Sequential Optimal Portfolio Performance: Market and Volatility Timing

Michael Johannes Nicholas Polson Jon Stroud*

First draft: March 2000

This draft: February 2002

Abstract

This paper studies the economic benefits of return predictability by analyzing the impact of market and volatility timing on the performance of optimal portfolio rules. Using a model with time-varying expected returns and volatility, we form optimal portfolios sequentially and generate out-of-sample portfolio returns. We are careful to account for estimation risk and parameter learning. Using S&P 500 index data from 1980-2000, we find that a strategy based solely on volatility timing uniformly outperforms market timing strategies, a model that assumes no predictability and the market return in terms of certainty equivalent gains and Sharpe ratios. Market timing strategies perform poorly due estimation risk, which is the substantial uncertainty present in estimating and forecasting expected returns.

*Johannes is at the Graduate School of Business, Columbia University, 3022 Broadway, NY, NY, 10027, mj335@columbia.edu. Polson is at the Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago IL 60637, ngp@gsbngp.uchicago.edu. Stroud is at The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6302, stroud@wharton.upenn.edu. We would like to thank seminar participants at Columbia and NYU.

1 Introduction

There is strong empirical evidence that equity returns have predictable components. Standard theoretical results based on Merton (1971) show how investors should alter their optimal portfolio holdings to take advantage of time-varying investment opportunities. For example, investors should increase their allocation to risky assets in periods of high expected returns (market timing) and decrease their allocation in periods of high volatility (volatility timing).

In this paper, we evaluate the economic benefits of equity return predictability. While theory shows that an investor gains from market and volatility timing, the nature and magnitude of the potential benefits is largely unknown. Our approach takes the perspective of an investor who exploits predictability by using a model of time-varying expected returns and volatility to sequentially form optimal portfolios. To measure the economic benefits of predictability, we compare our investor's returns to those based on a model without predictability and to the market.

We find that a strategy based on solely volatility timing leads to significant economic gains. On the other hand, market timing strategies based on time-varying expected returns always perform worse than the volatility timing strategy. Moreover, expected return strategies often perform worse than the no predictability strategy or the market. For example, an investor with a risk aversion of 4 who follows the volatility timing strategy attains an annualized Sharpe ratio of 0.71, compared to 0.49 for the market, 0.39 for the no predictability strategy and 0.31 for a expected returns based strategy. Relative to the model with no predictability, the volatility timing strategy results in an annualized certainty equivalent gain of 3.26%, which shows the substantial economic gains generated by volatility predictability.

Our approach differs from the existing literature in two major ways. First, instead of measuring the benefits of predictability via ex-ante calibration,[1] we measure economic benefits through the *out-of-sample* performance of the optimal portfolio holdings. This provides a practical evaluation of the economic benefits and is especially important in light of recent ev-

[1] Ex-ante calibration quantifies the population properties of optimal portfolios. Typically, calibration computes the expected utility gains of predictability and reports the sensitivity of the optimal portfolio rule to changes in the state variables or parameters. See, for example, Balduzzi and Lynch (1999), Campbell and Viceira (1999), Campbell, Chan and Viceira (2002), Lynch (2000), Gomes (2001) or Chacko and Viceira (2002).

idence that questions the nature of return predictability. For example, while there is evidence for *in-sample* expected return predictability based on scaled price variables, lagged returns or interest rate based variables, there is little evidence for any out-of-sample expected return predictability.[2]

Second, our investor does not observe the parameters or state variables that characterize the return distribution, thus we have to account for *estimation risk* and *parameter learning* when forming optimal portfolios. Estimation risk is the static uncertainty that is inherently present when estimating parameters or state variables. Parameter learning is the dynamic counterpart to estimation risk and corresponds to the process of revising beliefs about parameters as more data arrives. Ignoring estimation risk or parameter learning typically leads to misleading allocations (Brennan (1998), Stambaugh (1999) and Barberis (2000)).

To form optimal portfolios sequentially, our investor computes the filtering and predictive distribution of returns, while simultaneously accounting for estimation risk and parameter learning. To do this, we extend the current econometrics literature by developing a new Markov Chain Monte Carlo (MCMC) algorithm that solves the optimal filtering and prediction problem under parameter uncertainty. To study the role of parameter uncertainty in forecasting expected returns and volatility, we compare both filtered and smoothed estimates of expected returns and volatility. We also extend the current literature by analyzing sequential parameter estimates which characterize parameter learning over time.

Our model specifies that expected returns and volatility are stochastic and mean-reverting. This closely coincides with the idealized stochastic setting described in Brennan (1998) and the model used in Brandt and Kang (2002). For expected returns, our mean-reverting model is motivated by Merton (1971) and our log-volatility specification is standard. Our specification provides a flexible and parsimonious model for incorporating predictability that reflects our agnosticism over the source of the return predictability.

We focus on the single-period problem portfolio problem, since the multi-period portfolio problem with time-varying expected returns or volatility in the presence of estimation risk or parameter learning is computationally intractable. The difference between the single and multi-period problems is hedging demands. Recent evidence suggests that ignoring hedging

[2]See, for example, Bossaerts and Hillion (1999), Cremers (2001), Goyal and Welch (2002) and Hahn and Lee (2001).

demands is not a major concern, as they are typically a small component of asset demands (see Brandt (1999), Ang and Bekaert (2000), Aït-Sahalia and Brandt (2001), Gomes (2001) and Chacko and Viceira (2002)). Hedging demands are important components only when investors have very long investment horizons. To derive the optimal portfolio rule, we use an extension of Stein's lemma that applies when return predictability is generated by stochastic expected returns and volatility (Gron et al. (2001)).

Given the portfolio rule, we compute out-of-sample portfolio returns for a number of models incorporating time-varying expected returns, volatility or both. Evaluating the economic benefits of these dynamic strategies is a difficult problem and we therefore use a number of different metrics. First, we report average portfolio returns and volatility, and the modified Sharpe ratio of Graham and Harvey (1997) which accounts for the shortcomings of the traditional Sharpe ratio in dynamic settings, see Leland (1997). Second, in terms of utility, we compute the certainty equivalent gain (or loss) that a given portfolio strategy generates over the returns from the no predictability case and the market strategy.

As mentioned earlier we find that the optimal portfolio strategy based on a model with only time-varying volatility uniformly outperforms the other strategies in terms of the Sharpe ratio and certainty equivalent gains. This holds for a range of risk aversions and with or without portfolio leverage. Perhaps even more striking is the performance of an optimal portfolio based solely on time-varying expected returns. *If* the investor knew the full-sample parameter estimates, the portfolio gains from market timing would be on par with the volatility timing model. However, once estimation risk and parameter learning are accounted for, we find that expected returns based strategies perform poorly. For example, its modified Sharpe ratio is always worse than the volatility timing strategy and nearly always worse than the market return. The market timing strategy in some cases provides certainty equivalent gains over a model with no predictability. This indicates that it may be advantageous to account for time-varying expected returns when forming optimal portfolios even in presence of estimation risk.

Finally, we find that strategies based on a model with both time-varying expected returns and volatility perform similarly to those based solely on time-varying expected returns. Once again, estimation risk is responsible. This highlights the difference between in-sample statistical model fitting and out-of-sample portfolio performance. While, the model with time-varying expected returns and volatility certainly provides a better fit in-sample, the model provides little economic benefit to an investor.

Our conclusions regarding the economic benefits of volatility timing strategies are related to the findings in Fleming, et al. (2001a, b), although there are a number of major differences. First, while they do use out-of-sample volatility forecasts, they condition on full-sample estimates of expected returns which introduces a look-ahead bias. We find the expected estimation risk is of first-order importance in calculating the economic benefits. Second, we find economic benefits to volatility timing using only a single risky asset, without relying on diversification or time-varying correlations (Fleming et al. (2001a, b) use three risky assets (gold, T-bonds and S&P 500 futures)). This is of particular concern given the well-known extreme sensitivity of optimal allocations to estimates of expected returns (see, e.g., Best and Grauer (1991)).

The rest of the paper is outlined as follows. Section 2 discusses the evidence on predictability, introduces our model and discusses optimal portfolio formation. Sections 3 and 4 describe our estimation methodology and estimation results. Section 5 summarizes the portfolio performance. Section 6 concludes.

2 The Investment Set and Optimal Portfolios

2.1 A Model of Time-Varying Expected Returns and Volatility

While it is commonly accepted that both expected returns and volatility have predictable components, the exact nature of the predictability is less certain. Take for example, expected return predictability. Many authors have found evidence supporting expected return predictability based on variables such as the dividend-price ratio, lagged returns, interest rates and interest rate spreads. Despite this evidence for *in-sample* return predictability, there is little evidence for any *out-of-sample* expected return predictability. This distinction is especially important in portfolio applications where, presumably, an investor seeks to predict returns out-of-sample.

Specifically, recent papers by Bossaerts and Hillion (1999), Cremers (2001) and Goyal and Welch (2002) analyze the ability of standard predictors to forecast out-of-sample returns.

Goyal and Welch (2002) argue there is no out-of-sample predictability based on dividend-price ratios due to the structural instability in the relationship between returns and the dividend-to-price ratio.[3] Bossaerts and Hillion (1999) and Cremers (2001) study the sensitivity of model choice (the number and type of linear predictors) on predictability and find that while there is some evidence for in-sample predictability, there is little evidence for out-of-sample predictability.[4] Additionally, as pointed out by Hodrick (1992) and Stambaugh (1999), there are statistical issues regarding tests of predictability based on scaled price variables.

Our approach to modeling return predictability is motivated by Merton (1971) and allows for predictable time-variation in expected returns and volatility by specifying that both are mean-reverting stochastic processes. We consider the following model of continuously compounded returns (r_t):

$$r_t = \mu_t + \sqrt{V_t}\varepsilon_t \tag{1}$$

$$\mu_{t+1} = \alpha_\mu + \beta_\mu \mu_t + \sigma_\mu \varepsilon^\mu_{t+1} \tag{2}$$

$$\log(V_{t+1}) = \alpha_v + \beta_v \log(V_t) + \sigma_v \varepsilon^v_{t+1} \tag{3}$$

where μ_t is the time-varying and stochastic expected return, $\sqrt{V_t}$ is the stochastic volatility and $\varepsilon_{t,}$ ε^v_t and ε^μ_t are potentially correlated normal random variables.

This specification provides a flexible model of time-varying expected returns and volatility. It embodies our agnostic beliefs about the specific sources of predictability as expected returns and volatility are stochastic and latent. The specification for expected returns captures the common view that expected returns have a mean-reverting component.[5] One practical advantage of this specification is that our portfolio returns will not be the result of our particular choice of predictor variables.

[3] In particular, they find that the predictive ability of dividend ratios was due to two influential observations in the 1970s.

[4] The variables in Bossarts and Hillion (1999) and Cremers (2001) include a January dummy, lagged returns, excess bond returns, Treasury yields and term spreads, stock market price level, dividend yield, price-to-earnings, volume-to-price, industrial production, inflation, changes in inflation and credit spreads. Recently, Lettau and Ludvigson (2000) find evidence that the consumption to wealth (CAY) ratio provides out-of-sample predictability, although Hahn and Lee (2001) question the recent predictive ability of the consumption to wealth ratio.

[5] For example, this specification is consistent with Campbell and Shiller (2001) who argue that dividend-to-price and price earnings ratios are mean-reverting components of expected returns.

Our expected returns specification is identical to that used in continuous-time portfolio problems by Merton (1971), Kim and Omberg (1996), Liu (1999) and Wachter (1999).[6] These papers focus on solving for the optimal rule, while ours analyzes the dynamic properties of the optimal portfolio returns. Following Merton (1971) we consider a partial equilibrium setting and allow for the possibility of negative expected returns. The log-variance model is popular for empirical applications (see Jacquier, et al. (1994, 2001)). As shown by Chernov et al. (2001) and Andersen et al. (2001), it provides a fit similar to that of the square-root model which is commonly used in portfolio applications.

Our specification does not include any lagged interaction between μ_t and V_t. This implies, for example, that there is not a variance risk premia term in the expected returns equation:

$$\mu_{t+1} = \alpha_\mu + \beta_\mu \mu_t + \beta_v V_t + \sigma_\mu \varepsilon_{t+1}^\mu.$$

This is consistent with the current literature that finds no evidence that lagged volatility explains expected returns or vice versa (see Brandt and Kang (2002) and Koopman and Uspensky (1999)). We do, however, allow for a contemporaneous interaction between expected returns and volatility via a correlation between the shocks to the latent variables, $cov(\varepsilon_t^v, \varepsilon_t^\mu) = \rho$. This effect proxies a leverage-type effect and allows expected returns and volatility to move together. Brandt and Kang (2002) find that a correlation between ε_t and ε_t^μ is not significant.

As noted by Whitelaw (1994) and Harvey (2001), the relationship between the levels and dynamics of μ_t and V_t are certainly extremely complicated. Harvey (2001) argues that the findings will depend crucially on the conditioning information used, on the data set and the model used. Given the large number of papers that analyze this issue and the lack of consensus,[7] we intentionally chose the simplest specification that captured the time-variation in expected returns and volatility.

Finally, we note that adding additional lagged variables and correlations may generate identification problems and significantly complicates estimation (see Brandt and Kang (2002)

[6] These papers specify that $d\mu_t = \kappa\,(\theta - \mu_t)\,dt + \sigma dW_t$. Since this SDE implies a Gaussian process for μ_t, the specifications are the same if we set $\beta_\mu = e^{-k}$, $\alpha_\mu = \theta(1 - e^{-k})$ and $\sigma_\mu^2 = \frac{\sigma^2}{2k}(e^{2k} - 1)$.

[7] A partial list of the papers analyzing the relationship between levels and/or dynamics of expected returns and volatility includes French, Schwert and Stambaugh (1987), Breen, Glosten and Jaganathan (1989), Glosten, Jaganathan and Runkle (1993), Whitelaw (1994, 1997), Harrison and Zhang (1999), Brandt and Kang (2001) and Harvey (2001). Despite this large literature, there are few if any robust findings regarding the relationship between expected returns and volatility.

for a discussion of some of the issues involved). We will discuss these issues later in more detail, but it is not clear if parameters in more general models can be identified from the observed data. Moreover, filtering and forecasting expected returns and volatility in more general models may not be possible.

