Optimal Estimation of the Risk Premium for the Long Run and Asset Allocation: A Case of Compounded Estimation Risk

Éric Jacquier
HEC Montréal, CIRANO, CIREQ

Alex Kane
University of California San Diego

Alan J. Marcus
Boston College

First version: October 2002
This version: October 2004

ABSTRACT:

It is well known that an unbiased forecast of the final value of a portfolio requires compounding at the arithmetic mean return over the investment horizon. However, the maximum likelihood practice, common with academics, of compounding at the *estimated* mean return, produces upward biased and highly inefficient estimates of long-term expected returns. We derive analytically two estimators of long-term expected returns for given sample sizes and horizons, one unbiased and one efficient in small samples. Both entail penalties that reduce the annual compounding rate as the investment horizon increases. The unbiased estimator, which is far lower than the compounded arithmetic average, is still very inefficient, often more so than a simple geometric estimator known to practitioners. Our small-sample efficient estimator is even lower. These results compound the sobering evidence in recent work that the equity risk premium is lower than suggested by post-1926 data. Our methodology and results are robust to extensions such as predictable returns. We also confirm analytically that parameter uncertainty, properly incorporated, produces optimal asset allocations in stark contrast to conventional wisdom. Longer investment horizons require *lower*, not higher, allocations to risky assets.

KEYWORDS:

Risk premium, asset allocation, long-term returns, arithmetic mean, geometric mean, maximum likelihood, mean squared error, estimation risk, small sample.

The paper had benefited from comments from Warren Bailey, Bryan Campbell, Francis Diebold, René Garcia, Éric Renault, Sy Smidt, and seminar participants at HEC Montréal, University of Montréal, Cornell, Simon Fraser, and the Frankfurt CFS conference on new directions in risk management. We are especially grateful for the comments of two referees and the editors. Address correspondence to Eric Jacquier at eric.jacquier@hec.ca

Estimates of expected long-term returns, or the risk-premium, are crucial inputs in empirical asset pricing and especially portfolio theory. First, it is important to know what wealth a portfolio is expected to generate over the long term, for example, in the context of pension funds and retirement policy. Second, this forecast is an implicit input to the asset allocation decision, the optimal mix of risky and risk-free assets. Portfolio theory is most always derived under the assumption that the agent knows the parameters of the return distribution. While sophisticated models of time-varying opportunity sets are well developed, the effect of parameter uncertainty on the forecast and the optimal investment decision are not as often discussed.

Recent reconsiderations of the equity risk premium have put the issue of parameter estimation back on the front burner of academic research. Fama and French (2002) or Jagannathan et al. (2000) make the case that the risk premium, i.e., the mean return in excess of the risk-free rate, is less than implied by post-1926 average returns. In addition, more inclusive databases also result in lower historical risk premia. For example, Dimson et al. (2002) and Jacquier et al. (2003) show that including returns from pre-1926 periods reduces historical average returns. Our results compound this sobering evidence. We show that a downward penalty for estimation risk should be applied to the per-period estimate of the risk premium when it is used for long-term forecasts. This argument is absent in the current literature on the declining equity premium and its implications for long-term investment.

At the core of the discussion is the magnitude of the estimation error in mean returns even with long samples. Of course, given the true parameters of the asset-return distribution, an unbiased forecast of the terminal value of a portfolio obtains by compounding the initial value at the arithmetic mean rate of return over the investment horizon. Substituting a maximum likelihood, arithmetic average for this true mean is of course asymptotically efficient and often advocated, e.g., Campbell (2001). Yet, practitioners often prefer geometric averages to simulate future portfolio values, arguing that the arithmetic average produces upward-biased long-term forecasts.

This paper studies the effect of estimation error when this unbiased and efficient one-period maximum-likelihood estimator is used for multi-period forecasting. We show that the usual asymptotic arguments for the consistency and efficiency of this common practice, although maximum likelihood, do not apply. This is because the forecasting horizon is often sizable relative to the sample size. Hence, the small-sample bias and inefficiency due to Jensen's inequality and the estimation error of the mean, do not vanish. Asymptotics here require the ratio of the length of the estimation period to that of the forecasting horizon also to be large.

First, for the unbiased estimator, initially discussed by Blume (1974), we review the analytic derivation detailed in Jacquier et al. (2003).[1] Then we derive an analytical formula for a small-sample efficient estimator, based on the minimization of mean squared errors. This is an appropriate replacement for the maximum likelihood procedure, here the arithmetic estimator, when asymptotic conditions are not met. Both these new estimators of expected long-term returns compound initial value at a *weighted average* of the arithmetic and geometric average returns. The weights depend on the ratio of the estimation period to the investment horizon. The small-sample efficient estimator is always lower than the unbiased estimator, itself always smaller than the arithmetic estimator. Using the small-sample efficient estimator results in a considerable efficiency gain over the unbiased estimator. In fact, even the practitioner-favored geometric estimator, while biased, is more efficient than the unbiased estimator for large investment horizons.

It is important to recognize that simulation methods of future returns are not immune to these bias and inefficiency problems. Most simulation methods that use sample estimates as inputs to a scenario analysis only reproduce the problems which we describe analytically.

Our initial analysis focuses on the simple case of i.i.d. log-normal returns, so we verify the robustness of our methodology and results to generalizations of this simple data generating process. Namely, extending our optimal estimator to allow for predictable returns requires only a simple modification of the initial formula. We show that predictability has little effect on the compounding of estimation risk. This is because predictability modifies the data generating process of both past and future data in ways that partially cancel in the computation of the estimator. We also discuss robustness to heteroskedasticity and the estimation of the variance.

To illustrate the practical import of these results, we consider an application to a classic problem of optimal asset allocation. We show that, even with i.i.d. returns, the optimal allocation to the risky asset must be reduced compared to the known-parameter case to properly incorporate estimation risk; the longer the horizon, the greater the reduction. This is in stark contrast to conventional wisdom that recommends increasing the allocation for longer horizons. For realistic values of the inputs, we demonstrate that these effects are large and should be taken into account. Again, we provide easy-to-implement analytical results.

The next section reviews the biases of the geometric and arithmetic estimators of expected long-term returns, and the derivation of an unbiased estimator. Section 2 derives a small-sample efficient estimator which should be used instead of the maximum likelihood estimator. We show that the efficiency gain compared to the maximum likelihood and the unbiased estimators are considerable. Section 3 relaxes the i.i.d. assumption and discusses the robustness of the estimators proposed, providing specific results for autocorrelated returns. Section 4 discusses the implications of these results for long-term asset allocation. Section 5 concludes.

1 UNBIASED ESTIMATION OF LONG-TERM EXPECTED RETURNS

1.1 Biases of the arithmetic and geometric estimators

We now show that the two most common competing estimators of expected future long-term returns are both biased. Assume that the one-period return, R_t, of a stock portfolio is log-normally distributed. That is, the log-return $r_t = \ln(1 + R_t)$ is i.i.d. normal with mean μ, standard deviation σ.[2] Therefore, the multi-period log-return, a sum of one-period log-returns over a future investment horizon of H periods, is normal with mean $H\mu$ and variance $H\sigma^2$. For an investment of \$1, V_H, the future portfolio value in H periods, can be written as

$$V_H = \$1 \times \exp(\mu H + \sigma \sum_{i=1}^{H} \varepsilon_{t+i}), \qquad \text{where } \varepsilon_{t+i} \sim \text{i.i.d. N}(0,1) \tag{1}$$

By the properties of the log-normal distribution, the expected return over H periods is:

$$E(V_H) = e^{(\mu + \frac{1}{2}\sigma^2)H} = [1 + E(R)]^H \tag{2}$$

Equation (2) is the basis of the standard practice of forecasting portfolio value by compounding at the expected rate of return. For example, the Ibbotson Associates publication *Stocks, Bonds, Bills and Inflation Annual Yearbook* simulates future portfolio values using arithmetic averages of past returns. Campbell (2001), who addresses the impact of autocorrelation on long-term forecasts, recommends the substitution of the sample average $\overline{R}$ for $E(R)$ when returns are i.i.d.

