$\Pr[S_{t+1} = j|y_T] = \sum_{l=0}^{1} \Pr[S_{t+1} = l, S_{t+1} = j|y_T]. \tag{A.30}$$

Finally, we obtain smoothed inferences about b_t conditional on information up to time T by iterating backwards through the following two equations and using the smoothed probabilities given in equations (A.29) and (A.30) to collapse across states as in equations (A.22)-(A.25):

<u>Smoothing</u>

$$b_{t|T}^{(j,l)} = b_{t|t}^{(j)} + \tilde{P}_t^{(j,l)}(b_{t+1|T}^{(l)} - b_{t+1|t}^{(j,l)}) \tag{A.31}$$

$$P_{t|T}^{(j,l)} = P_{t|t}^{(j)} + \tilde{P}_t^{(j,l)}(P_{t+1|T}^{(l)} - P_{t+1|t}^{(j,l)})\tilde{P}_t'^{(j,l)}, \tag{A.32}$$

where $\tilde{P}_t^{(j,l)} \equiv P_{t|t}^{(j)} P_{t+1|t}^{(j,l)-1}$.

A.3 Markov-Switching Risk Premium with Learning Model

For the model specified in equations (1.4) and (1.5) with the Markov-switching variance assumption for the error term in equation (1.4), Hamilton's (1989) filter is given by a more complicated version of the same three steps presented in the previous subsection:

<u>Step 1a:</u> Calculate the joint probability of S_t, S_{t-1}, S_{t-k}, and $S_t^* \equiv \sum_{j=2}^{k-1} S_{t-j}$ and solve for $\Pr[S_t = 1 | y_{t-1}]$ by summing across all possible values of S_{t-1}, S_{t-k}, and S_t^*:

$$\Pr[S_t = j, S_{t-1} = i, S_{t-k} = h, S_t^* = m | y_{t-1}] = \Pr[S_t = j | S_{t-1} = i] \cdot \Pr[S_{t-1} = i, S_{t-k} = h, S_t^* = m | y_{t-1}] \quad \text{(A.33)}$$

$$\Pr[S_t = 1 | y_{t-1}] = \sum_{i=0}^{1} \sum_{h=0}^{1} \sum_{m=0}^{k-2} \Pr[S_t = 1, S_{t-1} = i, S_{t-k} = h, S_t^* = m | y_{t-1}]. \quad \text{(A.34)}$$

<u>Step 1b:</u> Calculate the conditional density of r_t:

$$f(r_t | S_t = j, S_{t-1} = i, S_{t-k} = h, S_t^* = m, y_{t-1}) = \frac{1}{\sqrt{2ps_{S_t}^2}} \exp\left\{ -\frac{1}{2s_{S_t}^2} (r_{S_t}^* - b(k) \sum_{j=1}^{k} r_{S_{t-j}}^*)^2 \right\}, \quad \text{(A.35)}$$

where $r_{S_t}^* \equiv r_t - m_0 - m_1 S_t - g \cdot \Pr[S_t = 1 | y_{t-1}]$. (Note: given r_t, S_t, S_{t-1}, S_{t-k}, S_t^*, y_{t-1}, and particular values for the parameters, we observe all the elements of (A.35) since

$$\sum_{j=1}^{k} r_{S_{t-j}}^* = \sum_{j=1}^{k} r_{t-j} - m_0 \cdot k - m_1 \sum_{j=1}^{k} S_{t-j} - g \sum_{j=1}^{k} \Pr[S_{t-j} = 1 | y_{t-j-1}]$$

and $\sum_{j=1}^{k} S_{t-j} = S_{t-1} + S_t^* + S_{t-k}$).

Step 2: Calculate the joint density of r_t, S_t, S_{t-1}, S_{t-k}, and S_t^* and collapse across all possible states to find the marginal density of r_t:

$$f(r_t, S_t = j, S_{t-1} = i, S_{t-k} = h, S_t^* = m | y_{t-1}) = \\ f(r_t | S_t = j, S_{t-1} = i, S_{t-k} = h, S_t^* = m, y_{t-1}) \times \\ \Pr[S_t = j, S_{t-1} = i, S_{t-k} = h, S_t^* = m | y_{t-1}] \quad \text{(A.36)}$$

$$f(r_t | y_{t-1}) = \sum_{j=0}^{1} \sum_{i=0}^{1} \sum_{h=0}^{1} \sum_{m=0}^{k-2} f(r_t, S_t = j, S_{t-1} = i, S_{t-k} = h, S_t^* = m | y_{t-1}). \quad \text{(A.37)}$$