2.2 Portfolio Allocation

Beginning with Merton (1971), a number of authors have solved for exact, closed-form solutions in optimal portfolio problems assuming expected returns or stochastic volatility evolve continuously through time.[8] There are also a number of papers that provide numerical or approximate closed-form solutions.[9] Our problem is noticeably harder due to the presence of time-varying μ_t and V_t, estimation risk and parameter learning. To understand the issues involved in solving this problem, we first consider the problem of portfolio choice in the complete information economy, and then in a partial information setting.

In a complete information economy, there is no estimation risk which implies that μ_t, V_t and any parameters, Θ, governing their evolution are observed. In the general multi-period problem, an investor chooses at time t a sequence of portfolio weights $\{\omega_s\}_{t\leq s\leq T}$ to maximize utility at time T:

$$J\left(W_t,\mu_t,V_t,\Theta\right)=\max_{\{\omega_s\}_{s=t}^{T}} E\left[U\left(W_T\right)|W_t,\mu_t,V_t,\Theta\right]$$

subject to the usual budget constraint. If we consider a continuous-time diffusion model for expected returns and volatility the optimal portfolio is

$$\omega_t=\left(\frac{-J_{WW}}{J_W W}\right)\frac{\mu_t-r_f}{V_t}$$

where r_f is the risk-free rate, J_W and J_{WW} are partial derivatives of the value function. For simplicity, this rule assumes the shocks to expected returns and volatility are independent of the shocks to returns.

[8]See, e.g., Kim and Omberg (1996), Liu (1999) and Wachter (1999) in the case of time-varying expected returns and Liu (1999), Longstaff (2000) and Liu, Longstaff and Pan (2001) in the case of stochastic volatility.

[9]In either discrete or continuous-time settings, Brennan, et al. (1997), Ang and Bekaert (1999), Balduzzi and Lynch (1999), Campbell and Viceira (1999), Chacko and Viceira (1999), Campbell, Chan and Viceira (2001) and Lynch (2001) use approximate or numerical solutions to study optimal portfolio holdings in the presence of expected return or volatility predictability.

However, given the uncertain nature of return predictability, it is unrealistic to assume that μ_t, V_t and Θ are observed by the investor. In practice, all must be estimated and this raises the issue of estimation risk and parameter learning. In general, assuming parameters are known when they are in fact estimated with noise data can lead to drastically misleading optimal portfolio conclusions as the investor ignores a source of risk (see, e.g., Kandel and Stambaugh (1996), Stambaugh (1999), Barberis (2000) and Pastor (2000)).

When μ_t, V_t or Θ are not observed, the investor solves the optimal portfolio problem in two stages. In the first stage, the investor estimates the parameters and state variables and then computes the predictive distribution of wealth given observed data, $p\left(W_T|R^t\right)$. Given this density, the investor solves

$$\begin{aligned} J\left(R^t\right) &= \max_{\{\omega_s\}_{s=t}^T} E\left[U\left(W_T\right)|R^t\right] \\ &= \max_{\{\omega_s\}_{s=t}^T} \int U\left(W_T\right) p\left(W_T|R^t\right) dW_T \end{aligned}$$

where R^t is a vector of observed, discretely compounded returns up to time t and the maximization is subject to the usual budget constraint.

There are two difficult components to this problem. First, solving for $p\left(W_T|R^t\right)$ requires the investor to estimate parameters and forecast future expected returns and volatility. The solution to this highly nonlinear filtering and estimation problem is $p\left(\mu_T, V_T, \Theta|R^t\right)$. Second, conditional on $p\left(W_T|R^t\right)$ or $p\left(\mu_T, V_T, \Theta|R^t\right)$, solving for the optimal portfolio rule is extremely difficult. In general, since the filtering density is of the same dimension as the entire history of returns, R^t, solving this dynamic program either analytically or numerically is currently intractable.

In some highly stylized cases, there are tractable solutions in continuous-time to a multi-period investment problem when parameters or state variables are unobserved (see Brennan (1998), Brennan and Xia (2000), Comon (2001) or Xia (2001)). Unfortunately, this approach is limited along two dimensions. First, only certain parameters or state variables can be unobserved. For example, expected returns can be unobserved but the current volatility state or parameters driving the volatility of returns cannot by the quadratic variation process. Second, when there are multiple unobserved states, the computational demands for filtering grow exponentially. Thus solving the multi-period problem with estimation risk or parameter learning in realistic settings is currently not possible.

To solve the optimal portfolio problem, we follow the existing literature and simplify the allocation problem by considering a single-period problem:

$$J\left(R^t\right) = \max_{\omega_t} E\left[U\left(W_{t+1}\right)|R^t\right] = \max_{\omega_t} \int U\left(W_{t+1}\right) p(R_{t+1}|R^t) dR_{t+1},$$

where $p(R_{t+1}|R^t)$ is the predictive distribution of future returns and next periods wealth is $W_{t+1} = W_t[R_f + \omega_t(R_{t+1} - R_f)]$. This is the approach taken in, for example, Kandel and Stambaugh (1996), Stambaugh (1999), Barberis (2000), Pastor and Stambaugh (2000), Pastor (2000) and Wang (2002).

The difference between single and multi-period problems is hedging demands. As shown by Brandt (1999), Ang and Bekaert (2000), Aït-Sahalia and Brandt (2001), Gomes (2001) and Chacko and Viceira (2002), hedging demands are typically extremely small components of the optimal portfolio allocation and have been found to be important only for long-horizon investors such as the infinitely lived investors in Campbell and Viceira (1999, 2000).

To derive the optimal portfolio rule, assume that $U\left(W_{t+1}\right)$ is twice differentiable, strictly increasing and concave in the portfolio weight. Given this, the optimal portfolio is characterized by the first order condition

$$E\left[U'(W_{t+1})(R_{t+1} - R_f)|R^t\right] = 0,$$

where the expectation is taken over the predictive distribution of future returns. Applying the definition of covariance implies that

$$cov\left[U'(W_{t+1}), R_{t+1} - R_f|R^t\right] + E\left[U'(W_{t+1})|R^t\right] E\left[(R_{t+1} - R_f)|R^t\right] = 0. \tag{4}$$

To separate utility effects from those of risk and return, we use the fact that the predictive distribution of returns is a stochastic volatility mixture distribution. In this case, a generalization of Stein's lemma allows us to re-write the covariance term as

$$\begin{aligned} cov\left[U'(W_{t+1}), R_{t+1} - R_f|R^t\right] &= E^Q\left[U''(W_{t+1})|R^t\right] cov\left(W_{t+1}, R_{t+1}|R^t\right) \\ &= \omega_t E^Q\left[U''(W_{t+1})|R^t\right] var\left[R_{t+1}|R^t\right], \end{aligned}$$

where Q is the size-biased volatility-adjusted distribution (see Gron, et al. (2000)). Solving for the optimal portfolio, we find that

$$\omega_t = \frac{1}{\gamma} \frac{E\left[R_{t+1} - R_f|R^t\right]}{Var\left[R_{t+1}|R^t\right]} \tag{5}$$

where $\gamma^{-1} = -E^Q \left[U''(W) | R^t \right] / E \left[U'(W) | R^t \right]$. This provides a justification for using a mean-variance rule, where the risk-aversion is modified to take into account the fact that returns are generated by a fat-tailed stochastic volatility distribution. Alternatively, the rule can be viewed just as the usual conditional mean-variance rule.

We use a second order Taylor series expansion of e^{r_t} to convert from discretely to continuously compounded returns. Alternatively we could assume that $R_{t+1} \approx \mu_{t+1} + \frac{1}{2} V_{t+1} + \sqrt{V_{t+1}} \varepsilon_{t+1}$. This implies that

$$\omega_t = \frac{1}{\gamma} \frac{E_t(\mu_{t+1}) - r_f}{Var_t(r_{t+1})} + \frac{1}{2\gamma}. \tag{6}$$

We implement this optimal rule. This is the same rule that Kandel and Stambaugh (1996) use to approximate a constant relative risk aversion utility function.

It is important to recognize potential problems that may arise in this setting. For example, with constant relative risk aversion utility, $U(W_{t+1}) = \frac{W_{t+1}^{1-A}}{1-A}$, expected utility can be infinite for $A > 1$ when the investor either shorts the risky asset or takes a levered position, as noted by Kandel and Stambaugh (1996). The problem in their setting is that the expected value of wealth may not exist because the predictive density of returns has a fat-tailed t-distribution, as the variance has an inverted gamma prior. In our empirical implementation, our prior is effectively a mixture of normals, which leads to finite expected utility in cases such as power utility.

3 Optimal Sequential Learning and Estimation Risk

The optimal portfolio rule depends on the investor's perception of the predictive distribution of future returns. When forecasting expected returns and variance, our Bayesian investor is careful to integrate out all of the uncertainty in estimating Θ, μ_t and V_t. This section discusses the inferential problems our investor solves prior to making portfolio decisions.

To understand the difficulties in sequentially forecasting returns, consider the following factorization of the predictive distribution of returns:

$$\begin{aligned} p(r_{T+1}|r^T) &= \int p(r_{T+1}, V_{T+1}, \mu_{T+1}, \Theta|r^T) d\Theta d\mu_{T+1} dV_{T+1} \\ &= \int p(r_{T+1}|V_{T+1}, \mu_{T+1}) p(V_{T+1}, \mu_{T+1}|\Theta, r^T) p\left(\Theta|r^T\right) d\Theta d\mu_{T+1} dV_{T+1} \end{aligned}$$

This decomposition shows that the predictive density incorporates three different components: (1) the conditionally normal model specification for returns, $p(r_{T+1}|V_{T+1}, \mu_{T+1}) = \mathcal{N}(\mu_{T+1}, V_{T+1})$; (2) latent variable filtering and forecasting, $p(V_{T+1}, \mu_{T+1}|r^T, \Theta)$, and (3) the parameter estimation problem $p\left(\Theta|r^T\right)$. We briefly discuss each of these components.

The marginal parameter posterior distribution, $p\left(\Theta|r^T\right)$, provides parameter inference and quantifies the uncertainty regarding the values of the parameters. It is important to note that our investor treats the parameters as random. If estimation risk is not taken into account, this distribution is a point mass evaluated at a set of parameters, $\widehat{\Theta}$. Previous research documents that in both i.i.d. and regression based settings, parameter uncertainty tends to alter the predictive density of returns and can have a major impact on portfolio allocation. [10]

If the investor knew the parameter values, $p(V_{T+1}, \mu_{T+1}|r^T, \Theta)$ provides the predictive distribution of expected returns and volatility. We can represent this density as

$$p(V_{T+1}, \mu_{T+1}|r^T, \Theta) = \int p(V_{T+1}, \mu_{T+1}|\mu_T, V_T, \Theta, r^T) p(\mu_T, V_T|r^T, \Theta) d\mu_T dV_T. \qquad (7)$$

The intuition is that, conditional on the parameters, prediction of the future state variables involves two steps. The first step specifies the evolution of the state variables over time. If the investor knew the parameters and the current expected returns and volatility then the distribution of next periods expected returns and volatility, $p(V_{T+1}, \mu_{T+1}|\mu_T, V_T, \Theta, r^T)$, is again normal. The final component, $p(\mu_T, V_T|r^T, \Theta)$, solves the classical filtering problem of estimating the latent state variables given parameters and observed data. Together, these components account for the uncertainty in estimating the latent state variables.

[10] In the i.i.d. settings, see for example, see Bawa, Brown and Klein (1979) or Polson and Tew (2000). In settings that include regression parameters, see Kandel and Stambaugh (1996), Pastor and Stambaugh (1999), Pastor (1999) and Barberis (2000).

3.1 Estimation and Posterior Simulation

To estimate the model and implement the optimal portfolio allocation, we need to solve the smoothing, filtering, sequential learning and forecasting problems which are summarized by the following distributions:

$$\begin{aligned} \text{Smoothing} &: p\left(\Theta, \mu^T, V^T | r^T\right) \\ \text{Filtering} &: p\left(\mu_t, V_t | r^t\right) \ \ t = 2, ..., T \\ \text{Sequential Learning} &: p\left(\Theta | r^t\right) \ \ t = 2, ..., T \\ \text{Forecasting} &: p\left(\mu_{t+1}, V_{t+1} | r^t\right) \ \ t = 2, ..., T \end{aligned}$$

where $r^T = (r_1, \ldots, r_T)$, $\mu^T = (\mu_1, \ldots, \mu_T)$, and $V^T = (V_1, \ldots, V_T)$ as the full observation and state vectors.

While these three distributions are clearly related, they use different amounts of information and address different issues. The smoothing distribution summarizes information about the parameters and the entire paths of expected returns and volatility conditional on the entire data set. For example, the smoothing distribution generates $p\left(\Theta | r^T\right)$ which summarizes in-sample parameter inference. While reasonable for static inference problems, this is not useful for practical portfolio problems as it uses future information (up to time T) to estimate the state variables at time t. The filtering and sequential learning distributions solve for the distribution of the current latent states or parameters conditional only on available information at time t. We now discuss these estimation problems in turn.

3.2 The Smoothing Problem

To solve the smoothing problem, we generate samples from $p\left(\Theta, \mu^T, V^T | r^T\right)$. Since the dimension of $p\left(\Theta, \mu^T, V^T | r^T\right)$ is $2T + K$, where K is the number of elements in Θ, it is not possible to directly sample from the smoothing distribution and we must rely on numerical schemes to characterize the distribution.

We use Markov Chain Monte Carlo (MCMC) methods to sample from $p\left(\Theta, \mu^T, V^T | r^T\right)$. MCMC is very popular for estimating latent variable models due to their generality and the speed of implementation. Recent papers in finance using MCMC include Barberis (2000), Pastor (1999), Pastor and Stambaugh (2000) and Eraker, Johannes and Polson (2002). Johannes and Polson (2002) provide a general overview and description of these methods.