However, the sample-average holding period return, $\overline{R}$, is an unbiased but of course noisy estimator of $E(R)$. Jensen's Inequality implies therefore that

$$E([1+\overline{R}]^H) > [1+E(\overline{R})]^H = [1+E(R)]^H = E(V_H). \tag{3}$$

Because of the estimation error in $\overline{R}$, the compounded sample-average return gives an upward-biased estimate of the expected future portfolio value. That the bias should vanish in large sample is probably at the root of the typical practice of substituting the maximum likelihood estimate of $E(R)$ into (2). We will see that this asymptotic intuition is misleading when H is sizeable compared to the sample size T. Blume (1974) first discusses this bias. However, because he assumes that returns are normal rather than log-normal, he does not obtain exact formulas for expected values or bias. Cooper (1996) analyzes the bias in the context of discount factors for capital budgeting purposes. He concludes that the arithmetic mean is usually nearly appropriate, even accounting for estimation error. However, because discount factors involve powers of the *reciprocal* of the rate of return, the biases he finds differ drastically from those considered here. In our case, Jacquier et al. (2003) provide an analytical expression for the bias and an unbiased estimator. We review them briefly.

Denote by $\hat{\mu}$, the sample average computed over T past periods of log-returns, i.e., the maximum likelihood estimator of μ.. It can be written as

$$\hat{\mu} = \frac{1}{T}\sum_{i=1}^{T}\ln(1+R_{-i}) = \frac{1}{T}\left(\mu T + \sigma\sum_{i=1}^{T}\varepsilon_{-i}\right).$$

As is well known, $\hat{\mu}$ is unbiased with standard deviation $\sigma/\sqrt{T}$. It is conveniently written as

$$\hat{\mu} = \mu + \omega\,\sigma/\sqrt{T}\,, \qquad where\ \ \omega \sim \mathrm{N}(0,1). \tag{4}$$

It is well known, e.g., Merton (1980), that for i.i.d. returns, the precision of the estimator of μ depends only on the calendar span of the historic sample period. That is, $\hat{\mu}$ cannot be made more precise by sampling the data more often, only by sampling for a longer time period. This is at the core of the imprecision of standard estimates of mean returns and market risk premia. See Fama and French (2002) and Jagannathan et al. (2000) for recent discussions. In contrast, the estimate of variance can be made arbitrarily precise by sampling more frequently within a given sample period. Therefore, as high-frequency returns data have been available for at least 40 years, the estimation error in σ is a second order effect. For our purposes, we will ignore the estimation error in σ.

A first estimator for $E(V_H)$, based on the arithmetic average return, replaces $E(R)$ in (2) by $\bar{R}$. Alternatively, one can insert $\hat{\mu}$ in (2), estimating $1+E(R)$ by $e^{(\hat{\mu} + \frac{1}{2}\sigma^2)}$. $(1 + \bar{R})^H$ and $e^{(\hat{\mu} + \frac{1}{2}\sigma^2)H}$, while not exactly equal in small sample, have the same probability limit. They are both Maximum Likelihood Estimators (MLE) of $E(V_H)$. Asymptotically, the MLE is invariant to transformation, so substituting $\hat{\mu}$ or $\bar{R}$ appropriately, results in a MLE estimator of $E(V_H)$. Because the sampling distribution of $(1 + \bar{R})^H$ is only asymptotically lognormal and requires simulations, we would prefer to use $e^{(\hat{\mu} + \frac{1}{2}\sigma^2)H}$ which has an analytical distribution. We must first convince ourselves, however, that the two estimators have similar properties for realistic values of T and H.

Insert Figure 1 here

Figure 1 plots the ratios of the sample means and standard deviations of the two estimators for investment horizons, H, from 1 to 40 years, historical sample size $T = 75$ years, $\mu = 0.1$, $\sigma = 0.2$. These values are typical of a scenario where μ is estimated from a long sample of equity returns. For example, the mean and standard deviation of the log-return on the S&P500 from 1926 to 2001 are .099 and .196. Figure 1 shows clearly that the two estimators are very similar even for long horizons. The sample means of the two estimators are within 0.5% of each other even for horizons up to 40 years. Their standard deviations are within 4% of each other. For the rest of the paper, we will refer to $A = e^{(\hat{\mu} + \frac{1}{2}\sigma^2)H}$ as the *arithmetic average estimator*.

The bias of the arithmetic average (MLE) estimator A follows from (4) . Rewrite A as:

$$A = e^{(\hat{\mu} + \frac{1}{2}\sigma^2)H} = e^{(\mu + \omega\sigma/\sqrt{T} + \frac{1}{2}\sigma^2)H} = e^{(\mu + \frac{1}{2}\sigma^2)H}\, e^{(\omega\sigma/\sqrt{T})H}$$

Substituting from equation (2), we obtain

$$E(A) = E(V_H)\; E[e^{\omega\sigma H/\sqrt{T}}] = E(V_H)\; e^{\frac{1}{2}\sigma^2 H^2/T} \qquad (5)$$

Hence, the arithmetic average estimator is always biased upward by a factor of $e^{\frac{1}{2}\sigma^2 H^2/T}$.

Note that $\hat{\mu}$ is the logarithm of the geometric mean return. The "geometric average" estimator, denoted G, compounds at the exponential of $\hat{\mu}$ rather than $\hat{\mu} + \frac{1}{2}\sigma^2$. Practitioners often advocate it as an alternative to the arithmetic average estimator A. The argument is often made as follows. *"As G = $(P_1 / P_{-T})^{1/T}$, compounding at G into the future is the only way to generate forecasts that match the growth rate observed in the past, e.g. from T periods ago until now."* We can write the expectation of G as:

$$\begin{aligned} E(G) &= E(e^{\hat{\mu}H}) = E[e^{(\mu + \omega\sigma/\sqrt{T})H}] = e^{\mu H + \frac{1}{2}\sigma^2 H^2/T} \\ &= E(V_H)\, e^{\frac{1}{2}\sigma^2(H/T - 1)H} \qquad (6) \end{aligned}$$

Therefore the intuition in the above argument is correct only in the razor's edge case for which $H = T$. Then the geometric average estimator results in an unbiased forecast of expected cumulative returns. In all other cases, G is biased. The investment horizon H is of course exogenously set. There is no reason to restrict the estimation period T to be equal to H even if this were feasible. This would amount to giving up precision in $\hat{\mu}$, just for the sake of removing bias. One clearly wants to use the largest T available no matter the investment horizon considered.

1.2 Unbiased Forecasts of Long-Term Expected Returns

For any T and H, one can design an unbiased estimator of $E(V_H)$. In fact, the direct inspection of equations (5) or (6) shows that compounding at a rate $\hat{\mu} + \frac{1}{2}\sigma^2(1 - H/T)$ removes the bias.

Formally, we consider estimators in the class

$$C = e^{(\hat{\mu} + \frac{1}{2}k\sigma^2)H}. \tag{7}$$

These estimators nest the geometric average estimator G ($k = 0$) and the arithmetic average estimator A ($k = 1$). They can be interpreted as using a compounding rate that is a linear combination of A and G with weight on A equal to k.[3] The unbiased estimator U, obtains by solving for the value k_U that sets $E(C)$ in (7) equal to $E(V_H)$. The result is

$$k_U = 1 - H / T. \tag{8}$$

The unbiased estimator, U, is always smaller than the arithmetic estimator A since k_U is less than 1. Only as the forecast horizon H becomes small relative to the estimation period, T, does the maximum likelihood estimator become unbiased. The geometric average estimator, G, is biased downward when k_U is positive, that is, when $H < T$. For investment horizons *longer* than the estimation period, i.e., $H > T$, G is biased upward, like A. G is unbiased only for $T = H$.

The bias **in *A*** can persist even as $\hat{\mu}$ converges to μ when the sample size T increases, as long as the ratio H/T remains sizeable. For the estimator of expected terminal wealth to converge to its true value, both T and T/H must be large.