Step 3: Update the joint probability of S_t, S_{t-1}, S_{t-k}, and S_t^* given r_t and solve for the joint probability of S_t, S_{t-k+1}, and S_{t+1}^*:

$$\Pr[S_t = j, S_{t-1} = i, S_{t-k} = h, S_t^* = m | y_t] = \\ \frac{f(r_t, S_t = j, S_{t-1} = i, S_{t-k} = h, S_t^* = m | y_{t-1})}{f(r_t | y_{t-1})} \quad \text{(A.38)}$$

$$\Pr[S_t = j, S_{t-k+1} = g, S_{t+1}^* = n | y_t] = \\ \sum_{i=0}^{1} \sum_{h=0}^{1} \Pr[S_t = j, S_{t-1} = i, S_{t-k} = h, S_{t-k+1} = g, S_{t+1}^* = n | y_t], \quad \text{(A.39)}$$

where

$$\Pr[S_t = j, S_{t-1} = i, S_{t-k} = h, S_{t-k+1} = g, S_t^* = m | y_t] = \\ \Pr[S_{t-k} = h | S_{t-k+1} = g] \times \Pr[S_t = j, S_{t-1} = i, S_{t-k} = h, S_t^* = m | y_t] \\ \Rightarrow \Pr[S_t = j, S_{t-1} = i, S_{t-k} = h, S_{t-k+1} = g, S_{t+1}^* = n | y_t]$$

since $S_{t+1}^* = S_{t-1} + S_t^* - S_{t-k+1}$.

Our approach is much the same as before. Given $\Pr[S_{-i^*} = 1]$ for $i^* = 0, \ldots, k-1$ and $\Pr[S_0 = i, S_{-k+1} = h, S_1^* = m]$ for $i = 0,1$, $h = 0,1$, and $m = 0, \ldots, k-1$, we iterate through (A.33)-(A.39) for $t = 1, \ldots, T$ to obtain the filtered probability $\Pr[S_t = 1 | y_{t-1}]$. Also, as in the previous section, we use the unconditional probability given in (A.28) for each $\Pr[S_{-i^*} = 1]$. As for $\Pr[S_0 = i, S_{-k+1} = h, S_1^* = m]$, deriving unconditional joint probabilities in terms of *p* and *q* is impractical for large values of *k*. Instead, we treat these initial probabilities, denoted $\Pi = (p_1, \ldots, p_{4(k-1)})$, as $4 \times (k-1)$ additional parameters to be estimated.

We use the marginal density of r_t given in equation (A.37) to find maximum likelihood estimates of the parameters as follows:

$$\max_q l(q) = \sum_{t=1}^{T} \ln(f(r_t | y_{t-1})), \tag{A.40}$$

where $q = (b, s_{e0}, s_{e1}, q, p, m_0, m_1, g, \Pi)$.[3] Note that in this case we are able to include all observations in calculating the likelihood function.

As before, we use Kim's (1994) smoothing algorithm given in (A.29) and (A.30) to obtain the smoothed probability $\Pr[S_t = 1 | y_T]$.

A.4 Forecast Comparison

For the four models presented in this paper, construction of *j*-period-ahead forecasts is somewhat complicated. Specifically, following Doan, Litterman, and Sims (1984), we need to employ an iterative procedure to calculate multi-period forecasts given a one-

[3] Since we are not particularly interested in making inferences about the startup probabilities, we do not report their estimates. Also, for practical reasons, we treat their values as known for calculation of asymptotic standard errors based upon second derivatives. We consider this approach reasonable since inferences about the other parameters are robust to the use of alternative startup probabilities, such as an equal probability for each possible initial state.

period dependent variable in equations (1.1), (1.2), and (1.4). For the basic Jegadeesh model, the law of iterated expectations implies that the resulting forecast represents $E[\sum_{i=1}^{j} r_{t+i} | y_t]$. However, for the extension models, the law of iterated expectations does not apply since multi-period forecasts are nonlinear functions of past information.