Our MCMC algorithm consists of repeated sampling from the following set of conditional distributions:

$$\begin{aligned} \text{Regression Parameters} &: p(\,\phi\,|\,\mu^T, V^T, \Sigma, r^T\,) \\ \text{Innovation Covariance} &: p(\,\Sigma\,|\,\mu^T, V^T, \phi, r^{\,T}) \\ \text{Expected Returns} &: p(\,\mu^T\,|\,V^T, \phi, \Sigma, r^T\,) \\ \text{Volatility} &: p(\,V^T\,|\,\mu^T, \phi, \Sigma, r^T\,) \end{aligned}$$

where $\phi = (\alpha_\mu, \beta_\mu, \alpha_v, \beta_v)$ denotes the "regression" parameters and Σ denotes the innovation covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_\mu^2 & \rho\,\sigma_\mu\sigma_v \\ \rho\,\sigma_\mu\sigma_v & \sigma_v^2 \end{pmatrix}.$$

Given initial values for the parameters (ϕ^0, Σ^0) and state variables $\left(\left(\mu^T\right)^0, \left(V^T\right)^0\right)$, the algorithm updates iteratively by sampling from the conditional distributions given above. For example, the first two steps of the algorithm would be

$$\begin{aligned} \phi^1 &\sim p(\,\phi\,|\,\left(\mu^T\right)^0, \left(V^T\right)^0, \Sigma^0, r^T\,) \\ \Sigma^1 &\sim p(\,\Sigma\,|\,\left(\mu^T\right)^0, \left(V^T\right)^0, \phi^1, r^{\,T}). \end{aligned}$$

Continuing in this manner, the algorithm produces a Markov Chain,

$$\left\{\phi^g, \Sigma^g \left(\mu^T\right)^g, \left(V^T\right)^g, \right\}_{g=1}^G$$

whose distribution converges to $p\left(\Theta, \mu^T, V^T | r^T\right)$ under a number of different metrics.

To derive conditional distributions for the parameters, we need to specify the prior distributions. We assume conjugate prior distributions on the parameters which implies that both ϕ and Σ can be updated directly by drawing from the posterior conditional. The prior distribution for ϕ is multivariate normal and the prior for Σ is inverted Wishart ($\mathcal{W}^{-1}$). Under these assumptions, the full conditional posterior distribution of ϕ is also normal, $p(\,\phi\,|\,\mu^T, V^T, \Sigma, r^T\,) \sim \mathcal{N}(m^*, S^*)$, and the conditional posterior of Σ is $\mathcal{W}^{-1}(A^*, d^*)$ (the formulas for the starred parameters are given in the Appendix).

The complete conditionals for the latent states are slightly more complicated. The conditional for μ is given by

$$p(\, \mu^T \mid V^T, \phi, \Sigma, r^T \,) \propto p(r^T|\mu^T, V^T)p(\mu^T|V^T, \phi, \Sigma)$$

and by inspection is proportional to the likelihood, $p(r|\mu, V)$, multiplied by the evolution distribution $p(\mu|V, \phi, \Sigma)$. These densities combine to form a heteroskedastic Gaussian state-space model, and so we can sample the entire vector μ jointly using the *Forward-Filtering Backward-Sampling* (FFBS) algorithm of Carter and Kohn (1994). As in most cases, the volatility update presents the only difficulty in our MCMC scheme as $p(V|\mu, \phi, \Sigma, r)$ is unrecognizable and cannot be directly sampled from. However, we avoid this problem by using the mixture approach of Carter and Kohn (1994, 1996) (see also Kim, Shephard and Chib (1998)).

3.3 The Filtering Problem

The filtering distribution, taking into account parameter uncertainty, is defined as

$$p(\mu_t, V_t|r^t) \text{ for } t = 2, \ldots, T.$$

and provides the distribution of the current expected return and volatility states given current information. One way to calculate this distribution is to marginalize the parameters from the joint density $p(\, \mu_t, V_t \,, \phi, \Sigma|\, r^t \,)$. Obtaining this distribution for one time period is not difficult (it is the same as the smoothing problem using data up to time t), but computing this distribution for every day in our sample is problematic due to computational demands. Computing the smoothing distribution a single time takes about 20 minutes using our S&P 500 data which has over 3000 observations. Repeating this for every day in our sample would be prohibitively expensive in terms of computing time.

Our filtering problem is more difficult than the typical problem because we account for parameter uncertainty. Typical filtering exercises compute $p(\mu_t, V_t|\widehat{\Theta}, r^t)$, ignoring the fact that the parameters are not known, while we require $p(\mu_t, V_t|r^t)$. This implies that standard methods such as approximate Kalman filtering, whereby a non-linear, non-Gaussian model is approximated by a linear Gaussian model, are not applicable.

To compute the filtering distribution, we use two approximations, one with regard to parameter uncertainty and the other with regard to the filtering distribution of the state variables conditional on the parameters. Ideally, the filtering algorithm would update inference on the parameters as every additional data point arrives. However, it is unlikely that the parameter posterior would change with the addition of a single data point if a reasonable number of observations have already been observed. Because of this, we update the parameter posterior every 50 days, that is, we recompute $p(\mu^t, V^t, \phi, \Sigma | r^t)$. Our results show that the parameters tend to change very slowly with the exception of a couple of periods like the crash of 1987. In this case, it is clear that the parameters would change even if they were updated as every new data point arrived.

Second, we use an approximate sampling procedure to draw the state variables μ_t and V_t for the periods between parameter refreshing. Conditional on the parameters, we have that the filtering density of the latent states is $p(\mu_{t-k}^t, V_{t-k}^t | r^t, \phi, \Sigma)$, where in our empirical implementation, $k = 250$ (1-year), $\mu_{t-k}^t = \{\mu_{t-k}, ..., \mu_t\}$ and $V_{t-k}^t = \{V_{t-k}, ..., V_t\}$. This density, by the Markov property, can be written as the integral

$$\int p(\mu_{t-k}^t, V_{t-k}^t | r_{t-k}^t, \mu_{t-k}, V_{t-k}, \phi, \Sigma) p\left(\mu_{t-k}, V_{t-k} | r^t, \phi, \Sigma\right) d\mu_{t-k} dV_{t-k}$$

where $r_{t-k}^t = \{r_{t-k}, ..., r_t\}$.

Our approximation uses the MCMC draws $\left\{\mu_{t-k}^g, V_{t-k}^g\right\}_{g=1}^G$ from the last periods filtering problem to initialize the states, (μ_{t-k}, V_{t-k}). That is, we use the MCMC draws from $p\left(\mu_{t-k}, V_{t-k} | r^{t-1}, \phi, \Sigma\right)$ instead of $p\left(\mu_{t-k}, V_{t-k} | r^t, \phi, \Sigma\right)$. This approximation will only have an effect if the current return observation dramatically affects the estimate of expected returns and volatility from $k-$periods ago. In practice with $k = 250$, it is unlikely that this has any substantive impact. Given these draws, we sample the unobserved states from

$$p(\mu_{t-k}^t, V_{t-k}^t | r_{t-k}^t, \mu_{t-k}^g, V_{t-k}^g, \phi, \Sigma)$$

which provides draws from $p(\mu_{t-k}^t, V_{t-k}^t | r^t, \phi, \Sigma)$. We also randomize over parameters to account for parameter uncertainty. The algorithm is extremely fast as the smoothing step requires computing only k steps. In the limit, as we refresh parameters every day and increase the fixed length to the entire history, our algorithm converges the true filtering distribution. To our knowledge, this is the only computationally feasible algorithm that can perform filtering under parameter uncertainty. Polson, Stroud and Mueller (2002) provide further details of this filtering algorithm and simulation based evidence on its efficiency.

3.4 Prediction Problem

Given the filtering distribution, it is straightforward to calculate the forecasting distribution. The distribution of future returns conditional on μ_{t+1} and V_{t+1} is normal, given the model specification, thus the predictive distribution is

$$p\left(r_{t+1}|r^t\right) = \int N\left(\mu_{t+1}, V_{t+1}\right) p\left(\mu_{t+1}, V_{t+1}|r^t\right) d\mu_{t+1} dV_{t+1}.$$

To compute the predictive distribution, we need only to calculate $p\left(\mu_{t+1}, V_{t+1}|r^t\right)$.

The predictive distribution for next periods expected return and volatility state are given by the model specification:

$$\begin{aligned} \mu_{t+1} &= \alpha_\mu + \beta_\mu \mu_t + \sigma_\mu \varepsilon_{t+1}^\mu \\ \log\left(V_{t+1}\right) &= \alpha_v + \beta_v \log\left(V_t\right) + \sigma_v \varepsilon_{t+1}^v. \end{aligned}$$

The filtering problem from the previous section provides samples from $p(\mu_t, V_t, \phi, \Sigma | r^t)$ and implies that the predictive distribution of returns is

$$p\left(r_{t+1}|r^t\right) \simeq \frac{1}{G} \sum_{g=1}^{G} N\left(\mu_{t+1}^g, V_{t+1}^g\right).$$

From this, it would be straightforward, although numerical intensive, to compute the optimal allocations for a specific utility function.

Since our portfolio rule requires us only to know the conditional mean and variance, we only need to forecast those moments. Given the Gaussian nature of the problem we have that

$$\begin{aligned} E_t\left(\mu_{t+1}\right) &= E\left(\alpha_\mu + \beta_\mu \mu_t | r^t\right) \\ &= \int \left(\alpha_\mu + \beta_\mu \mu_t\right) p\left(\mu_t, \Theta | r^t\right) d\mu_t d\Theta \end{aligned}$$

where we are careful to integrate out the posterior uncertainty regarding the parameters and expected returns. The MCMC estimate of the forecasted expected returns is:

$$E_t\left(\mu_{t+1}\right) \approx \frac{1}{G} \sum_{g=1}^{G} \left(\alpha_\mu^g + \beta_\mu^g \mu_t^g\right)$$

The volatility is similarly estimated.

3.5 Prior Specification

To complete the specification, we need to specify the prior parameters. When analyzing optimal portfolio holdings using actual data, it is now standard in both i.i.d and regression settings to formally impose prior information on the form of the expected returns process (see, e.g. Black and Litterman (1991), Pastor (1999) and Pastor and Stambaugh (1999)).[11]

We place uninformative priors over the volatility parameters, α_v, β_v, σ_v and ρ and the volatility of expected returns, σ_μ, but we do impose some prior information on parameters that determine the average expected return and speed of mean reversion, α_μ and β_μ.

Our motivation for using informative priors is twofold. First, without some restrictions on the nature of expected returns, optimal portfolio positions will often be extreme.[12] To see why, consider taking simple 45-day moving average estimates for μ_t and V_t for the S&P 500 index. For example, from April 3 - May 20, 1997, the S&P 500 went from 760.60 to 841.66, a 10 per cent move (86 per cent annualized). Volatility, estimated similarly was about 15 per cent annualized. What is the mean-variance optimal portfolio rule? With a reasonable risk aversion of 4, the portfolio leverages wealth by a factor of 8.9. Even with a risk aversion of 8, the investors is leveraged more than 4 times wealth.

The second motivation for informative priors on α_μ and β_μ is statistical. Without informative priors on these parameters (especially β_μ), there are identification problems. For example, with uninformative priors, estimates quickly degenerate and the algorithm converges to a state where $\mu_t \approx r_t$, corresponding to a near perfect fit and an infinite likelihood. Expected returns are extremely volatile, oscillating from -20 to +40% over short periods of time. They are not persistent (β is close to zero) and expected returns explain most of the volatility returns. We have strong prior beliefs over the expected returns process that this is not correct.

This should not be a surprise as our model is a mixture of normal distributions and mixture

[11] A more subtle form of prior is specification of the expected returns process. Brandt and Kang (2001) use a specification for the expected returns process which is always positive which is a sufficiently strong prior specification to avoid identification issues.

[12] See Jobson and Korkie (1980) or Best and Grauer (1991) for details on the sensitivity of mean-variance portfolios on estimates of expected returns.

Table 1: Annualized summary statistics for the S&P 500 index returns.

Mean	SD	Mean/SD	Sharpe Ratio
13.6	14.9	0.91	0.49

models often have identification problems. For example, a mixture of normal distributions (regime-switching models) results in an unbounded likelihood function without parameter restrictions. In a related model with constant variance, Lamoreaux and Zhou (1996) also place informative priors on the latent process driving the conditional expected value of the dependent variable.

Our approach to prior specification is intuitive and pragmatic. For β_μ, we consider three different values, $\beta_\mu = 0.98$, 0.99 and 0.995, which varies the persistence of the process. We have also considered lower values, however, the expected returns process degenerates and is extremely volatile. Given the mean-reverting specification, it is easy to calculate the persistence of expected returns at other horizons. For example, the persistence with $\beta_\mu = 0.99$ over the monthly horizon is $(.99)^{20} = 0.81$, which is reasonable. Our prior for α_μ must depend on β_μ as the unconditional mean of the volatility process is $E(\mu_t) = \frac{\alpha_\mu}{1-\beta_\mu}$. Increasing the persistence naturally increases the average expected return. We center our prior for α_μ at 10% annualized with a reasonably large standard deviation.

4 Empirical Results: S&P 500

This section provides estimation results. The S&P 500 is our risky asset and we use daily data from January 1, 1980 to December 31, 2000.[13] Table 1 provides S&P 500 summary statistics. For interest rate data, we use daily 3-month Treasury bill rates over the same time period and the mean T-Bill rate was 6.186%

We estimate and evaluate portfolio performance for the following models: (SMVC) the general model with stochastic μ_t and V_t and correlated shocks; (SMV) the special case of SMVC with no correlation, $\rho = 0$; (SV) the model with stochastic V_t but constant expected

[13] We also considered the Nasdaq 100 index from January 1, 1986 - December 31, 2000. The results were similar and we have omitted these results for brevity.

Table 2: Parameter estimates for the S&P 500 index for the various models. We report the parameter posterior mean and the standard deviation is reported below.