2 SMALL-SAMPLE EFFICIENT ESTIMATION

Asymptotics in our case is in *T/H,* not in *T*. The previous section showed that for relevant values of T and H, the MLE is biased; large sample conditions are not met. One can then strongly suspect that it not efficient either in the relevant small-sample conditions. Rather than counting on asymptotic efficiency, i.e., asymptotic minimum mean-squared error, one needs to construct a small-sample-efficient estimator for the sample size and horizon considered.

Further, note that unbiasedness is not in itself an estimation goal. Rather, it is typically used if needed to reduce a universe of possible estimators to a manageable set. This can lead to inadmissible estimators, as noted for example by Stein (1956). Instead, estimators should be set to minimize a loss function, a measure of average distance to the true parameter. When the problem at hand does not supply a specific loss function, a natural candidate is the mean squared error (MSE), the expectation of the squared deviation of the estimator from the true parameter. MSE can be written as the sum of the variance and squared bias. If tractability is not an issue, there is no reason to impose unbiasedness when searching for a minimum mean squared error estimator.

A minimum MSE estimator of expected future wealth, again within the class of estimators C in (7), is easily derived. Recall that the class nests A, G, and U, for $k = 1$, 0 and $1 - H/T$. The MSE of an estimator in class C is:

$$MSE(C) = E[C - E(V_H)]^2 = E\left(e^{\hat{\mu}H + \frac{1}{2}k\sigma^2 H} - e^{\mu H + \frac{1}{2}\sigma^2 H}\right)^2$$

$$= E\left(e^{2\hat{\mu}H + k\sigma^2 H} - e^{2\mu H + \sigma^2 H} - 2\, e^{\hat{\mu}H + \frac{1}{2}k\sigma^2 H + \mu H + \frac{1}{2}\sigma^2 H}\right) \qquad (9)$$

Substituting $\hat{\mu} = \mu + \omega\,\sigma/\sqrt{T}$, from (4) and evaluating the expectation yields:

$$MSE(C) = e^{2\mu H + 2\sigma^2 H^2/T + k\sigma^2 H} + e^{2\mu H + \sigma^2 H} - 2\, e^{2\mu H + \frac{1}{2}\sigma^2 H^2/T + \frac{1}{2}k\sigma^2 H + \frac{1}{2}\sigma^2 H}$$

Minimizing this expression over k results in the minimum MSE estimator, denoted M, of $E(V_H)$:

$$k_M = 1 - 3\, H/T \qquad (10)$$

Hence, for realistic values of T and H, the two popular estimators, arithmetic and geometric, and the unbiased estimator sometimes proposed in the literature, are all sub-optimal in terms of mean squared error, the most commonly used risk function. Equation (10) also shows that the best estimator of expected cumulative returns is even lower than the unbiased estimator. Table 1 compares the estimators for relevant ranges of T and H.

-- Insert Table 1 here --

Do these different values of k lead to very different estimates of $E(V_H)$? To illustrate the long-term effect of these differences in compounding rates, Panel (a) of Figure 2 plots the various estimates of $E(V_H)$ versus the investment horizon H. The estimates of final wealth diverge dramatically across estimators for longer horizons. For $T = 75$, the arithmetic average A predicts that \$1 will compound to \$120 in 40 years, the unbiased estimator forecasts \$80, while the minimum MSE forecasts only \$25. When T is 30 years, U and especially M dramatically penalize predictions for investment horizons longer than T. The optimal estimator M penalizes long-horizon forecasts far more than U, as soon as H becomes sizeable relative to T.

Insert Figure 2 here

Panel (b) of Figure 2 plots the effective annual compounding rates of *A, G, U,* and M versus H, for $T = 75$ years (the length of the CRSP monthly data base) and 30 years (a shorter sample size more relevant for an emerging market). The two horizontal lines in Panel b, A ($e^{.1 + ½ \times .2^2}$ = 12.7%) and G ($e^{.1}$ = 10.2%), are unaffected by H, while U and M "penalize" the increasing H by linearly decreasing the compounding rate. The effects are large. Even for a lengthy sample period of $T = 75$ years, a 30-year investment horizon, broadly appropriate for a retirement fund, calls for the compounding rate to decline from the arithmetic 12.7%, to about 10% for M. Necessary corrections become dramatic with shorter sample periods. For $T = 30$, the appropriate compounding rate is about 8% even for a 20-year investment horizon. These results compound the sobering message in Fama-French (2002) and Jagannathan et al. (2000), that the per-period estimate of the equity premium is lower than once thought.

One may argue that the results for M are specific to the risk function chosen. While this is of course correct, the loss function chosen here has some wide appeal. Mainly, it is the small sample implementation of the efficiency criterion invoked to justify the maximum likelihood principle. Also, the same estimator is obtained whether one minimizes sampling loss in terms of dollars or returns. Estimators optimal with respect to alternative loss functions, e.g., asymmetric, can be derived in a similar fashion, see Jacquier (2004).

Because the predictions of the various estimators of final wealth differ so considerably, so must their predictive accuracies. Panel A of Figure 3 plots the root mean squared error of each estimator as a multiple of the value of expected final wealth, for $\mu = .1$, $\sigma = .2$, $T = 60$, as a function of H. By construction of course, M is most precise at all horizons, and therefore is the lowest curve. We first note that the precision of the MLE estimator, ***A***, largely favored in the academic community, is astonishingly poor. At a horizon of 40 years, its RMSE is nearly 2.5 times the true expected value of final wealth. Such a magnitude renders estimates nearly useless. The second poorest estimator in terms of precision is clearly *U*, the unbiased estimator. In fact, surprisingly enough, it is the geometric estimator *G*, which comes second best. The RMSE of *G* doesn't diverge much from that of the efficient estimator for horizons below 25 years. Even at a horizon of 40 years, the RMSE of *G* is only about 30% higher than that of *M*.

Insert Figure 3 here

Panel b of Figure 3 gives us further information on the relative precisions of these estimators. There, we plot the percentage improvement in RMSE of *G*, *U*, and *M* over *A* as a function of H/T. Again, by construction, the efficient estimator *M* establishes an upper bound on the improvement in RMSE. At very short horizons, *A*, *M*, and *U* are all virtually identical with values of $k \approx 1$ [see equations (8) and (10)], so the improvement over *A* is negligible. As the horizon extends, the improvement of *M* becomes dramatic, surpassing 60 percent at a 40-year horizon. The geometric estimator does its best at mid-range horizons. Its curve achieves tangency with that of *M* at $H/T = 1/3$. This follows from equation (10), which shows that at this point, $k_M = 0$, making *G* achieve efficiency. Close to this point, on either side of it, *G* is close to efficient. It is noteworthy that as soon as H/T is greater than about .2, the unbiased estimator *U* has a substantially higher RMSE than *G*. In fact, the geometric estimator is less precise than *A* and *U*, with a negative "improvement" in RMSE, only at shorter horizons. This is because, while the estimators *M*, *A*, and *U* converge for very low values of H/T, *G* remains downward biased. Note that the crossing of the curves for *U* and *G* (around $H/T = 0.2$) occurs where *G* compensates for its squared bias (*U* has no bias) with an equal reduction in variance. For larger H/T's, the variance of *U* is just too large to make its unbiasedness an interesting feature.

To summarize, the catastrophic lack of precision of *A*, the relatively disappointing imprecision of *U*, and strong performance of the geometric estimator in the middle range of investment horizons are the striking features of Figure 3.

3 ROBUSTNESS

To produce clear analytical results, we have made so far a number of simplifying assumptions. This section shows that the results are robust to these assumptions.

