

# Uncertainty in Second Moments: Implications for Portfolio Allocation

David Daewhan Cho\*

*SUNY at Buffalo, School of Management*

September 2003

## Abstract

This paper investigates the uncertainty in variance and covariance of asset returns. It is commonly believed that these second moments can be estimated very accurately. However, time varying volatility and nonnormality of asset returns can lead to imprecise variance estimates. Using CRSP value weighted monthly returns from 1926 to 2001, this paper shows that the variance is less accurately estimated than the expected return. In addition, a mean variance investor will incur significant certainty equivalent loss due to the uncertainty in second moments. Applying the Fama French 3 factor model to 25 size, BE/ME sorted portfolios from 1963 to 2001, the loss due to the variance estimation can be shown to be as large as the loss due to the expected return estimation. Moreover, as the number of assets in the portfolio increases, the loss due to the variance uncertainty becomes larger. This provides a possible explanation to the home bias puzzle.

*JEL classifications:* C22; D81; G11

*Keywords:* *Portfolio allocation; Estimation risk; Time varying volatility; Home bias; Bayesian analysis*

---

\*School of Management, SUNY at Buffalo, Buffalo NY 14221. ddcho@buffalo.edu. I gratefully acknowledge the helpful comments of George C. Tiao, Jeffrey Russell, Ruey Tsay and John Cochrane. I thank Kenneth French for providing the data. All errors and omissions remain my responsibility.

# 1 Introduction

It is commonly believed that the second moments of asset returns are accurately measured. Hence most financial theories are based on the assumption that the second moments are given, or at least measured precisely. Risk-return models try to explain the cross sectional returns by risk, which is measured by the second moments. Specifically, risk based asset pricing models explain the expected excess return in terms of factors and the size of risk, measured by the covariance between the individual return and the factor return. This line of models includes the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965), intertemporal CAPM of Merton (1973) and the Arbitrage Pricing Theory of Ross (1976).

For derivative pricing, the second moment parameter also plays a crucial role. Since volatility is typically measured by the second moment, models with volatility parameters require estimating the second moment. For example, Black and Scholes (1973) introduced the Black-Scholes formula for option pricing, where the volatility is the key input variable in the formula. The volatility parameter is also essential in term structure models. Examples of term structure models include those developed by Ho and Lee (1986) and Black, Derman and Toy (1990). These models assume volatility is known, but one needs to estimate it in order to implement the models. Another area where the second moment plays an important role is Value at Risk (VaR) calculations. Since 1995, banks must compute VaR to meet the regulatory requirements. Specifically, regulators require a bank's capital to be at least three times 10-day 99% VaR. If an econometric model building approach is used for calculations, then the second moment may be required to compute the VaR.

There are two main reasons why second moment estimates are believed to be accurate. First, the distribution of returns is often assumed to be an i.i.d. normal distribution. In such situations, the variance of the second moment estimate becomes smaller as the sample size grows. In practice, most statistical software assume an i.i.d. normal distribution and calculate the confidence interval for the mean and the variance. Assuming an i.i.d. normal distribution, the confidence interval for the mean is roughly 40% wider than that for the standard deviation. The second reason for an accurate second moment estimate is due to the idea proposed by Merton (1980). Assuming a continuous time geometric Brownian motion, he showed that a precise estimate of the variance can be achieved, if the sampling interval approaches zero. The increasing availability of high frequency data may enable one to get a precise estimate of variance, if Merton's assumptions are correct.

Mandelbrot (1963) and Fama (1965), however, reported that the distribution of asset returns departs from an i.i.d. normal distribution. They found that large kurtosis and time varying volatility are two prominent features in the return data. Since then, these two phenomena have been the subject of time series analysis. Student t distributions have been used to explain the leptokurtic behavior but have failed to explain the time varying volatility. The Autoregressive Conditional Heteroskedastic (ARCH) model proposed by Engle (1982) provides a simple framework to explain the time varying volatility. Bollerslev (1986) proposed a GARCH model, which is an extension to the ARCH model. Many other extensions of the ARCH model have been suggested to explain other features such as leverage effect. Even though these models typically assume a normal distribution for the innovation, they can generate a kurtosis larger than what an i.i.d. normal distribution would suggest. The kurtosis implied by these models using a conditional normal distribution, however, does not match the large kurtosis found in the data. Bollerslev (1987) proposes to use a standardized t distribution for the innovation to generate a fatter tailed distribution. Nevertheless, the kurtosis implied by this model is greater than the sample kurtosis. Bai, Russell and Tiao (2002) use a mixture of normal distributions and match the implied kurtosis with the sample kurtosis. Most of these works try to build a simple model to explain non-normality features found in the financial data.

Bai, Russell and Tiao (2001a) argue that Merton's idea of using high frequency data to estimate the second moment may not be as useful as one might hope. They show that the large kurtosis and time varying volatility features in high frequency data can result in imprecise variance estimates even using a large amount of high frequency data. Recent works indicate that second moments can be less precisely estimated than what one would expect. Specifically, the large kurtosis and time varying volatility features will result in imprecise variance estimates.

If one is concerned not only about point estimates but also about confidence intervals, then the uncertainty in second moments has economic significance. For example, most derivative pricing theories are developed under the assumption that the true parameters are known. When implementing the theories, however, one needs to replace the parameters with the estimates. Suppose an investor uses the Black-Scholes formula to price an option. The only unknown parameter is volatility and hence an estimate is needed. Once the volatility estimate is used for pricing, the option price has a distribution because the volatility estimate is a random variable. If a confidence interval is calculated for the option price, then it will depend on the variance of the volatility estimate. Using an imprecise volatility estimate results in a wider interval and hence there may exist an incentive to use a more precise

volatility estimate. If one only cares about the point estimate of the option price, however, the uncertainty in variance will not be an issue. Since most people use point estimates rather than confidence intervals, uncertainty in second moments may have little role in practice. But, once people consider the accuracy of their calculations, the uncertainty in second moments deserves more attention.