		Posterior						
Prior		SMVC		SMV		SV	SM	
β_μ		0.98	0.995	0.98	0.995	.	0.98	0.995
α_μ	$.04(1-\beta_\mu)$	0.837	0.248	0.868	0.260	.	0.829	0.228
$\cdot 1000$	0.100	0.094	0.073	0.092	0.072	.	0.094	0.074
σ_μ	0.0160	0.0091	0.0064	0.0087	0.0064	.	0.0096	0.0068
	0.0121	0.0017	0.0011	0.0023	0.0011	.	0.0021	0.0014
α_v	0.1205	-0.0080	-0.0093	-0.0100	-0.0103	-0.0086	.	.
	0.1001	0.0029	0.0031	0.0031	0.0032	0.0029	.	.
β_v	0.9804	0.9802	0.9780	0.9764	0.9759	0.9789	.	.
	0.0996	0.0048	0.0051	0.0050	0.0053	0.0050	.	.
σ_v	0.1921	0.1474	0.1558	0.1580	0.1607	0.1465	.	.
	0.1416	0.0164	0.0159	0.0151	0.0166	0.0167	.	.
ρ	-0.3591	-0.4916	-0.2307	.	.	.	.	.
	0.4520	0.2198	0.2122	.	.	.	.	.

return, $\mu_t = \mu$; (SM) the model with a stochastic μ_t and constant volatility, $V_t = V$ and (Constant) the model with constant mean and variance: $\mu_t = \mu$ and $V_t = V$. Note that in all of the models, any constant parameters are assumed unknown and we account for parameter estimation risk.

We use a daily sampling frequency for two reasons. First, if we view our model as a time-discretization of a continuous-time model, we know from Merton (1981) that accurate volatility estimation requires high frequency data. The high sampling frequency does not aversely effect expected returns estimates as their accuracy is a function of the total time span of the data. Second, daily data allows for daily rebalancing of optimal portfolios which comes close to the continuous-rebalancing of continuous-time portfolio problems.

4.1 Smoothing Results

4.1.1 Parameter Estimates

Table 2 provides summary statistics of the marginal posterior distribution for the parameters for S&P 500 returns, in the cases where $\beta_\mu = 0.98$ and 0.995. We report the posterior mean and posterior standard deviation for each parameter and specification. Estimates for the case $\beta_\mu = 0.99$ are not reported, as there is little additional insight beyond the other two cases. In all cases, the posterior standard deviation is substantially smaller than the prior standard deviation for all of the parameters, which indicates the uninformative nature of our priors. For example, the prior standard deviation of σ_μ, α_v, β_v and σ_v is about 10 times as large as the posterior. We can be uninformative regarding these parameters due to the fact that we estimate conditional on the value for β_μ. The posterior mean for α_μ implies a long run mean of, $252 * \alpha_\mu/(1 - \beta_\mu)$, which is equal to about 10.5%.

Notice that as the persistence of expected returns increases from $\beta_\mu = 0.98$ to 0.995, expected returns become closer to a random walk and both the σ_μ and ρ decline. Both effects are expected. Since volatility is often rapidly moving, if the correlation between expected returns and volatility shocks were high, rapid movements in volatility would induce large rapid declines in expected returns. In order to keep expected returns very persistent, expected returns cannot rapidly move and so the correlation must fall. Third, there is substantial posterior uncertainty regarding ρ, whose posterior mean is negative, but there is significant posterior mass for $\rho > 0$. This uncertainty does not disappear when expected returns become more persistent.

Finally, our estimation results indicate that the estimates of α_v, β_v and σ_v are effectively unchanged by the model specification. Since the returns data are so informative about the volatility state, the volatility parameter estimates are nearly independent of the expected returns specification. When volatility is constant in the SM model, the volatility of expected returns increases because time-varying expected returns must generate more volatility.

4.1.2 Smoothed Expected Returns and Volatility

While the parameter estimates provide insight into the roles played by the persistence and model specification, smoothed estimates of the spot expected returns and volatility provide a more intuitive view of their relative impact. Figures 1 and 2 provide smoothed expected returns and volatility estimates in a number of different cases.

Figure 1 displays smoothed expected returns estimates, $E\left(\mu_t|r^T\right)$, for different persistence levels and models. The top panel, panel A, shows the impact of persistence in the SMVC model with the solid line showing the smoothed expected returns in the case $\beta_\mu = 0.995$ and the dashed line the case where $\beta_\mu = 0.98$. Clearly, increasing the persistence smooths out the expected returns path, but it is important to note that the differences can be quite large.

Panel B compares expected returns paths in the SMVC and SMV model for $\beta_\mu = 0.98$ and shows the impact of the correlation between the shocks to the latent state variables, ρ. There are large downward spikes in expected returns in the SMVC model that are driven by rapid increases in volatility. Rapid increases in volatility signal rapid declines in expected returns as $\rho < 0$. These spikes downward are not present in the SMV model as there is no link between shocks in μ_t and V_t. In the SMV model, expected returns do not drastically change during period of high volatility such as the episodes in 1987, 1997 and 1998. For example, in 1987, the SMVC models estimate μ_t is less than -10% during the crash, while the SMV model has expected returns close to its mean, about 10%. In a sense, during these periods, the noise (volatility) is so great that it is difficult to extract the signal (expected returns) unless there is a prior link between the two. Finally, panel C displays smoothed expected returns for the SMV and SM model for $\beta_\mu = 0.98$. In the SM model, the expected returns process must move more rapidly to generate large shocks, such as those in 1987.

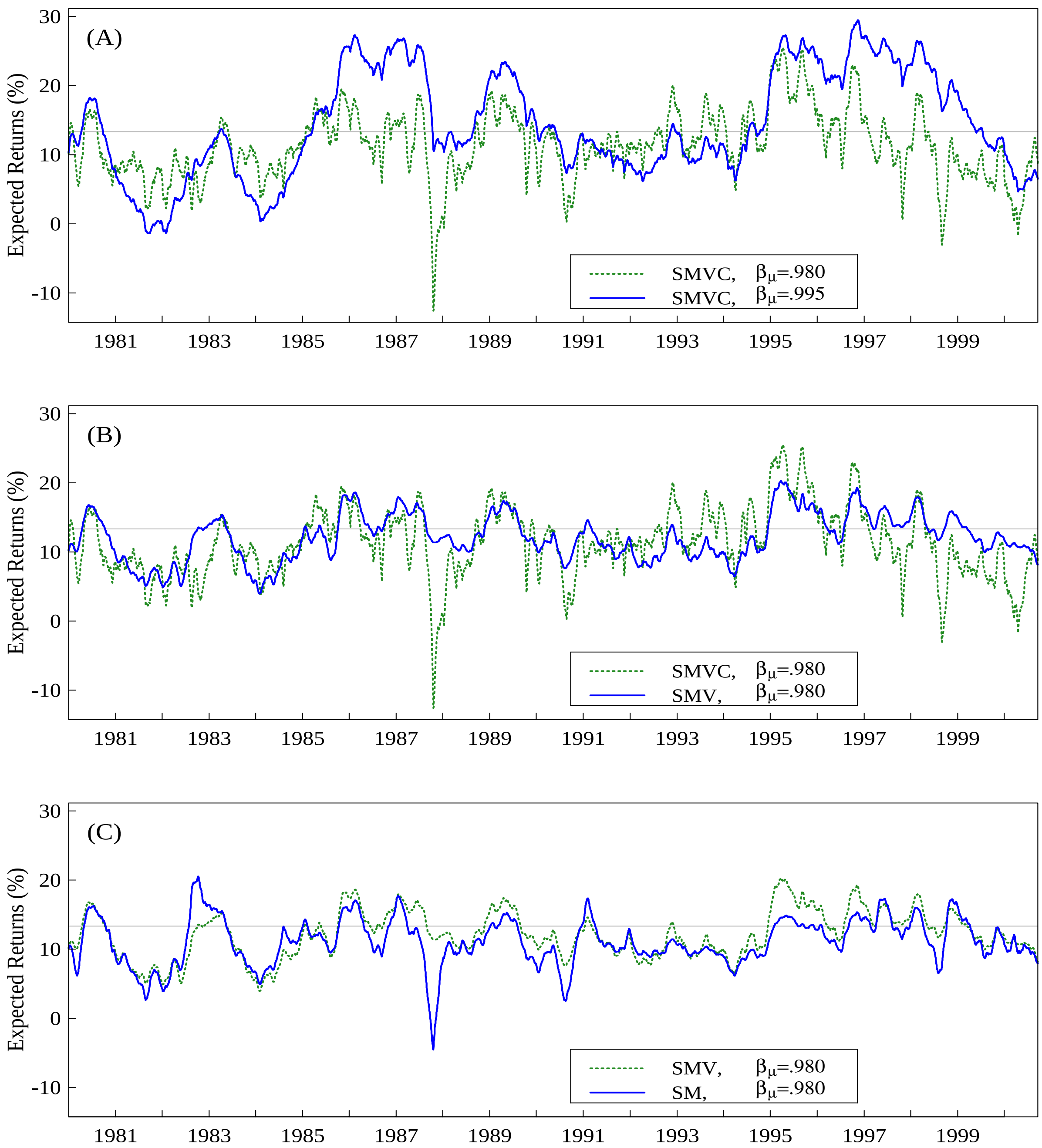

Figure 1: This figure show the effect of expected return persistence (changing β_μ) and different specifications.

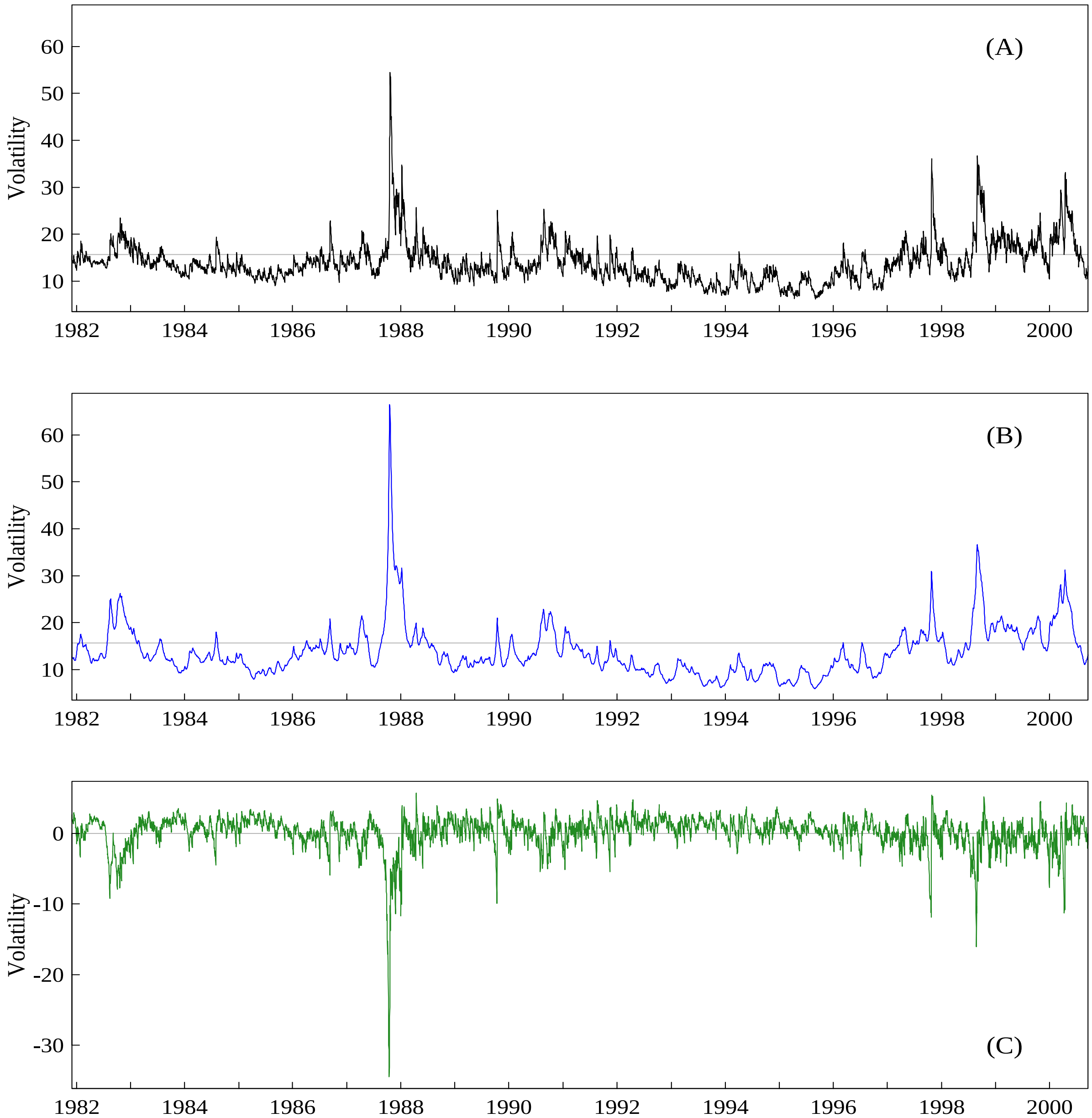

Figure 2: This figure shows the difference between filtered and smoothed estimates of volatility in the SMVC model for $\beta_\mu = 0.98$. Panel A shows filtered volatility estimates, Panel B shows smoothed volatility estimates and Panel C shows their difference.