A large literature beginning with Summers (1986), Poterba and Summers (1988), and Fama and French (1988) has addressed the autocorrelation of long-term returns. Estimates point to negative autocorrelations at lags in business cycle range. While the strength of the evidence is somewhat disputed, our analysis easily adjusts to autocorrelated returns. Autocorrelation enters the analysis through the variance of two sums of returns, namely H future returns for the forecast and T past returns for the estimate of μ. Given an autocorrelation structure, we introduce the correlation matrix $\boldsymbol{C}$ and vectors of ones, $\boldsymbol{i}$, of dimensions T or H as appropriate. The variance of a sum of log-returns is then $\sigma^2 \boldsymbol{i}'\boldsymbol{C}\boldsymbol{i}$, instead of $T\sigma^2$ or $H\sigma^2$. This affects $E(V_H)$ in equation (2), where the exponential term becomes $H(\mu + ½\ \boldsymbol{i}'\boldsymbol{C}_H\boldsymbol{i}\ \sigma^2/H)$. Similarly, $\hat{\mu}$ in equation (4) becomes $\mu + \omega\ \sigma \sqrt{\boldsymbol{i}'\boldsymbol{C}_T\boldsymbol{i}/T}$. The analysis then follows. Defining $F_T = \boldsymbol{i}'\boldsymbol{C}_T\boldsymbol{i}/T$ and $F_H = \boldsymbol{i}'\boldsymbol{C}_H\boldsymbol{i}/H$, the unbiased and minimum mean-squared error estimators require respectively

$$k_U = 1 - \frac{H}{T} \times \frac{F_T}{F_H} \tag{11}$$

$$k_M = 1 - \frac{3H}{T} \times \frac{F_T}{F_H} \tag{12}$$

The correction is similar for other non-i.i.d. specifications. The ratio F_T/F_H more generally can be replaced with the ratio of the variance of the sum of the T returns from the estimation period divided by T, to the variance of the sum of the H returns from the forecast period divided by H. It would be straightforward to generalize this approach for other ARMA processes. These corrections require little modification of our basic formulas. One suspects that, for most stable forms of predictability, the new ratios in (11) and (12) may not be far from 1.

Insert Figure 4 here

Figure 4 plots the unbiased and efficient estimators with and without the corrections for autocorrelation. We estimated an MA(4) process on the S&P 500 annual log-returns from 1926-2001 and computed k_U and k_M as per equations (11) and (12). Figure 4 shows that the correction is really a second-order effect given the autocorrelation in the data. Forecasts are barely affected by the correction. Other realistic long-term autocorrelations produced the same results.

For similar reasons, heteroskedasticity poses little problems for these estimators. Heteroskedasticity enters our computation through the variance of the sum of T or H returns. For long investment horizons, this is essentially captured by the unconditional variance, provided that variance is stationary. Empirical estimates of heteroskedasticity such as stochastic volatility models imply that conditionality in variance essentially vanishes for forecasting horizons beyond a few years. Note however, that a given form of heteroskedasticity in annual returns would warrant a modification of the standard estimator of μ, as well as the variance of the sum of the next H returns if H is small. Evidence of heteroskedasticity in annual returns is however weak.

Our results can also be used under alternative estimation procedures. For example, Fama and French (2002) note that use of the dividend discount model may provide more precise estimates of the market risk premium than historical averages. This efficiency gain is easily incorporated in our model. Simply convert the variance reduction in the estimation into an equivalent increase in estimation period, T. Namely, for a more efficient estimator, a reduction in sampling error variance by a factor of E is equivalent to an increase in notional sample size by a factor of E. This is also an intuition for the correction for autocorrelation in equations (11) and (12), which effectively adjusts H/T for the ratio of "average variance" of the returns in the estimation period to that of the forecasting horizon.

One may also worry that σ is in fact unknown and also estimated. Estimation error in σ introduces non-normality in the predictive distribution of log-returns (see for example Bawa et al., 1979). The variance of the predictive distribution is inflated by a factor of $\nu/(\nu-2)$, where ν is the sample size used to estimate the variance. Again, as pointed out before, the sampling distribution of $\hat{\sigma}$ converges to σ when the sampling frequency increases. One will therefore benefit from the use of higher frequency returns for the purpose of estimating σ. The inflation factor will then be extremely close to 1, and the induced non-normality negligible.

4 OPTIMAL FORECAST VS. OPTIMAL ALLOCATION

How important are these issues to economic decisions? We consider in this section an application to a classic problem in finance: a portfolio allocation between a risk-free asset and a risky asset. Samuelson (1969) or Merton (1969) show that for investors with power utility functions, the optimal allocation to the risky portfolio is $w^* = \dfrac{\alpha - r_0}{\gamma\sigma^2}$ where $\alpha \equiv \mu + ½\,\sigma^2$ is the expected rate of return on the risky asset, r_0 denotes the risk-free return, and γ is the investor's measure of relative risk aversion. The optimal allocation in these models is independent of time horizon, and the portfolio is rebalanced continuously to maintain the optimal weights.

To re-visit this problem under parameter uncertainty, consider an investor with power utility function who maximizes the expectation of utility of final wealth given by:

$$U(V_H) = \frac{V_H^{1-\gamma}}{1-\gamma} = \frac{1}{1-\gamma}\exp[(1-\gamma)\ln(V_H)] \tag{13}$$

Given a capital allocation of w to the risky portfolio and $(1-w)$ to the risk free asset, where the weight w is maintained constant through continuous rebalancing, the portfolio value at the horizon date is log-normal with parameters:[4]

$$\ln(V_H) \sim \mathrm{N}(\mu_H, \sigma_H^2) \equiv N[(r_0 + w(\alpha - r_0) - ½\,w^2\sigma^2 H,\ Hw^2\sigma^2] \tag{14}$$

Log-normality with these moments of $\ln(V_H)$ implies that expected utility is:

$$E[U(V_H)] = \frac{1}{1-\gamma}\exp\{(1-\gamma)H[r_0 + w(\alpha - r_0) - ½\,w^2\sigma^2 + ½(1-\gamma)w^2\sigma^2]\} \tag{15}$$

Maximizing (15) with respect to w yields the well-known optimal allocation $w^* = \frac{\alpha - r_0}{\gamma\sigma^2}$. For i.i.d. returns, the result is independent of the horizon, an implication extensively discussed in the literature.

In contrast, conventional advice is to increase the allocation with the horizon. This advice is often motivated in the literature by allowing predictability in expected returns, e.g., Campbell and Viceira (1999), Detemple et al. (2003), Wachter (2002) and others, or by invoking non-portfolio sources of income such as labor as in Bodie, Merton, and Samuelson, (1992).

The Samuelson-Merton result assumes knowledge of the parameters of the return distribution. Bawa et al. (1979) discuss a "variance inflation" effect due to estimation error on asset allocation in a one-period framework. The effect is not very dramatic for a single period. We now show that it is far greater when the ratio H/T is non-trivial. Barberis (2000) discusses it, but focuses on the interaction of asset allocation with learning, whereas we are more concerned with implications of estimation uncertainty and alternatives to unbiased estimators of expected return, even if no learning or predictability occurs. We now show analytically that the interaction between estimation error and (long) forecasting horizon is very important.

From a decision theoretic perspective, the mere substitution of a point estimate in the optimal allocation in place of the unknown α is incorrect. Rather, the investor, after estimating μ (hence α) has a distribution that represents its uncertainty. This may be a sampling distribution or, for a Bayesian econometrician, the posterior distribution. Therefore, $E[U(V_H)]$ in (15) is random, as a non-linear function of a random variable α. Following decision theory, the proper expected utility to maximize is that resulting from first integrating α out of equation (15); see Bawa, Brown, and Klein (1979) and Berger (1985) for Bayesian analysis and decision theory, and also Chamberlain (2000) for a recent study. This integration produces the expected utility of wealth given the *data,* $E[U(V_H \mid D)]$, to be optimized by the investor. Specifically:

$$E[U(V_H) \mid D] = \int E[U(V_H)/\alpha] \; p(\alpha \mid D) \, d\alpha \tag{16}$$

Given a sample size of length T, and no prior information, the posterior distribution of α is simply $N(\hat{\alpha}, \sigma^2/T)$. The result of the integration in (16) is

$$E[U(V_H \mid D)] = \frac{1}{1-\gamma} \exp\{(1-\gamma)H[r_0 + w(\hat{\alpha} - r_0) - ½\, w^2\sigma^2 + ½(1-\gamma)w^2\sigma^2(1+\frac{H}{T})]\} \quad (17)$$

While α in (15) is replaced with $\hat{\alpha}$, the last term in (17) is new. It reflects the variance inflation due to the estimation of α. The maximization of (17) yields the optimal asset allocation:

$$w^* = \frac{\hat{\alpha} - r_0}{\sigma^2[\gamma + \frac{H}{T}(\gamma - 1)]} \quad (18)$$

For $H<<T$, equation (18) collapses to the standard optimal allocation. Otherwise, for $\gamma >$ 1, the allocation to the risky asset is decreased (relative to the known-parameter case) in favor of a higher allocation to the risk-free asset. Recognition of the uncertainty in the estimate of α leads investors to shy away from risky assets, the more so the greater the ratio H/T. First, this result is precisely contrary to the common advice to invest more in stocks for longer horizons. Second, it happens even if returns are unpredictable. The only exception is log-utility investors, for whom γ = 1. Because log-utility is linear in α, estimation uncertainty and the horizon H do not affect the location of the optimum.