For all four models, the iterative approach works as follows. First, we find the one-period-ahead forecast implied by the model. For example, the basic Jegadeesh model given in equation (1.1) produces the following one-period forecast:

$$r_{t+1|t} = \hat{m} + \hat{b}(k)\sum_{i=0}^{k-1}(r_{t-i} - \hat{m}) . \tag{A.41}$$

Then, we treat the forecast $r_{t+1|t}$ as though it were actual realized data in order to obtain $r_{t+2|t}$. For the basic Jegadeesh model, this involves iterating equation (A.52) forward one period and substituting $r_{t+1|t}$ in for r_{t+1} on the right hand side. Repeating this process, we obtain the *j*-period-ahead forecast by simply summing up all of the one-period-ahead forecasts over the *j* periods as follows:

$$R_{t+j|t} = \sum_{i=1}^{j} r_{t+i|t} . \tag{A.42}$$

For the three extension models, the only difference in approach is how to calculate each one-period-ahead forecast. The time-varying coefficient model and time-varying coefficient and Markov-switching variance model produce the following one-period forecast:

$$r_{t+1|t} = \hat{m} + b_{t+1|t}\sum_{i=0}^{k-1}(r_{t-i} - \hat{m}) , \tag{A.43}$$

where, for the time-varying coefficient and Markov-switching variance model, $b_{t+1|t}$ is calculated from the collapsed $b_{t|t}$ given in equation (A.24). Again, we iterate the forecast equation forward one period to find $r_{t+2|t}$, and so forth, in order to obtain $R_{t+j|t}$. Note that the random walk specification of $b_t(k)$ conveniently means that $b_{t+i|t} = b_{t|t}$ for $i = 1, \ldots, j$.

Appendix B

B.1 The Equivalence of an MA(j-1) Process and the Sum of Changes in Expectations

For illustration, we present the case where $j = 2$. The results generalize to larger values of *j*.

Letting $g_i \equiv E[(e^{2,k}_{t,t+1} + e^{1,k}_{t+1,t+2})(e^{2,k}_{t+j,t+j+1} + e^{1,k}_{t+j+1,t+j+2})]$, the unconditional moments of the sum of changes in expectations when $j = 2$ are given as follows:

$$g_0 = s^2_{e2}(S_{t+1}) + s^2_{e1}(S_{t+2}), \tag{B.1}$$

$$g_1 = s_{e12}(S_{t+2}), \tag{B.2}$$

$$g_i = 0,\ i > 1. \tag{B.3}$$

That is, only the variance and the first-order autocorrelation coefficient are nonzero. Meanwhile, if we let $g_i \equiv E[(e_{t+2} + y_1 e_{t+1})(e_{t+j+2} + y_1 e_{t+j+1})]$, the unconditional moments for the MA(1) process are given as follows:

$$g_0 = s^2_e(S_{t+2}) + y_1{}^2 s^2_e(S_{t+1}), \tag{B.4}$$

$$g_1 = y_1 s^2_e(S_{t+2}), \tag{B.5}$$

$$g_i = 0,\ i > 1. \tag{B.6}$$

Again, only the variance and the first-order autocorrelation coefficient are nonzero. Therefore, the MA($j-1$) process represents a reduced-form model of the sum of changes in expectations.

B.2 Kim's (1994) Filter Modified for Independent Switching

For illustration, we present the state-space form and estimation details for the AR(1) with switching intercept specification and a *j*-period forward rate horizon. The other term premia specifications represent restricted cases of this specification.

Equations (3.2) and (3.5) imply the following observation equation:

$$f_t^{j,k} - r_{t+j}^k = \begin{bmatrix} 1 & 1 & y_1 & \cdots & y_{j-1} \end{bmatrix} \begin{bmatrix} q_t^{j,k} \\ e_{t+j} \\ e_{t+j-1} \\ \vdots \\ e_{t+1} \end{bmatrix}, \tag{B.7}$$

or, more compactly,

$$y_t = H\boldsymbol{b}_t. \tag{B.7'}$$

Equations (3.5), (3.12), and (3.13) imply the following state equation:

$$\begin{bmatrix} q_t^{j,k} \\ e_{t+j} \\ e_{t+j-1} \\ \vdots \\ e_{t+1} \end{bmatrix} = \begin{bmatrix} c(S_t) \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} f & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} q_{t-1}^{j,k} \\ e_{t+j-1} \\ e_{t+j-2} \\ \vdots \\ e_t \end{bmatrix} + \begin{bmatrix} v_t^{j,k} \\ e_{t+j} \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{B.8}$$

or, more compactly,

$$\boldsymbol{b}_t = m(S_t) + F\boldsymbol{b}_{t-1} + v_t. \tag{B.8'}$$

Equations (3.7) and (3.12) imply the following variance-covariance matrix for the state equation:

$$Q(S_t) = E[v_t v_t'] = \begin{bmatrix} s_{vj,k}^2 & 0 & 0 & \cdots & 0 \\ 0 & s_e^2(S_t) & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{B.9}$$