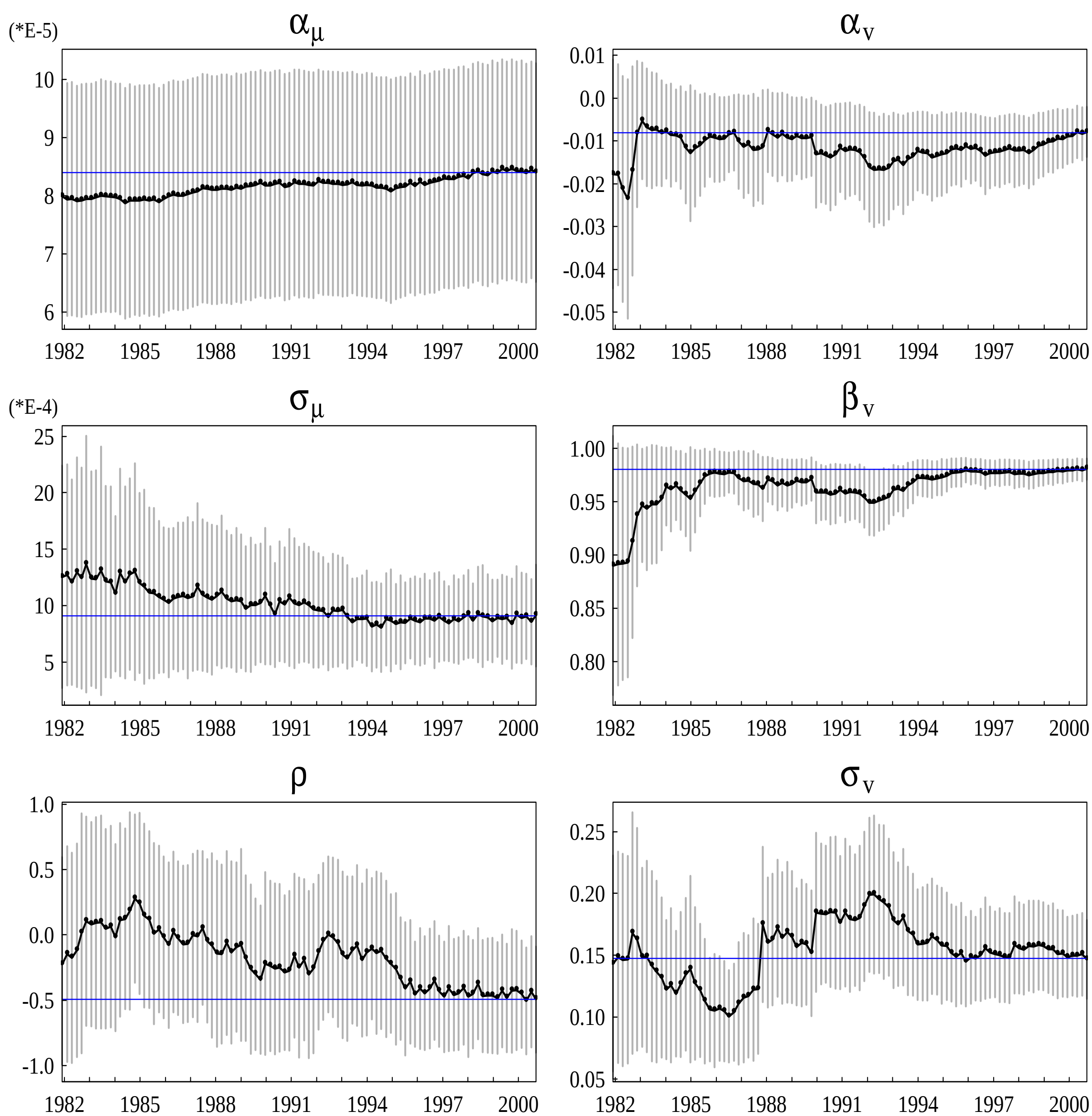

Figure 3: Sequential parameter learning plots for the SMVC model with $\beta_\mu = 0.98$. The straight line is the smoothed posterior mean, the solid line is the posterior mean at a given point in time and the shaded lines are a 2-standard deviation posterior confidence band.

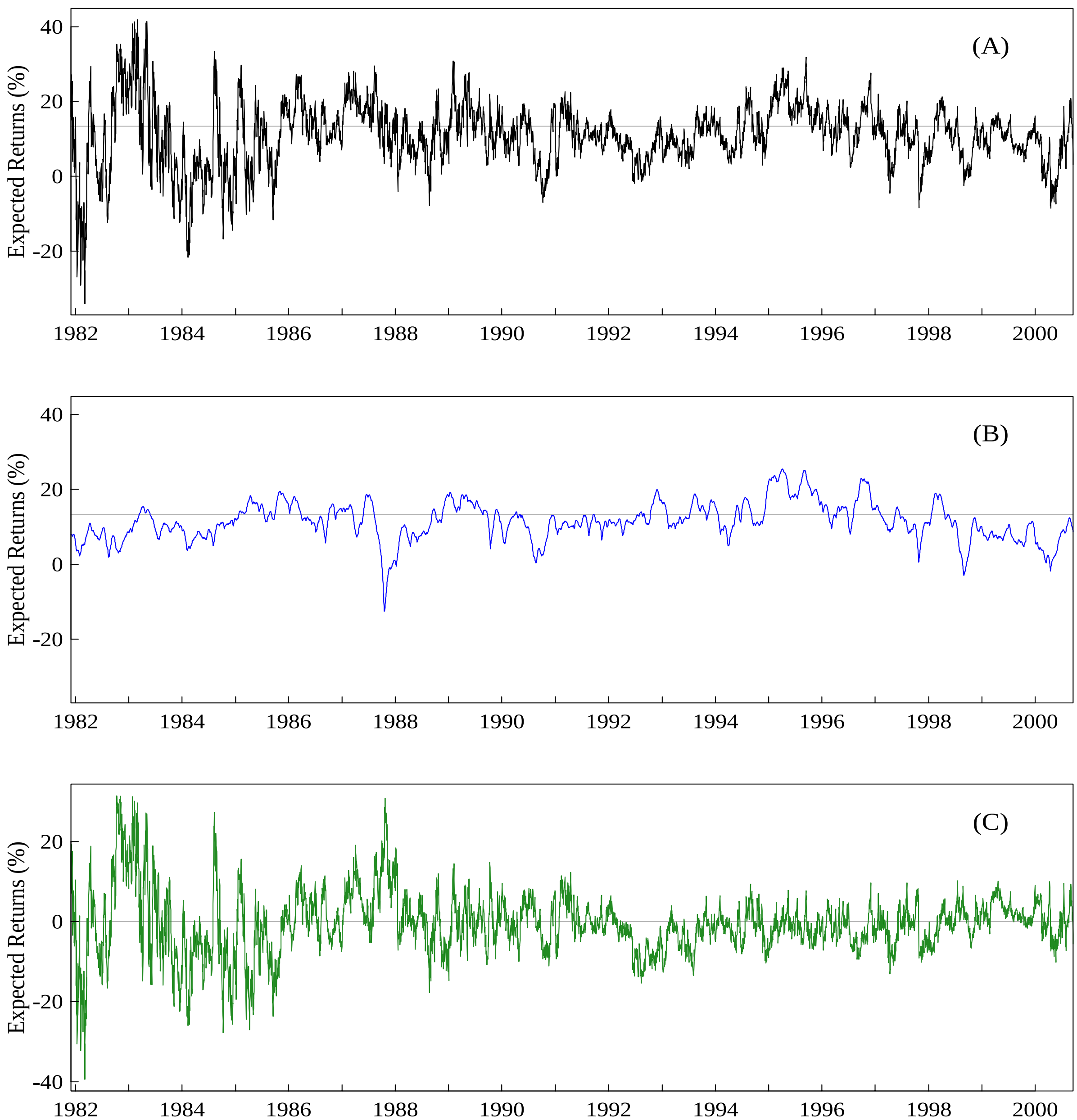

Figure 4: This figure shows the difference between filtered and smoothed estimates of expected returns in the SMVC model for $\beta_\mu = 0.98$. Panel A shows filtered estimates, Panel B shows smoothed estimates and Panel C shows their difference.

The top panel of Figure 2 displays smoothed estimates of volatility, $E\left(V_t|r^T\right)$, for the SMVC model. The smoothed estimates for the other models are similar and are therefore not reported.

4.2 Sequential Learning and Estimation

We now discuss sequential inference which is summarized by filtered estimates of expected returns, volatility and the parameters. These estimates are given by $E\left[\mu_t|r^t\right]$, $E\left[V_t|r^t\right]$ and $E\left[\theta|r^t\right]$. Since our sequential estimates incorporate estimation risk, we do not start filtering with the first observation, as there is no information regarding the parameters. Rather, we assume the investor uses the first 500 observations as a warm-up period and begins filtering with $t = 501$. To our knowledge, this paper is the first paper to report filtered estimates for expected returns and volatility and also to address sequential parameter learning problem using actual data.

4.2.1 Sequential Parameter Learning

As investors observe data, they naturally revise their beliefs about the parameter values. Figure 3 provides the sequential parameter estimates for the SMVC model,[14] the most general specification. Figure 3 reports the posterior means sequentially through time and a two standard deviation symmetric coverage interval for each of the parameters for the entire sample. That is, for each parameter θ we plot $E\left[\theta|r^t\right]$ and construct a 2-standard deviation confidence interval using $std\left[\theta|r^t\right]$. The solid constant line provides the full-sample posterior mean for each of the parameters, $E\left[\theta|r^T\right]$. Deviations from this line indicate the effect of sequential parameter learning. Note that the filtered estimates converge to the smoothed estimates as the sample size increases.

The most noticeable feature of the sequential learning is the substantial time-variation in parameter estimates, especially in σ_v and ρ. For example, prior to the crash in 1987, the posterior mean for σ_v was about 0.12 and after the crash it jumped upward to almost 0.17. This is indicative of a structural shift in volatility of volatility, as σ_v has remained high

[14] The sequential parameter estimates for the other models were similar and were therefore omitted.

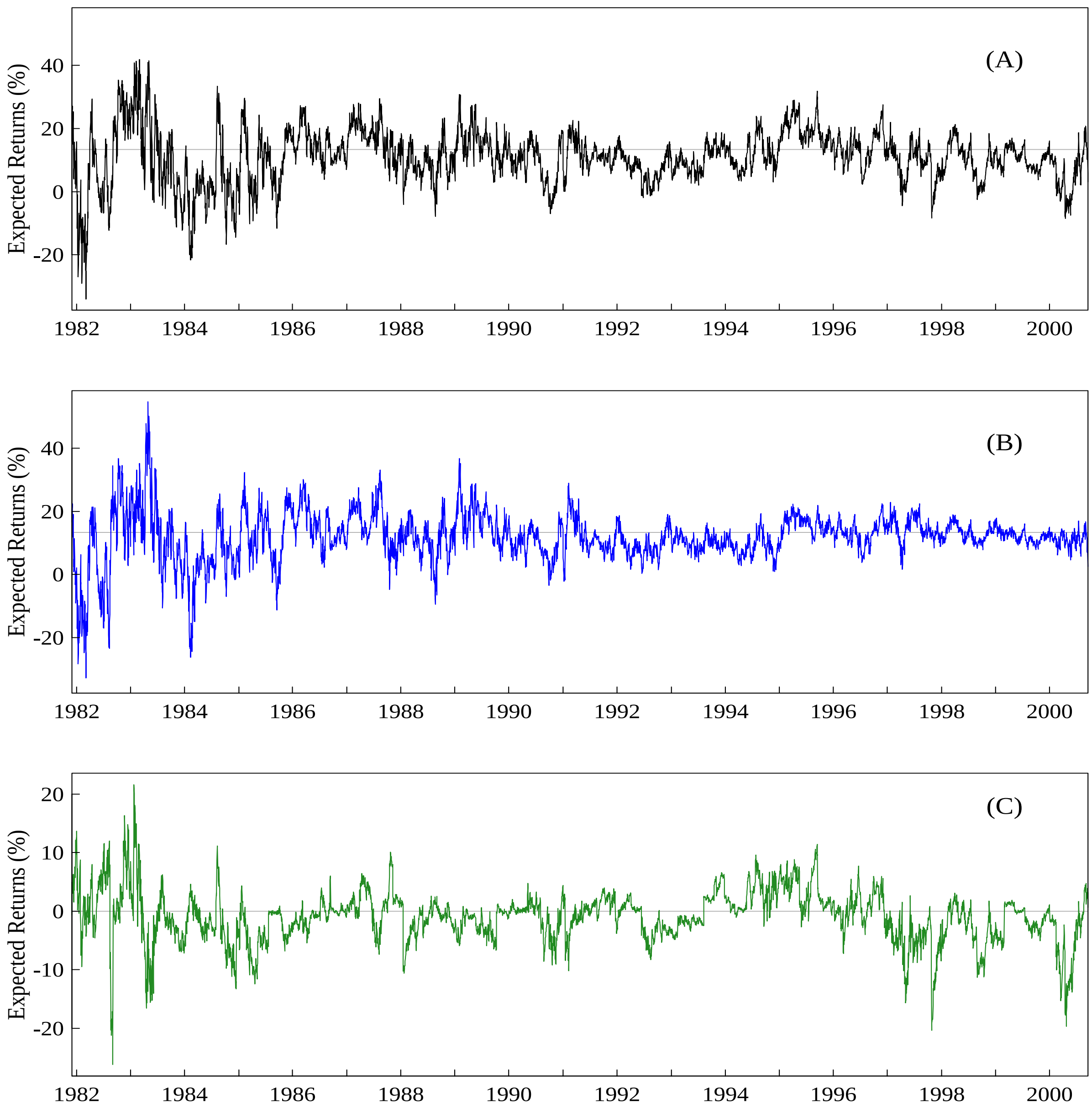

Figure 5: This figure shows the difference between filtered estimates in the SMVC and SMV models. Panel A displays filtered expected return estimates in the SMVC model for $\beta_{\mu} = 0.98$, Panel B displays filtered expected return estimates in the SMV model for $\beta_{\mu} = 0.98$and Panel C displays the difference.

throughout the decade of the 1990s.[15]

Even more striking is the variation in ρ. The full-sample estimate of ρ is about -0.5, but the sequential estimates vary dramatically. For example, the posterior mean of ρ has switched signs from positive in the early 1980s to nearly zero in the early 1990s. The sequential variation in ρ provides an explanation for why it is so difficult to pin down the relationship between expected returns and volatility: the investor's perception of this relationship changes drastically over time depending on how much data is available. This could, in part, explain the disparate results in the literature regarding the nature of the contemporaneous relationship between expected returns and volatility (see also Whitelaw (1994) and Harvey (2001)).