Note that the problem can easily be written where the investor has a proper prior. The optimal allocation in (18) will then involve the posterior mean rather than $\hat{\alpha}$, and a modified sample size accounting for the prior precision rather than T. One could then deduce from that optimal allocation what type of prior could be consistent with ignoring parameter uncertainty. It will have the feature that the prior mean increases, above the sample mean, with the forecasting horizon.

As in Barberis (2000) and Brennan (1998), our rationale on long-term asset allocation turns on an often-overlooked wrinkle in the argument concerning the proposition that stocks are less risky as long-run investments. The result here is driven by the fact that the impact of estimation uncertainty compounds as the investment horizon becomes more distant. Rebalancing does not eliminate this effect. This consideration might well dominate mild negative serial correlation as one evaluates the risk of stocks at different investment horizons. And, while the strength of serial correlation in market returns is still contested, there is no doubt about the considerable estimation error surrounding estimates of mean return.

Insert Figure 5 here

Figure 5 plots w^* as a function of H for the parameters used in Figure 2, namely $\hat{\mu} = .1$, $\sigma = .2$. We also use $r_0 = .04$, and $\gamma = 2$ and 4. The lines labeled "conventional" show the asset allocation in the known-parameter case. The downward sloping lines depict the optimal allocation to the risky portfolio when parameter uncertainty is properly incorporated. Figure 5 shows that estimation uncertainty has a dramatic interaction with the investment horizon. The optimal asset allocation may differ considerably for different investment horizons. For a long estimation period of 75 years, the optimal weight on the risky asset falls from 87%, under the "conventional" approach to 70% as the investment horizon expands from 1 to 40 years. The effect is far more pronounced when the estimation period is shorter, as it would be for an emerging market or if one believed that the U.S. was subject to structural breaks. For $T = 30$ years, the allocation falls to 53%.

More risk-averse investors have lower "conventional" allocation to the risky asset. Even for them, the effect of horizon is very strong as a fraction of the conventional allocation. For an investor with $\gamma = 4$, the 40-year conventional allocation is 43% while the optimal allocation is for T=30 is only 20%.

This example is admittedly only indicative in that we assume the choice of w^* is a once-and-for-all decision. A full-blown asset allocation problem [see Brennan (1998) and Barberis (2000)] would allow for dynamic updating of the estimate of μ and the allocation w^* as agents learn about the true distribution of returns from new data, and would give rise to intertemporal hedging demands against changes in the perceived opportunity set along the lines of Merton (1973). These issues would take us far afield from the forecasting problem, however.[5] Further, the evidence on the predictability of asset returns is far from uncontroversial. While we do not explicitly account for time-varying hedging demands, (18) does provide an analytical guideline to the impact of estimation uncertainty even in their presence. One may take (18) as indicative of the importance of estimation risk for simple asset allocation even in the presence of rebalancing.

5 CONCLUSION

We have considered the problem of forecasting portfolio values over long horizons when the return distribution is estimated from historic data. While recent papers address the estimation of the one-period expected return, few have focused on the formulation of long-term expected returns. Moreover, the literature focuses most exclusively on the arithmetic (maximum likelihood) and geometric estimators, rarely on unbiased estimators. It has so far ignored naturally efficient estimators such as those minimizing mean squared error (MSE), which can be seen as the small sample generalization of the principles justifying the maximum likelihood. We derive an analytic small-sample efficient estimator of long term expected returns. It is far lower than the MLE, the unbiased, and even the geometric average estimator when H is larger than $T/3$. The resulting efficiency gains are spectacular, not only on the MLE but also on the unbiased estimator for long horizons, and on the geometric estimator for the shorter horizons.

We show how the results are easily adjusted for serial correlation. However, realistic values of autocorrelation appear unimportant for the long-horizon forecasts that are the concern of this paper.

Strong cases are made in recent studies that the estimate of the market risk premium should be revised downward. Our result compounds this by stating that even these lower estimates of mean return should be adjusted further downward when used to predict long-term returns. Our results also show that alternative methods of estimation of the risk premium that can be shown to be more precise, e.g., Fama and French (2002), are especially valuable if the premium is to be used for long term forecasts.

We also analytically derive a striking implication for long-term asset allocation. Contrary to conventional wisdom, longer investment horizons imply *lower* allocations to risky assets in order to account for the fact that the estimation error gets compounded at the investor’s horizon.

REFERENCES

Barberis, Nicholas, 2000. “Investing in the Long Run when Returns are Predictable,” *Journal of Finance* 55 (February): 225-264.

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

Berger, James. (1985). *Statistical Decision Theory and Bayesian Analysis*, New York: Springer Verlag.

Blume, Marshall. (1974). “Unbiased Estimators of Long-run Expected Rates of Returns", *Journal of the American Statistical Association* 69 (Sept): 634-63.

Bodie, Zvie, Robert Merton, and William Samuelson. (1992). “Labor Supply Flexibility and Portfolio Choice in a Life Cycle Model,” *Journal of Economic Dynamics and Control* 16: 427-440.

Brennan, Michael. (1998). “The Role of Learning in Dynamic Portfolio Decisions.” *European Economic Review* 1, 295-306.

Campbell, John. (2001). "Forecasting U.S. Equity Returns in the 21st Century." Presentation to the Social Security Advisory Board, Washington, DC (July).

Campbell, John, and Luis Viceira. (1999). “Consumption and Portfolio Decisions when Expected Returns are Time Varying.” *Quarterly Journal of Economics* 114: 433-495.

Chamberlain, Gary (2000).”Econometrics and Decision Theory.” *Journal of Econometrics* 95(2): 255-283.

Cooper, Ian. (1996). “Arithmetic versus Geometric Mean Estimators: Setting Discount Rates for Capital Budgeting.” *European Financial Management* 2 (July): 157-167.

Dimson, Elroy, Paul Marsh, and Michael Staunton. (2002). *Triumph of the Optimists: 101 Years of Global Investment Returns*. Princeton, NJ: Princeton University Press.

Fama Eugene, and Kenneth French. (1988). “Permanent and Temporary Components of Stock Prices.” *Journal of Political Economy* 96: 246-273.

Fama Eugene, and Kenneth French. (2002). “The Equity Premium.” *Journal of Finance* 57 (April): 637-660.

Garcia, René, Jérome Detemple, and Marcel Rindisbacher. (2003). "A Monte Carlo Method for Optimal Portfolios." *Journal of Finance*, (February), 58:1, 401-446.

Jacquier, Éric. (2004). “Estimates of Long Term Expected Returns: Economic vs. Statistical Rationales”. Working Paper, HEC Montréal.

Jacquier, Éric, Alex Kane, and Alan Marcus. (2003). “Geometric or Arithmetic Mean : a Reconsideration.” Financial Analysts Journal 59, Nov/Dec.

Jagannathan, Ravi, Ellen McGrattan, and Anna Scherbina. (2000). “The Declining U.S. Equity Premium.” *Federal Reserve Bank of Minneapolis Quarterly Review* 24 (Fall): 3-19.

Merton, Robert. (1969). “Lifetime Portfolio Selection Under Uncertainty: The Continuous-time Case.” *Review of Economics and Statistics*.” 51: 247-257.