Then, in terms of the compact notation, the Kalman filter portion of Kim's (1994) filter is given as follows:

$$b_{t|t-1}(S_t) = m(S_t) + Fb_{t-1|t-1}, \tag{B.10}$$

$$P_{t|t-1}(S_t) = FP_{t-1|t-1}F' + Q(S_t), \tag{B.11}$$

$$h_{t|t-1}(S_t) \equiv y_t - Hb_{t|t-1}(S_t), \tag{B.12}$$

$$f_{t|t-1}(S_t) = HP_{t|t-1}(S_t)H', \tag{B.13}$$

$$b_{t|t}(S_t) = b_{t|t-1}(S_t) + K_t(S_t)h_{t|t-1}(S_t), \tag{B.14}$$

$$P_{t|t}(S_t) = P_{t|t-1}(S_t) - K_t(S_t)HP_{t|t-1}(S_t), \tag{B.15}$$

where $b_{t|t-1}(S_t)$ is the expectation of b_t conditional on information up to time *t-1* and S_t; $P_{t|t-1}(S_t)$ is the variance-covariance matrix of $b_{t|t-1}(S_t)$; $h_{t|t-1}(S_t)$ is the conditional forecast error; $f_{t|t-1}(S_t)$ is the variance-covariance matrix of $h_{t|t-1}(S_t)$; and $K_t(S_t) \equiv P_{t|t-1}(S_t)H'f_{t|t-1}(S_t)^{-1}$ is the Kalman gain.

The Hamilton (1989) filter portion of Kim's (1994) filter is given as follows:

$$f(y_t|S_t) = \frac{1}{\sqrt{2pf_{t|t-1}(S_t)}} \exp\left(-\frac{1}{2} \cdot \frac{h_{t|t-1}(S_t)^2}{f_{t|t-1}(S_t)} \right), \tag{B.16}$$

$$f(y_t) = (1-p) \cdot f(y_t|S_t = 0) + p \cdot f(y_t|S_t = 1), \tag{B.17}$$

with the updated probability that $S_t = i$, $i = \{0,1\}$ given by

$$\Pr[S_t = i | y_t] = \frac{f(y_t | S_t = i)}{f(y_t)}. \tag{B.18}$$

Then, to complete the modified version of Kim's (1994) filter, we need to collapse $b_{t|t}(S_t)$ and $P_{t|t}(S_t)$ across all states of the world. That is,

$$b_{t|t} = \sum_{i=0}^{1} \Pr[S_t = i | y_t] \cdot b_{t|t}(i), \tag{B.19}$$

$$P_{t|t} = \sum_{i=0}^{1} \Pr[S_t = i | y_t] \cdot \left\{ P_{t|t}(i) + (b_{t|t} - b_{t|t}(i))^2 \right\}. \tag{B.20}$$

Our approach to estimation of this model is to iterate through equations (B.10)-(B.20) for $t = 1,2,\ldots T$. Note that for starting up the filter, we use unconditional expectations of variables. Then, we use the marginal density function of y_t given in equation (A.17) to find maximum likelihood estimates of the parameters. That is,

$$\max_{\Pi} l(\Pi) = \sum_{t=1}^{T} \ln(f(y_t)), \tag{B.21}$$

where Π represents a vector of the parameters.

Finally, we obtain smoothed inferences by iterating the following two equations backwards from T to 1:

$$b_{t|T}(S_t) = b_{t|t} + \tilde{P}_t(S_t)(b_{t+1|T} - b_{t+1|t}), \tag{B.22}$$

$$P_{t|T}(S_t) = P_{t|t} + \tilde{P}_t(S_t)(P_{t+1|T} - P_{t+1|t}(S_t))\tilde{P}_t(S_t), \tag{B.23}$$

where $\widetilde{P}_t(S_t) \equiv P_{t|t} P_{t+1|t}(S_t)^{-1}$ and collapsing across states occurs each period as in equations (B.19) and (B.20).

James Morley received a Bachelor of Arts in Honours Economics from the University of British Columbia in May 1995 and a Masters of Arts in Economics from the University of Washington in June 1998. He has accepted a tenure track position in the Economics Department at Washington University to begin in Autumn 1999.