Why do these sequential parameter estimates change so drastically? A compelling argument is that, in this setting with unobserved expected returns and volatility, 3000 data points is still a relatively small sample. The reason for this is that we are estimating the parameters that drive expected returns and volatility, both of which are persistent processes and it is difficult to estimate the parameters of persistent processes.

4.2.2 Filtering Expected Returns and Volatility

Figures 2, 4, 5 and 6 provide filtered expected returns and volatilities to highlight the difference between filtered and smoothed estimates, the effect of estimation risk, the impact of learning and the effect of correlation. Specifically, Figure 2 displays filtered volatilities (panel A), smoothed volatilities (panel B) and their difference (panel C) for the SMVC model with $\beta_\mu = 0.98$. Filtered volatilities are far noisier than smoothed and the difference between smoothed and filtered volatilities can be quite large, especially in periods of market stress such as 1987, 1997 and 1998.

Figure 4 displays filtered expected returns (panel A), smoothed expected returns (panel B) and their difference (Panel C) for the SMVC model with $\beta_\mu = 0.98$. These estimates take into account estimation risk. There are two effects. First, the filtered estimates do not use future returns to estimate the current expected return or volatility state, which explains why the smoothed estimates of μ_t are much "smoother" than their filtered counterparts. Second, estimation risk has a greater impact on the filtered estimates during the early portion of the

[15] Alternatively, this could be driven by model misspecification and the need to incorporate jumps in returns or in volatility, see, for example, Eraker, et al. (2001).

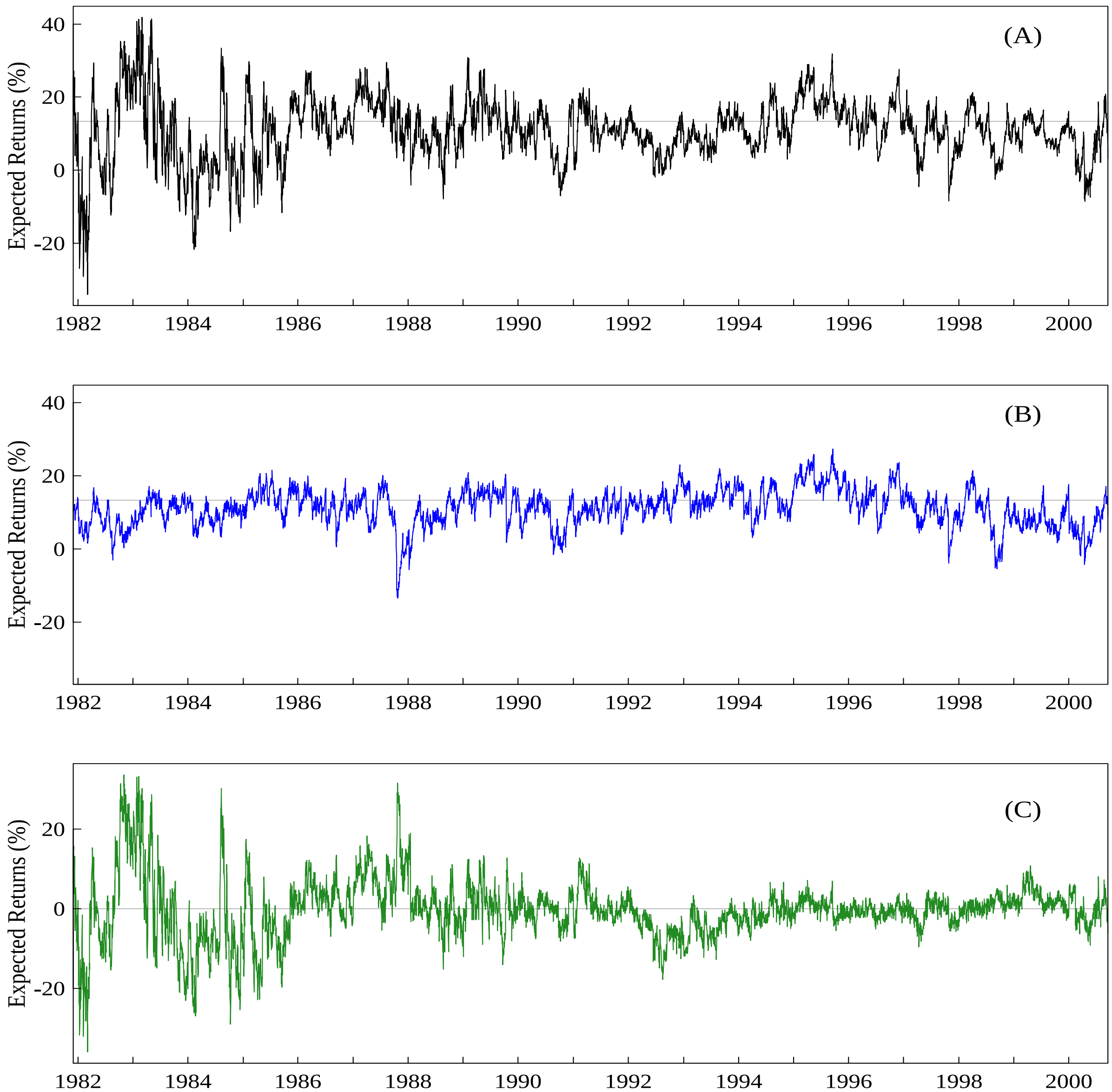

Figure 6: This figure shows the impact of estimation risk and parameter learning on forecasting expected returns in the SMVC model. Panel A displays filtered expected return estimates taking into account estimation risk, Panel B displays filtered expected return estimates conditional on the full-sample parameter estimates, and Panel C displays their difference.

Table 3: Characteristics of optimal portfolio returns for the SMVC model. The portfolio return mean, standard deviation and certainty equivalent are all in percentages.

	Leverage						No Leverage					
γ	Mean	SD	$\frac{\text{Mean}}{\text{SD}}$	S.R	CE1	CE2	Mean	SD	$\frac{\text{Mean}}{\text{SD}}$	S.R	CE1	CE2
						$\beta_\mu = 0.980$ Parameter Learning						
2	11.2	18.8	0.60	0.18	-0.52	-3.71	9.0	10.9	0.83	0.41	0.16	-3.60
4	8.8	14.2	0.62	0.20	-1.27	-4.40	8.4	9.4	0.89	0.48	0.44	-2.53
6	7.8	10.8	0.72	0.31	-1.09	-2.64	7.7	8.2	0.94	0.52	0.29	-1.26
						$\beta_\mu = 0.980$ No Parameter Learning						
2	12.7	17.2	0.74	0.32	1.55	-1.64	10.4	10.2	1.02	0.60	1.71	-2.02
4	9.9	13.1	0.76	0.34	0.43	-2.70	9.2	8.6	1.07	0.65	1.53	-1.44
6	8.9	10.3	0.86	0.44	0.03	-1.22	8.3	7.4	1.12	0.71	1.13	-0.28
						$\beta_\mu = 0.995$ Parameter Learning						
2	12.8	21.3	0.60	0.19	0.07	-3.12	12.8	21.3	0.60	0.19	0.61	-3.12
4	11.0	17.9	0.61	0.20	-1.45	-4.57	11.0	17.9	0.61	0.20	-1.60	-4.57
6	9.8	14.9	0.66	0.24	-2.25	-3.80	9.8	14.9	0.66	0.24	-2.25	-3.80

sample as there is more uncertainty regarding the parameters.

Figure 5 displays the difference between filtered expected returns in the SMVC and SMV models when $\beta_\mu = 0.98$. Panels A and B provide filtered expected returns for the SMVC and SMV models, respectively, and panel C displays their difference. While there are large differences for many periods, there is surprisingly little difference during certain periods, such as the crash of 1987. When discussing the smoothed estimates of μ_t, we found that the expected returns estimates in the SMVC and SMV models differed the most in periods of high volatility. This effect does not occur in the filtered estimates in 1987, as the investor perceived that the correlation between expected returns and volatility was in fact close to zero

Table 4: Characteristics of optimal portfolio returns for the SMV model. The portfolio return mean, standard deviation and certainty equivalent are all in percentages.

	Leverage						No Leverage					
γ	Mean	SD	$\frac{\text{Mean}}{\text{SD}}$	S.R	CE1	CE2	Mean	SD	$\frac{\text{Mean}}{\text{SD}}$	S.R	CE1	CE2
$\beta_\mu = 0.980$ Parameter Learning												
2	13.2	19.9	0.66	0.25	1.05	-2.14	11.0	11.7	0.94	0.53	1.97	-1.75
4	10.2	14.7	0.69	0.28	-0.16	-3.28	9.4	9.9	0.95	0.53	1.24	-1.72
6	9.0	11.2	0.80	0.39	-0.16	-1.70	8.4	8.5	0.99	0.57	0.84	-0.71
$\beta_\mu = 0.980$ No Parameter Learning												
2	15.8	19.4	0.81	0.40	3.85	0.65	13.0	12.3	1.05	0.64	3.83	0.11
4	11.5	12.8	0.90	0.48	2.18	-0.90	10.7	9.7	1.10	0.69	2.62	-0.34
6	9.9	9.3	1.06	0.65	1.91	0.37	9.3	7.8	1.19	0.78	2.10	0.54
$\beta_\mu = 0.995$ Parameter Learning												
2	12.8	21.3	0.60	0.19	0.07	-3.11	10.9	12.3	0.89	0.47	1.73	-1.99
4	11.0	17.9	0.61	0.20	-1.45	-4.46	9.8	11.0	0.89	0.48	1.19	-1.78
6	9.8	14.9	0.66	0.24	-2.25	-3.80	9.2	10.1	0.91	0.50	0.75	-0.80

(see Figure 3). During the periods of high volatility in 1997, 1998 and 2000, when the investor perceived that ρ was negative, the SMVC and SMV models have drastically different filtered expected returns.

Finally, Figure 6 provides the filtered expected returns with and without parameter learning and estimation risk in the SMVC model with $\beta_\mu = 0.98$. Panel A displays filtered expected returns taking into account estimation risk and parameter learning. Panel B displays filtered expected returns conditional on the full sample posterior means and Panel C their difference. The differences are quite large. During the beginning of the sample, parameter uncertainty is

Table 5: Characteristics of optimal portfolio returns for the SV model. The portfolio return mean, standard deviation and certainty equivalent are in percentages.

	Leverage						No Leverage					
γ	Mean	SD	$\frac{\text{Mean}}{\text{SD}}$	S.R	CE1	CE2	Mean	SD	$\frac{\text{Mean}}{\text{SD}}$	S.R	CE1	CE2
						Parameter Learning						
2	16.0	17.0	0.94	0.53	4.92	1.73	13.6	12.3	1.11	0.69	4.43	0.71
4	11.3	10.0	1.13	0.71	3.26	0.14	10.8	8.5	1.27	0.86	3.16	0.20
6	9.6	6.8	1.41	1.00	2.82	1.27	9.2	6.3	1.46	1.05	2.62	1.07
						No Parameter Learning						
2	16.1	17.0	0.95	0.53	5.02	1.83	13.5	12.2	1.11	0.69	4.36	0.63
4	11.2	10.2	1.10	0.68	3.08	-0.04	10.8	8.5	1.27	0.86	3.16	0.20
6	9.6	7.1	1.35	0.94	2.70	1.11	9.2	6.4	1.44	1.02	2.58	1.03

the highest and naturally this translates into more volatile expected returns estimates. Toward the end of the sample, the sequential parameter estimates are closer to the full sample estimates, but there are still substantial differences in filtered expected returns. This shows the strong effect of parameter estimation risk (as opposed to parameter learning) in estimating expected returns.

Table 6: Characteristics of optimal portfolio returns for the SM model. The portfolio return mean, standard deviation and certainty equivalent are in percentages.

	Leverage						No Leverage					
γ	Mean	SD	$\frac{\text{Mean}}{\text{SD}}$	S.R	CE1	CE2	Mean	SD	$\frac{\text{Mean}}{\text{SD}}$	S.R.	CE1	CE2
$\beta_\mu = 0.980$ Parameter Learning												
2	12.9	20.1	0.64	0.23	0.67	-2.52	10.2	11.3	0.90	0.49	1.27	-2.46
4	10.6	14.6	0.73	0.31	0.30	-5.04	9.3	10.0	0.93	0.51	1.10	-4.08
6	9.3	11.0	0.85	0.43	0.28	-5.71	8.6	8.6	1.00	0.58	0.99	-5.00
$\beta_\mu = 0.980$ No Parameter Learning												
2	15.5	18.3	0.85	0.43	3.96	-0.77	13.3	12.1	1.10	0.68	4.18	0.45
4	10.3	9.5	1.08	0.67	2.45	-0.65	10.5	9.1	1.15	0.73	2.65	-0.32
6	8.8	6.4	1.38	0.96	2.18	0.63	8.8	6.4	1.38	0.96	2.18	0.63
$\beta_\mu = 0.995$ Parameter Learning												
2	13.5	20.7	0.65	0.24	1.03	-2.16	10.0	11.0	0.91	0.49	1.13	-2.59
4	12.1	17.7	0.68	0.27	-0.21	-3.32	9.6	10.3	0.93	0.52	1.28	-1.68
6	10.4	14.5	0.72	0.30	-1.30	-2.85	9.2	9.6	0.96	0.54	1.04	-0.50

5 The Economic Benefits of Predictability

5.1 Measuring Economic Benefits

Our approach measures the benefits of return predictability by comparing the out-of-sample performance of optimal portfolios.[16] This provides an intuitive metric to measure the gains

[16]It is important to recognize that very few researchers compute truly out-of-sample *optimal* portfolio returns. While there is a large literature that analyzes the out-of-sample returns to technical and other ad-hoc timing rules (see, e.g., Brock, Lakonishok and LeBaron (1992) and Peseran and Timmermann (1995)), these results typically ignore the utility concerns of the investor.