Merton, Robert. (1973). “An Intertemporal Capital Asset Pricing Model.” *Econometrica* 41 (September): 867-887.

Merton, Robert. (1980). “On Estimating the Expected Return on the Market: An Exploratory Investigation.” *Journal of Financial Economics* 8 (December): 323-406.

Poterba, James, and Larry Summers. (1988). “Mean Reversion in Stock Prices: Evidence and Implications.” *Journal of Financial Economics* 22: 27-60.

Roll, Richard. (1983). “On Computing Mean Returns and the Small Firm Premium.” *Journal of Financial Economics* (November) 12: 371-386.

Samuelson, Paul. (1969). “Lifetime Portfolio Selection by Dynamic Stochastic Programming,” *Review of Economics and Statistics*.” 51: 239-246.

Stein, Charles. (1956). “Inadmissibility of the Usual Estimator of the Mean of a Multivariate Distribution.” in J. Neyman, ed., Proc. 3rd Berkeley *Symp. Math. Stat. Prob*., vol. 1 University of California Press.

Summers, Larry. (1986). “Does the Stock Market Rationally Reflect Fundamental Values?” *Journal of Finance* 41(3): 591-602.

Wachter, Jessica. (2002). “Portfolio and Consumption Decisions under Mean-Reverting Returns: an Exact Solution for Complete Markets.” *Journal of Financial and Quantitative Analysis* 37, 63-91.

Table 1

Properties of alternative estimators of cumulative portfolio return.

All estimators are of the general form $e^{(\hat{\mu} + ½k\sigma^2)H}$. $\hat{\mu}$ is the maximum likelihood estimator of the mean log-return from a sample of length T. The investment horizon is H years. The values for k are as follows:

Arithmetic estimator (A):	$k = 1$
Geometric estimator (G):	$k = 0$
Unbiased estimator (U):	$k_U = 1 - H/T$
Minimum MSE estimator (M):	$k_M = 1 - 3\,H/T$

Condition		Ordering	Bias < 0	Bias > 0
$T >> H$	$k_M \approx k_U \approx 1$	$M \approx U \approx A$	G	
$T > 3H$	$0 < k_M < k_U < 1$	$G < M < U < A$	G, M	A
$3H > T > H$	$k_M < 0 < k_U < 1$	$M < G < U < A$	M, G	A
$H > T$	$k_M < k_U < 0$	$M < U < G < A$	M	G, A

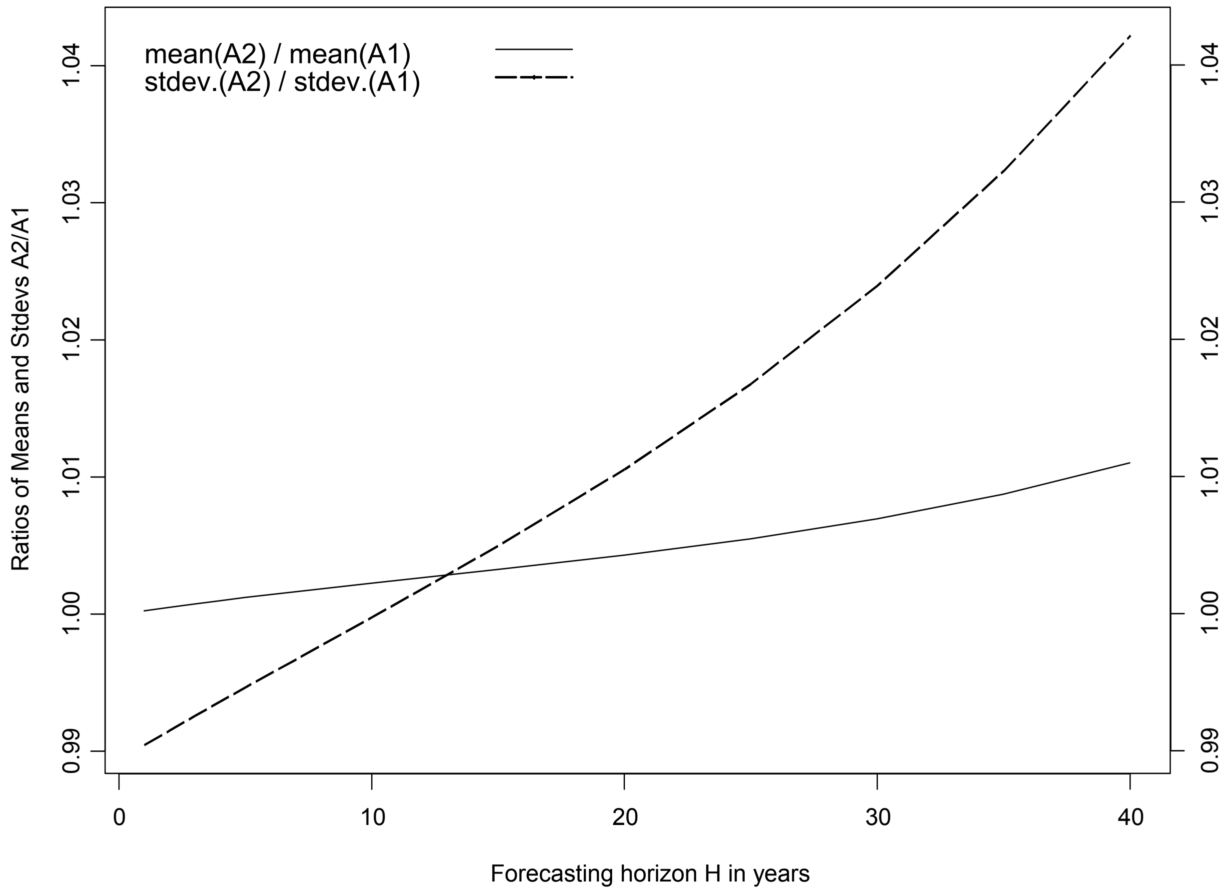

Figure 1: Ratios of sampling means and standard deviations of estimators $A_1 = (1 + \overline{R})^H$ and $A_2 = e^{(\hat{\mu} + \frac{1}{2}\sigma^2)H}$, $\mu = 0.1$, $\sigma = 0.2$, $T = 75$. The mean and standard deviation of A_1 are computed from 2 million draws of the asymptotic distribution of $\overline{R}$.