Table 7: Characteristics of optimal portfolio returns for the constant model, a model where expected returns and volatility are constant, but unknown parameters. The portfolio return mean, standard deviation and certainty equivalent are all in percentages.

	Leverage					No Leverage				
γ	Mean	SD	$\frac{\text{Mean}}{\text{SD}}$	S.R	CE2	Mean	SD	$\frac{\text{Mean}}{\text{SD}}$	S.R	CE2
Parameter Learning										
2	11.8	19.0	0.62	0.21	-3.19	9.0	11.6	0.78	0.36	-3.73
4	8.0	9.9	0.81	0.39	-3.12	8.0	9.5	0.84	0.42	-2.96
6	6.7	6.6	1.02	0.60	-1.55	6.7	6.6	1.02	0.60	-1.55
No Parameter Learning										
2	22.3	27.5	0.81	0.40	3.36	14.0	14.8	0.95	0.53	0.43
4	14.8	14.9	0.99	0.58	1.20	14.0	13.7	1.02	0.61	1.09
6	11.7	9.9	1.18	0.77	1.82	11.7	9.9	1.18	0.77	1.82

or losses of return predictability. Our approach for evaluating the economic benefits is similar to the approach of Brennan, et al. (1996) who examine out-of-sample returns for a timing strategy based on time variation in the expected returns in bonds, stock and cash.[17] They solve a multi-period portfolio problem, ignoring estimation risk and learning, two factors which we find to be important.

Given the filtered expected returns and volatility, we compute optimal portfolios daily and calculate portfolio returns. This provides a long time series of out-of-sample portfolio returns to evaluate the performance of the different models and the impact of estimation risk and parameter learning.

There are a number of alternative approaches for measuring the economic benefits of

[17] Fleming, Kirby and Ostdiek (2000, 2001) compute portfolio returns using out-of-sample volatility estimates but in-sample expected returns estimates.

predictability. Lynch (2001), for example, calculates the ex-ante benefits as measured by the amount an investor would pay to have access to certain variables which are assumed to predict future returns. This calibration or population approach provides a measure of economic costs and benefits of predictability but ignores the fact that the true nature of the predictability is uncertain. If parameters are unknown and must be estimated, the ex-ante optimal portfolio may have poor ex-post performance as the ex-ante measures assume the investor knows the true nature of the predictability.

An alternative, but related approach is used by Chapman, et al. (2001). Their investor chooses an optimal portfolio after running predictability regressions on returns that are simulated from a general equilibrium model. One of their main conclusions is that "calibration exercises based on point estimates can be very misleading" (Chapman, et al. (2001), p. 3) and they also find that estimation risk plays a crucial role in portfolio performance. Our approach provides a natural complement to these other population based approaches for studying the benefits of predictability.

Comparing the performance of dynamic trading strategies is difficult because of market timing, the potential for leverage and non-normal returns. Due to these factors, many standard metrics for evaluating portfolio performance, such as the Sharpe Ratio, cannot be used when evaluating dynamic trading strategies.[18]

We report a number of different summary statistics to evaluate the portfolio performance. First, we report the portfolio mean and standard deviation as well as their ratio. Second, we report the adjusted Sharpe ratio of Graham and Harvey (1997) which, to a certain extent, mitigates the effect of leverage and market timing, while still retaining the attractive intuition of the Sharpe ratio. This metric matches the portfolio return volatility to that of the benchmark (the S&P 500) by adjusting expected return up or down to obtain a better measure of return for a given level of risk (the market's risk). For example if a managed portfolio had a 8% average return and 10.5% volatility, we would lever/de-lever our holdings in the managed portfolio match the volatility of the S&P 500 (14.9%), by multiplying the expected returns by the ratio of S&P 500 to portfolio volatility. In this example, we would multiply the expected return by 1.41 (0.149/0.105) to get a risk-adjusted return of 11.3%. Given this risk adjusted

[18]Leland (1997, page 32) argues that "applying the Sharpe ratio to a portfolio with nonlognormal returns will, in general, produce nonsense as a measure of manageral ability."

return, we compute the usual Sharpe ratio on a constant volatility basis.

Finally, we use a utility based metric based on certainty equivalent return. On economic grounds, this is the most relevant metric as it quantifies benefits based on investor preferences. For each of the models we consider, we compute the annualized risk free return which would match utility with that of the no predictability or buy-and-hold (sub-optimal) benchmarks. CE1 is the certainty equivalent gain from the model with no predictability and CE2 is the certainty equivalent gain relative to the buy-and-hold strategy.

To insure that our results are due to portfolio leverage, we report results allowing for leverage (up to twice wealth, consistent with current margining practices) and without leverage. In both cases, for simplicity we do not allow for short selling, although the results are substantively unchanged when allowing for leverage.

5.2 Economic Benefits: S&P 500

Tables 3-7 summarize optimal portfolio performance for all of our models. We also analyze the impact of parameter uncertainty. In the case without parameter uncertainty by setting the parameters equal to the posterior means obtained using the full sample. We report the results for a range of risk aversions.

The most striking result is that volatility timing (the SV model) uniformly dominates the other strategies in terms of Sharpe ratios and certainty equivalent gain for all risk aversions and with or without leverage. Moreover, the SV model also outperforms (suboptimal) buy-and-hold strategy in terms of Sharpe ratios and certainty equivalent gains. For example, an investor with a risk aversion of 6 has an Sharpe ratio of 1.00 with leverage and 1.05 without leverage. The annualized certainty equivalent gain from the no predictability optimal strategy is 2.82% with leverage and 2.62% without leverage.

Volatility timing works well for two reasons. First, volatility timing is a successful portfolio strategy. Although it is difficult to outperform the market in terms of expected returns, reducing portfolio volatility is easier. The intuition for this is in the filtered volatility estimates in Figure 2. In periods of market stress, forecasted volatility is high and the investor naturally reduces their exposure in these periods. Second, the SV strategy is not adversely affected by estimation risk. In fact, the portfolio rule that accounts for estimation risk performs marginally better than the rule that ignores it.

Surprisingly, the market timing strategy performs worse than the buy-and-hold strategy in all cases. Table 6 provides portfolio returns in the SM model and shows that while the pure market timing strategy does worse than the SV model. The market timing strategy does provide some economic gains to predictability in terms of the certainty equivalent gain over the constant model (CE1). This implies that expected return predictability still provides economic benefits.

Unlike the simple stochastic volatility model, models with time-varying expected returns suffer severely from estimation risk. Table 6 compares the performance of the SM model with and without parameter uncertainty. If the parameters were known (set to the full-sample posterior means), the SM models performs almost as well as the SV model. For example, for $\beta_\mu = 0.98$ and $\gamma = 6$, the SM model attains a Sharpe ratio of 0.96, and an annualized certainty equivalent gain of 2.18% over the no predictability strategy. These fall drastically to a Sharpe ratio of 0.43 and certainty equivalent of 0.28% when parameter uncertainty is taken into account.

This result was foreshadowed earlier in Section 4.2.2 where Figure 6 showed the dramatic impact of parameter uncertainty on the filtered returns. Filtered expected returns substantively change depending on whether or not estimation risk is accounted for. This implies that conditioning on full-sample estimates when forming optimal portfolios leads to misleading results for the SM model, while it results in no change for the SV model.

Table 3 and 4 summarize the portfolio returns for the models incorporating both time-varying expected returns and volatility. The full model with correlation between the shocks to expected returns and volatility performs the worst in terms of certainty equivalent, with or without leverage and for nearly all risk aversions. The SMVC model is also outperformed by the SMV model even when taking into account the parameter uncertainty in ρ. These results clearly indicate that the simpler models outperform their more complicated counterparts, a result reminiscent of Occam's razor.

Table 7 summarizes the portfolio performance assuming the expected returns and volatility are constant, though unknown, parameters. Finally, Figure 7 displays the unconstrained portfolio weights for all the models with time-variation. The strategies based on time-varying expected returns look similar while the SV model portfolio weights (the second from the bottom panel) are quite different. These portfolio weight plots shows that the volatility timing strategy changes weights relatively smoothly and so transactions costs will have little economic effect. [19]

6 Conclusion

We study the economic benefits of market and volatility timing. To do this, we developed and implemented sequential techniques to obtain out-of-sample estimates for time-varying expected returns, volatility and parameters. Given these estimates, we investigated the performance of optimal portfolio rules and focused on the relative importance of time-varying state variables, estimation risk and parameter learning.

Our most important result is that a volatility timing strategy, based on simple stochastic volatility model, uniformly dominates market timing strategies. This is measured by the out-of-sample returns to optimal portfolios and holds for a range of plausible risk aversions. The reason for this is twofold. First, it is relatively easy to forecast volatility and, moreover, forecasting volatility is relatively immune to estimation risk. Second, as theory would suggest, volatility timing appears to be a useful recipe for increasing portfolio returns, and especially for decreasing portfolio volatility.

In contrast, market timing strategies provide few, if any, economic benefits. While theory suggests investors should time the market based on expected returns, in practice this is difficult due to the risk present in estimating the parameters that drive the expected returns process. For example, in complete information setting where parameters are known, a model with only stochastic expected returns outperforms the other models, while, when estimation risk is taken into account, its performance deteriorates leading to a significant economic performance reduction as measured by either Sharpe Ratios or certainty equivalents. Our results indicate that simpler models appear to outperform their more complicated counterparts, a result consistent with Occam's Razor.

[19] Moreover, in practice, the portfolios would be implemented in the futures market which has extremely small transactions costs. For example, current round-trip fees for a member of the Chicago Mercantile Exchange for an S&P 500 E-Mini futures contract which pays $25 times the index level is about $0.15 per round trip. We did not implement the portfolio rule using futures data because futures were only introduced in late 1982 and were not sufficiently liquid until a number of years later.

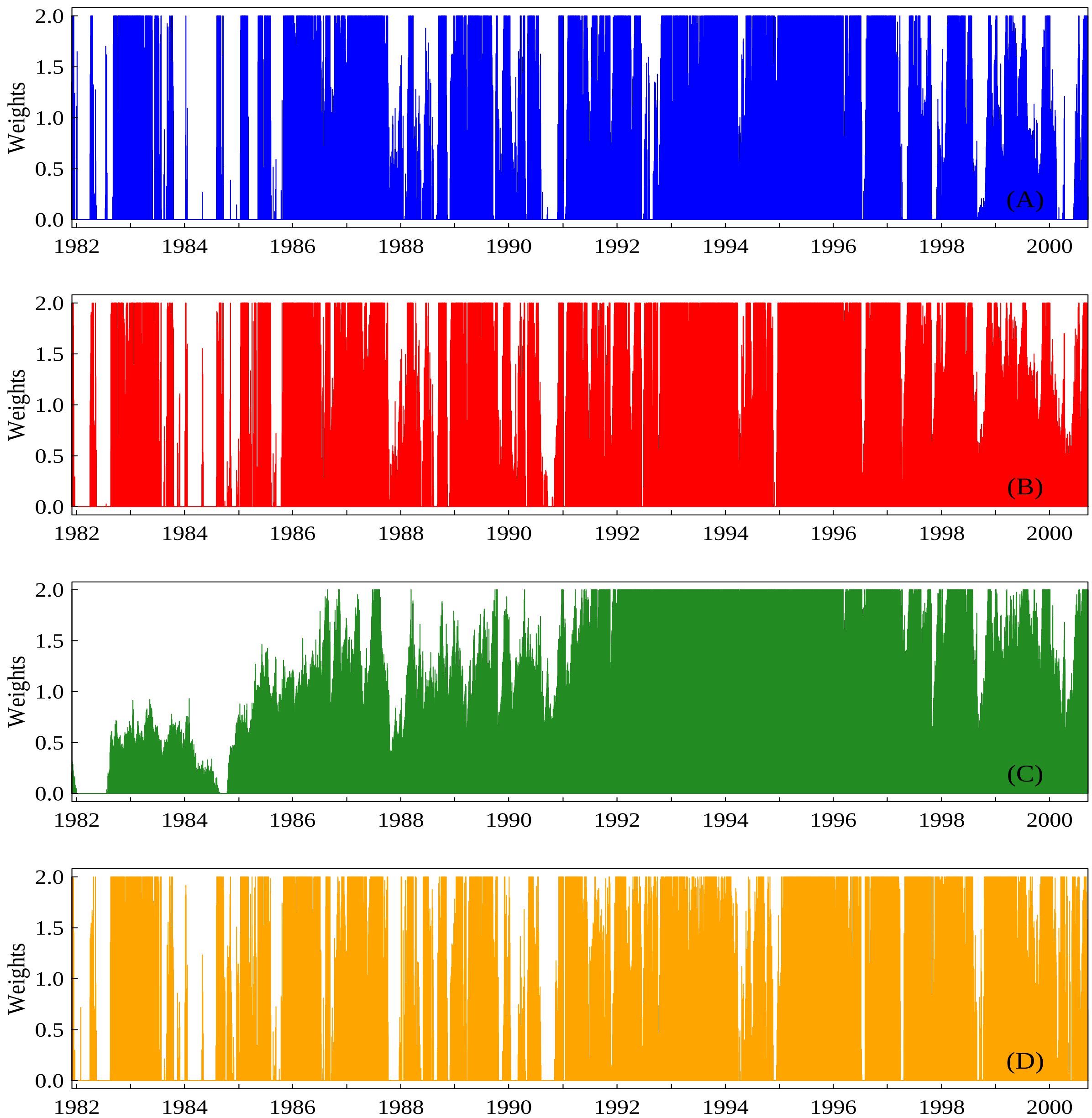

Figure 7: This figure displays the optimal portfolio holdings for the SMVC model (top panel) the SMV model (second panel), the SV model (third panel) and the SM model (bottom panel).

In future work, we plan a number of extensions. First, it is straightforward (although computationally more difficult) to add jumps in returns or a leverage effect (correlation between shocks to returns and volatility) to the model specification. Estimation and filtering proceed in the same manner, although, in the case of jumps, the portfolio rule would change as the model. Moreover, as noted by Liu, Longstaff and Pan (2002), it is never optimal to hold a levered portfolio with normally distributed jumps in returns. Second, it is also straightforward to extend the estimation and filtering methods to the multivariate setting, either via explicit models for time-varying correlations or via a factor model. Finally, it is possible to study the effect of rebalancing frequency on the performance of optimal rules.