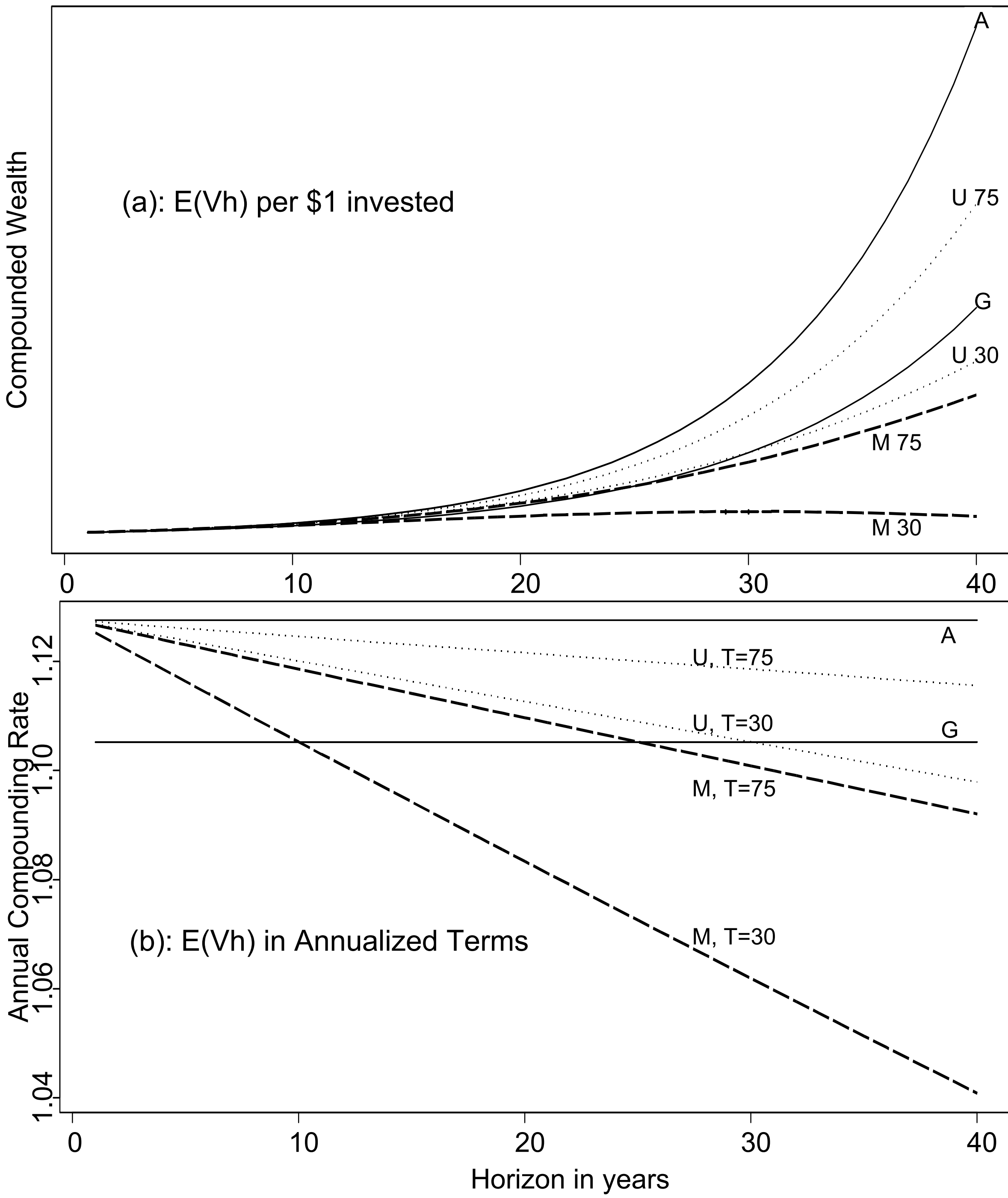

Figure 2: Estimators of Long-Term Expected Returns: *A, G, U, M*. $\hat{\mu} = 0.1$, $\sigma = 0.2$.

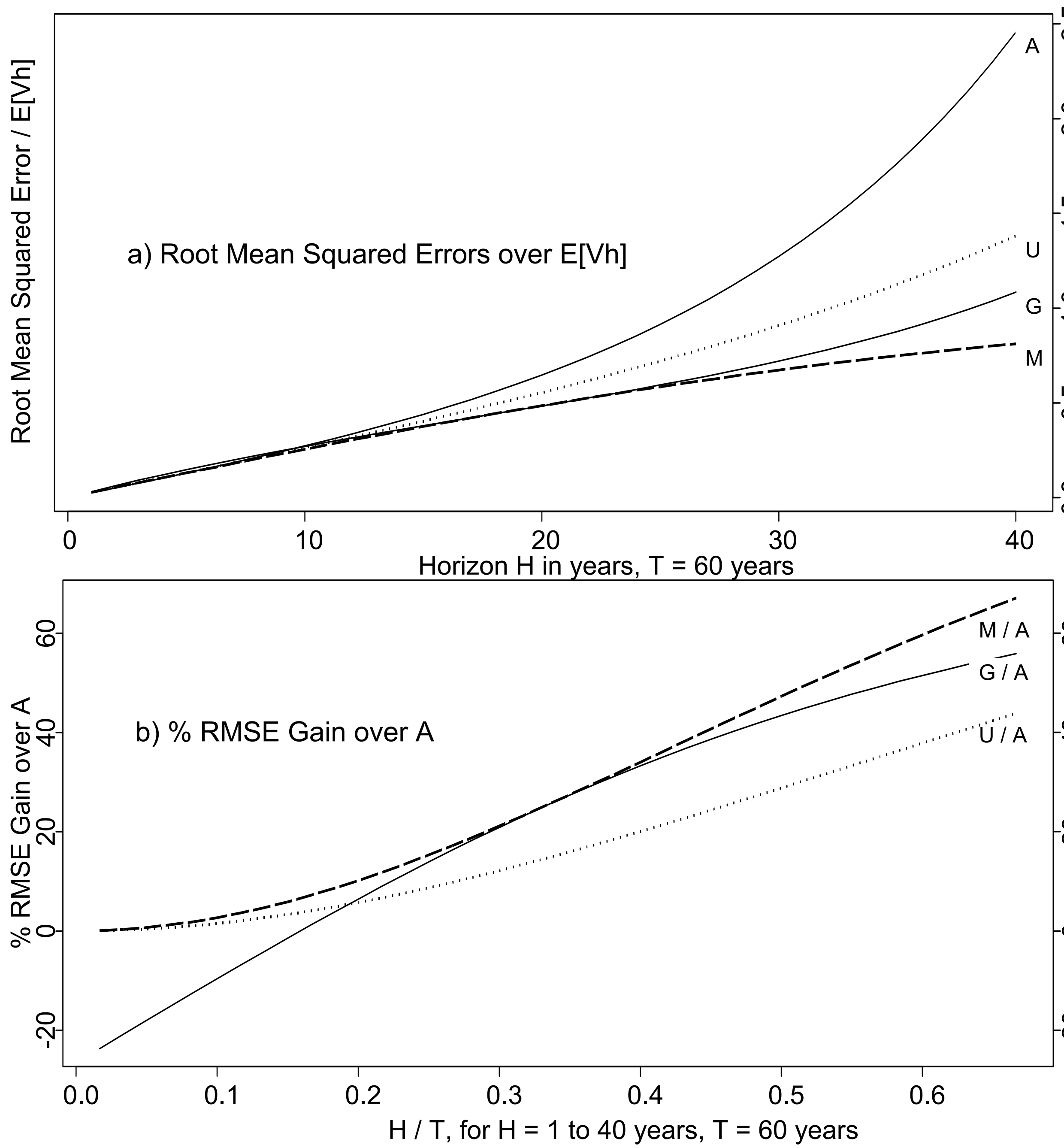

Figure 3: Root Mean Squared Errors of Estimators *A, G, U, M*. $\hat{\mu} = 0.1$, $\sigma = 0.2$, T=60 years.

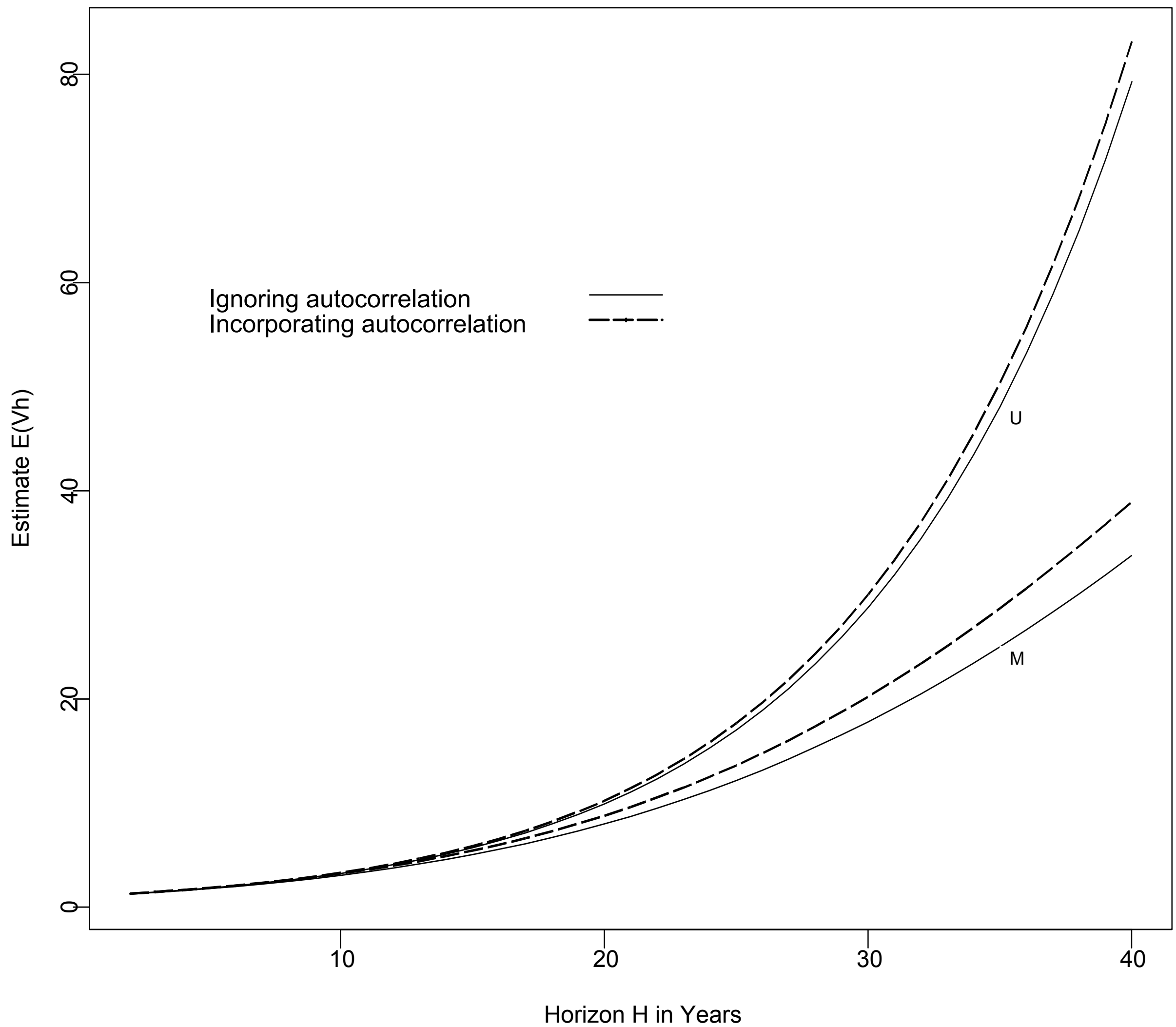

Figure 4: Effect of autocorrelation on estimators *U* and *M*, $\hat{\mu} = 0.1$, $\sigma = 0.2$, $T = 75$, MA(4) on annual S&P returns: $\theta = (-0.16, -0.02, -0.16, -0.08)$ estimated on 1926-2001

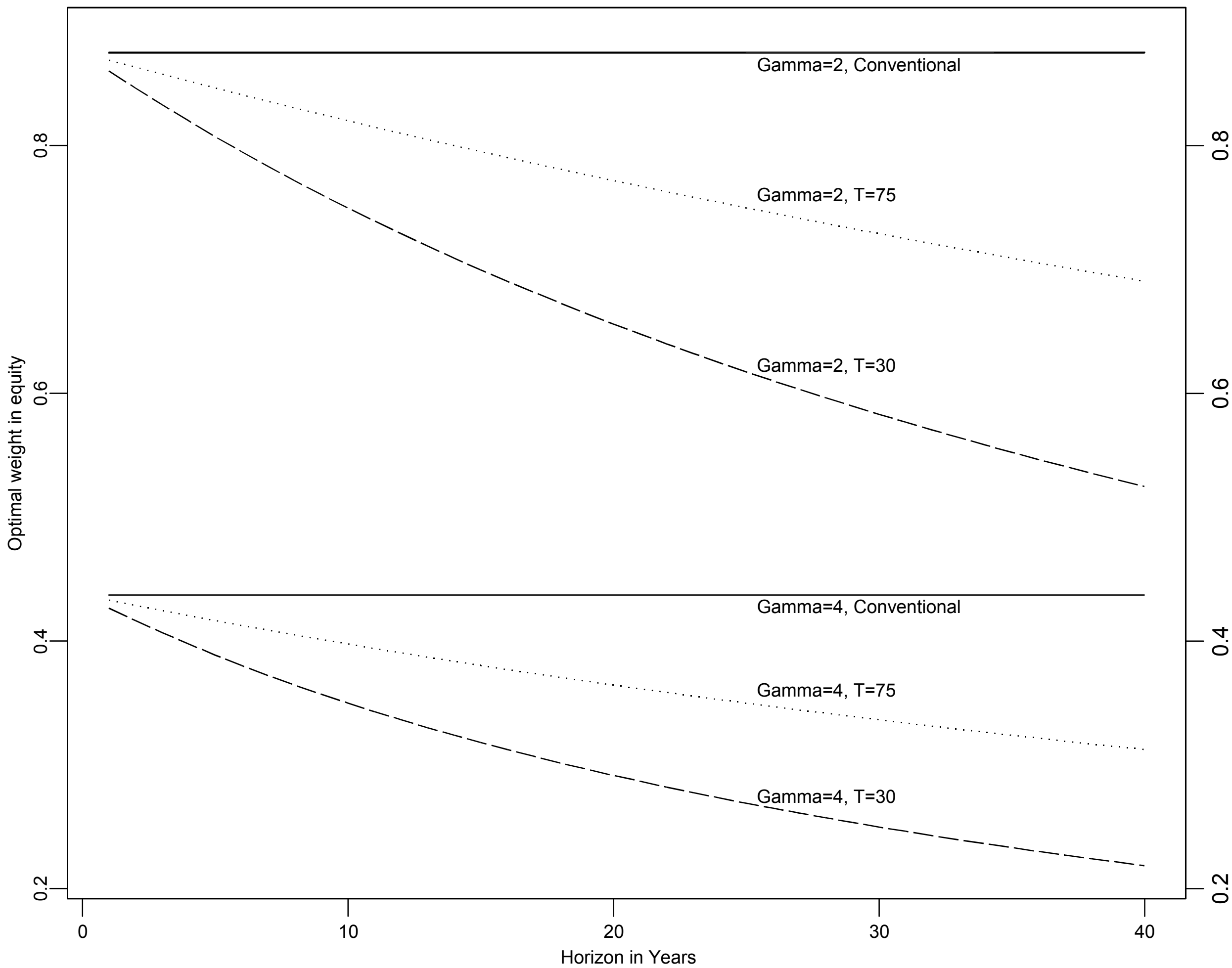

Figure 5: Combined effects of horizon and estimation error on optimal allocation, $\hat{\mu} = 0.1$, $\sigma = 0.2$

[1] See also Roll (1983) who discusses how, due to Jensen's inequality, different estimates of portfolio mean returns have different biases, e.g., first compound and then cross-sectionally average or vice-versa.

[2] Autocorrelation in returns modifies the variance of the long-term returns but not the spirit of the following discussion. Section 3 discusses the effect of autocorrelation and heteroskedasticity.

[3] Considering the seemingly larger class $e^{(k_1\hat{\mu} + ½\, k_2\sigma^2)H}$ adds no generality. One can show that any value $k_1 \neq 1$ leads to *infeasible* estimators, i.e., functions of the true parameter. We restrict our analysis to feasible estimators, i.e., functions solely of sample statistics. It is also easily verified that feasible estimators in the class, $w_1\, A + w_2\, G$, require $w_2 = 0$ and map one to one with those in equation (7).

[4] This can be shown formally using, for example, Ito's lemma. However, it is easier to note that under continuous rebalancing, i.e., with fixed portfolio weights, the instantaneous portfolio return is log-normally distributed with constant mean and variance, which implies that the full-period return remains log normal with the same parameters per unit time.

[5] Barberis (2000) presents optimal asset allocations both for buy-and-hold investors who never rebalance, and for investors who update parameter estimates as more data become available and periodically rebalance optimally. Our case is somewhere between these extremes in that our investors rebalance but do not update. The greater simplicity of this setting allows us to maintain the focus on forecasting issues as well as to derive a closed-form solution for asset allocation that analytically demonstrates the importance of the key ratio H/T. Symmetrically, Barberis (2000) presents detailed analysis of optimal asset allocation using numerical simulations, but, in contrast to our focus, devotes little attention to properties of alternative estimators of expected return.