References

Aït-Sahalia, Yacine and Michael Brandt, Variable Selection for Portfolio Choice, *Journal of Finance,* 56, 2001, 1297-1351.

Andersen, T., L. Benzoni and J. Lund. (2001) "Towards an Empirical Foundation for Continuous-Time Equity Return Models." Forthcoming *Journal of Finance.*

Ang, Andrew and Geert Bekaert, 2000, International Asset Allocation with Time-Varying Correlations, working paper, Columbia University.

Balduzzi, P. and A. Lynch, 1999, Transaction Costs and Predictability: Some Utility Cost Calculations, *Journal of Financial Economics* 52, 47-78.

Barberis, Nicholas, 2000, Investing for the Long Run when Returns are Predictable, *Journal of Finance 55*, 225-264.

Bawa, Vijay, Stephen Brown and Roger Klein, 1979, *Estimation Risk and Optimal Portfolio Choice*, (North Holland, Amsterdam).

Bernardo, Tony and Adrian Smith, 1994, *Bayesian Theory*, John Wiley & Son Ltd.

Best, Michael and R. Grauer, On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results, *The Review of Financial Studies*, 2, 315-342

Black, Fischer and Robert Litterman, 1991, Asset Allocation: Combining Investor Views with Market Equilibrium, *The Journal of Fixed Income, Sept. 1991.*

Bossaerts, Peter and Pierre Hillion, 1999, Implementing statistical criteria to select return forecasting models: What do we learn?, *Review of Financial Studies* 12, 405-428.

Brandt, Michael, 1999, Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach, *Journal of Finance,* 54, 1999, 1609-1646.

Brandt, Michael and Qiang Kang, 2002, On the Relationship Between the Conditional Mean and Volatility of Stock Returns: A Latent VAR approach, working paper, University of Pennsylvania.

Breen, William, Larry Glosten and Ravi Jaganathan, 1991, Economic Significance of Predictable Variations in Stock Index Returns, *Journal of Finance* 44, 5.

Brennan, Michael, 1998, The Role of Learning in Dynamic Portfolio Decisions, *European Finance Review,* 1, 295-306.

Brennan, Michael, Eduardo Schwartz and Ronald Lagnado, 1997, Strategic Asset Allocation, *Journal of Economic Dynamics and Control 21*, 1377-1403.

Brennan, Michael, Yihong Xia, 2001, Stock Return Volatility and Equity Premium, *Journal of Monetary Economics* 47, 249-283, April 2001)

Campbell, John, Yeung Lewis Chan and Luis Viceira, 2002, A Multivariate Model of Strategic Asset Allocation" forthcoming *Journal of Financial Economics.*

Campbell, John and Robert Shiller, 2001, Valuation Ratios and the Long-Run Stock Market Outlook: An Update, NBER Working Paper No. 8221.

Campbell, John and Luis Viceira, 1999, Consumption and Portfolio Decisions When Expected Returns are Time Varying, *The Quarterly Journal of Economics*, 114:433-495.

Chacko, George and Luis Viceira, 2002, Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets, working paper, Harvard University.

Chernov, M., E. Ghysels, R. Gallant and G. Tauchen (2001). Alternative Models for Stock Price Dynamics, Working paper, Columbia University.

Comon, Etienne, 2001, Extreme Events and the Role of Learning in Financial Markets, working paper, Harvard University.

Cremers, Martijn, 2002, Stock Return Predictability: A Bayesian Model Selection Perspective, forthcoming *The Review of Financial Studies.*

Eraker, Bjorn, Michael Johannes and Nick Polson, 2002, The Role of Jumps in Returns and Volatility, forthcoming, *The Journal of Finance.*

Fleming, Jeff, Chris Kirby and Barbara Ostdiek, 2001a, The Economic Value of Volatility Timing, *Journal of Finance* 56, 329-352.

Fleming, Jeff, Chris Kirby and Barbara Ostdiek, 2001b, The Economic Value of Volatility Timing Using Realized Volatility, working paper, Rice University.

French, Kenneth, G. William Schwert and Robert Stambaugh, 1987, Expected Stock Returns and Volatility, *Journal of Financial Economics* 19, 3-30.

Glosten Lawrence, Ravi Jagannathan and David Runkle, 1993, On the Relation Between Expected Vale and Volatility of the Nominal Excess Returns on Stocks, *Journal of Finance* 48, 1779-1802.

Gomes, Francisco, 2002, Exploiting Short-Run Predictability, working paper, LBS.

Goyal, Amit and Ivo Welch, Predicting the Equity Premium With Dividend Ratios, working paper, Yale University.

Graham, John and Campbell Harvey, 1997, Grading the Performance of Market-Timing Newsletters, *Financial Analysts Journal, 54-66.*

Gron, Anne, Derrick Jorgenson and Nick Polson, 2000, Portfolio Choice and Stochastic Volatility, working paper, University of Chicago.

Hahn, Jaehoon and Hangyong Lee, 2001, On the Estimation of the Consumption-Wealth Ratio: Cointegrating Parameter Instability and its Implications for Stock Return Forecasting, working paper, Columbia University.

Harrison, Paul, and Harold Zhang, 1999, An Investigation of the Risk and Return Relation at Long Horizons, *The Review of Economics and Statistics*, 81(3), 399-408.

Harvey, Campbell, 2001, The Specification of Conditional Expectations, working paper, Duke University.

Hodrick, Robert, 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies,* 3, 357-386.

Jacquier, E., Polson, N. and Rossi, P. (1994), "Bayesian analysis of Stochastic Volatility Models", (with discussion). *J. Business and Economic Statistics.*, 12, 4.

Jacquier, E., Polson, N. and Rossi, P. (2001), Models and Priors for Multivariate Stochastic Volatility, forthcoming, *Journal of Econometrics*.

Jobson, J.D. and Robert Korkie, 1980, Estimation for Markowitz Efficient Portfolios, *Journal of the American Statistical Association* 75, 544-554.

Johannes, Michael and Nick Polson, 2002, MCMC methods in Financial Econometrics, forthcoming, *Handbook of Financial Econometrics.*

Kandel, Shmuel, and Robert Stambaugh, 1996, On the Predictability of Stock Returns: An Asset Allocation Perspective, *Journal of Finance* 51, 385-424.

Kim, T.S. and Edward Omberg, 1996, Dynamic Non-Myopic Portfolio Behavior, *Review of Financial Studies*, 9, 141-161.

Koopman, Siem and Eugenie Uspensky, 1999, The Stochastic Volatility in Mean model: Empirical Evidence from International Stock Markets, working paper.

Lamoreaux, Christopher and Guofu Zhou, 1996, Temporary Components of Stock Returns: What Do the Data Tells Us?, *The Review of Financial Studies*, 43, 1033-1059.

Leland, Hayne, 1999, Beyond Mean-Variance: Performance in a Nonsymmetrical World, *Financial Analysts Journal,* 27-35.

Liu, Jun, 1999, Portfolio Selection in Stochastic Environments, Working paper, UCLA.

Liu, Jun, Francis Longstaff and Jun Pan, 2001, Dynamic Asset Allocation with Event Risk, forthcoming, *Journal of Finance.*

Longstaff, Francis, 2001, Optimal Portfolio Choice and the Valuation of Illiquid Securities, forthcoming, *Review of Financial Studies.*

Lynch, A., 2001, Portfolio Choice and Equity Characteristics: Characterizing the Hedging Demands Induced by Return Predictability, *Journal of Financial Economics* 62, 67-130.

Merton, Robert, 1971, Optimum Consumption and Portfolio Rule in Continuous Time, *Journal of Economic Theory* 3, 373-413.

Merton, Robert, 1981, On Estimating the Expected Return on the Market: An Exploratory Investigation, *Journal of Financial Economics* 8, 323-361.

Pástor, Luboš and Robert Stambaugh, 2000, Comparing Asset Pricing Models: An Investment Perspective, *Journal of Financial Economics*, 56, 335–381.

Pástor, Luboš, 2000, Portfolio Selection and Asset Pricing Models, *Journal of Finance*, 55, 179–223.

Polson, Nick and Bernard Tew, 2000, Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500, 1970:1996, *Journal of Business of Economic Statistics,* 18, 164-174.

Polson, Nick, Jonathan Stroud and Peter Mueller (2002), Practical Filtering with Parameter Learning, working paper, University of Pennsylvania.

Stambaugh, Robert, 1999, Predictive Regressions, *Journal of Financial Economics*, 54, 375-421.

Wachter, Jessica, 1999, Optimal Consumption and Portfolio Allocation under Mean-Reverting Expected Returns: An Exact Solution for Complete Markets, forthcoming, *Journal of Financial and Quantitative Analysis*.

Whitelaw, F. Robert, 1994, Time Variation and Covariation in the Expectation and Volatility of Stock Market Returns, *Journal of Finance 49, 3-53*.

Whitelaw, Robert, 1997, Time-Varying Sharpe Ratios and Market Timing, working paper, NYU.

Xia, Yihong, 2000, Learning about Predictability: The Effect of Parameter Uncertainty on Dynamic Asset Allocation, forthcoming, *Journal of Finance*.

Appendix: MCMC algorithm and the complete Conditionals of ϕ and Σ

The likelihood at each time period is Gaussian, $p(r_t|\mu_t, V_t) = N(r_t \,|\, \mu_t, V_t)$, and the transition density at each time is a bivariate normal:

$$p(\mu_{t+1}, V_{t+1}|\mu_t, V_t, \phi, \Sigma) \;=\; \mathcal{N}\left(\left(\begin{array}{c} \alpha_\mu + \beta_\mu\, \mu_t \\ \alpha_v + \beta_v\, \log V_t \end{array}\right), \Sigma\right)$$

The full likelihood can be written as $p(r^T|\mu^T, V^T) = \prod_{t=1}^T p(r_t|\mu_t, V_t)$, and the distribution on the latent components as

$$p(\mu^T, V^T|\phi, \Sigma) = \prod_{t=1}^{T} p(\mu_t, V_t|\mu_{t-1}, V_{t-1}, \phi, \Sigma)$$

due to the Markov specification. To simplify notation let $\phi = (\alpha_\mu, \beta_\mu, \alpha_v, \beta_v)$ denote the "regression" parameters and let Σ denote the innovation covariance matrix:

$$\Sigma = \left(\begin{array}{cc} \sigma_\mu^2 & \rho\, \sigma_\mu \sigma_v \\ \rho\, \sigma_\mu \sigma_v & \sigma_v^2 \end{array}\right).$$

The joint probability model can be specified in three parts: the likelihood, $p(r^T|\mu^T, V^T)$, the distribution on the latent components, $p(\mu^T, V^T|\phi, \Sigma)$, and the prior on the parameters, $p(\phi, \Sigma) = p(\phi)\, p(\Sigma)$. To complete the model, we assume conjugate priors for the unknown parameters ϕ and Σ:

$$\begin{aligned} p(\phi) &= N(\phi \,|\, m, S) \\ p(\Sigma) &= IW(\Sigma \,|\, A, d), \end{aligned}$$

where $IW(\cdot|A, d)$ denotes an inverse-Wishart distribution with matrix parameter A and scalar degrees of freedom d (see O'Hagan, p. 293).

Our MCMC smoothing algorithm generates samples from the joint posterior distribution (suppressing superscripts):

$$p(\phi, \Sigma, \mu, V|r) \;\propto\; p(r|\mu, V)\; p(\mu, V|\phi, \Sigma)\; p(\phi)\; p(\Sigma)$$

by the sequentially drawing from the full conditional posterior distributions.

In this section, we derive the complete conditional for the parameters ϕ and Σ given vector of the state variables, (μ^T, V^T), and the observed data. Conditional on the state vectors, (μ^T, V^T), the parameters (ϕ, Σ) are independent of the observed returns, r. Hence, we need to compute $\mathrm{p}\big(\phi, \Sigma|\mu^T, V^T\big)$ which is proportional to $p\left(\mu^T, V^T|\phi, \Sigma\right) p\left(\phi, \Sigma\right)$. The transition density $p(\mu_t, V_t|\mu_{t-1}, V_{t-1}, \phi, \Sigma)$, for $t = 1, \ldots, T$, can be viewed as a multiple linear regression model:

$$z_t = X_t\, \phi + \epsilon_t,$$
$$\epsilon_t \sim N(0, \Sigma),$$

where

$$z_t = \begin{pmatrix} \mu_t \\ \log V_t \end{pmatrix} \text{ and } X_t = \begin{pmatrix} 1 & \mu_{t-1} & 0 & 0 \\ 0 & 0 & 1 & \log V_{t-1} \end{pmatrix}.$$

Following standard regression theory (see, for example, Bernardo and Smith, 1995), the posterior distribution for $p(\phi|\mu, V, \Sigma)$ is $N(m^*, S^*)$, where

$$S^* = \left(\sum_{t=0}^{T} S_t^{-1}\right)^{-1} \text{ and } m^* = S^* \left(\sum_{t=0}^{T} S_t^{-1} m_t\right)$$

with $S_t = (X_t' \Sigma^{-1} X_t)^{-1}$ and $m_t = (X_t' \Sigma^{-1} X_t)^{-1} X_t' \Sigma^{-1} z_t$, for $t = 1, \ldots, T$ where the prior distribution for ϕ is $N(m_0, S_0)$. The prior distribution for Σ is

$$p(\Sigma) \propto |\Sigma|^{-(d_0+p+1)/2} \exp\{-\frac{1}{2}\, tr\, (\Sigma^{-1}\, A_0)\}$$

which implies that the to a conditional posterior distribution for Σ, $p\left(\Sigma|\phi, \mu, V\right)$ is $\mathrm{I}W(A^*, d^*)$, where

$$A^* = \sum_{t=0}^{T} A_t \text{ and } d^* = d_0 + 2T$$

with $A_t = (z_t - X_t\phi)(z_t - X_t\phi)'$.