A financial application where the uncertainty in second moments is essential can be found in a portfolio allocation problem. For a mean variance optimizing investor, the uncertainty in second moments will not affect the portfolio decision. As long as the expected returns and covariances are consistently estimated, he will use the point estimates to compute the optimal portfolio. Nevertheless, I will show that the loss due to the estimation risk depends on the uncertainty in the first two moments. This is a surprising result because it implies that the fourth moment will affect the utility of a mean variance investor whose concerns are limited to the first two moments.

Portfolio allocation problems under parameter uncertainty have been around for a long time. Zellner and Chetty (1965), Klein and Bawa (1976), and Brown (1979) provided an early application of Bayesian methods to such problems in the context of i.i.d. returns. Kandel and Stambaugh (1996) consider a problem under return predictability. They show that weak predictive regressions still result in variations in the portfolio choice of a short horizon investor. Barberis (2000) extends Kandel and Stambaugh's results to a long horizon problem. Pastor and Stambaugh (2000) explore the problem not only under parameter uncertainty but also under uncertainty in the correct asset pricing model. Most of the works in this line of literature emphasize the uncertainty in expected return. They assume the second moments are accurately measured and thus the uncertainty in second moments is of minor concern. Due to the large kurtosis and time varying volatility, the uncertainty in second moments will be larger than that under the i.i.d. normality assumption and may have economic significance to the investor. I will present a simple framework which measures the economic significance to a mean variance investor under parameter uncertainty. Furthermore, I will use 25 size and BE/ME sorted portfolio returns from 1963 to 2001 to show that the economic significance due to the uncertainty in variance can be as large as that due to the uncertainty in expected return.

The organization of the paper is as follows. In Section 2, I explain how to estimate the variance of a sample variance and related issues are discussed. In Section 3, a framework for calculating the certainty equivalent loss is introduced. I also show the empirical results using 25 size and BE/ME sorted portfolios under the sampling model and the Fama French 3 factor model. Section 4 concludes.

## 2 Measuring the Uncertainty in Variance

Second moments include variances and covariances. In this section, I present a framework to measure the uncertainty in sample variance, which is the most widely used estimator for second moments. I also evaluate this measurement for several financial data: the CRSP value weighted monthly excess returns, IBM daily returns, and half hour Dollar/Deutsch Mark exchange rate returns.

### 2.1 Framework

If a random variable  $x$  follows an i.i.d. normal distribution, then the variance of the sample variance  $s^2$  can be consistently estimated by,<sup>1</sup>

$$\widehat{\text{var}}(s^2) = \frac{2s^4}{T}, \quad (1)$$

where  $T$  is the number of observations.

Most asset return series, however, depart from this ideal distribution and have features such as large kurtosis and time varying volatility. In the presence of such features, one can estimate the variance of sample variance by the Heteroskedasticity Autocorrelation Consistent (HAC) estimation, which can be viewed as a special case of the GMM method proposed by Hansen (1982).<sup>2</sup>

$$\widehat{\text{var}}(s^2) = \frac{1}{T} \sum_{j=-\infty}^{\infty} \left[ \frac{1}{T} \sum_{t=1}^T (e_t^2 - s^2) (e_{t-j}^2 - s^2) \right], \quad (2)$$

where  $e_t = x_t - \bar{x}$ ,  $\bar{x} = \frac{1}{T} \sum_{i=1}^T x_t$ . Newey and West (1987) pointed out that the variance in Eq. (2) need not be positive semidefinite in small samples.<sup>3</sup> They suggest that one should down weight higher order correlations and propose a similar estimator which guarantees the covariance matrix to be positive semidefinite.

$$\widehat{\text{var}}(s^2) = \frac{1}{T} \sum_{j=-k}^k \left[ \frac{1}{T} \sum_{t=1}^T \frac{k-|j|}{k} (e_t^2 - s^2) (e_{t-j}^2 - s^2) \right] \quad (3)$$

---

<sup>1</sup>The exact variance of  $s^2$  is  $\frac{2\sigma^4}{T-1}$  and the unbiased estimator for  $\text{var}(s^2)$  is  $\frac{2s^4}{T+1}$ . Nevertheless, for a large sample, dividing  $2s^4$  by  $T+1$  or  $T$  does not make much difference.

<sup>2</sup>For detailed survey of HAC estimator, see Andrews (1991).

<sup>3</sup>Since  $s^2$  is a scalar, the variance in Eq. (2) is invertible. But later, we will extend this to more than one return series so the positive semidefiniteness becomes an issue.

Throughout this paper, I will refer to the GMM estimator as the variance estimated by Eq. (3). The GMM estimator is not only consistent but also robust because it does not assume any parametric model for estimation. Cochrane (2001), however, argues this nonparametric estimator can perform poorly in finite samples because it needs to estimate many high order correlations. He suggests imposing a parametric structure for the autocorrelation and heteroskedasticity can improve the estimation. Since the GARCH(1,1) model is a commonly used model for explaining time varying volatility, it is a natural choice for the parametric structure.

Suppose the deviation,  $e_t$ , follows a GARCH(1,1) model.<sup>4</sup>

$$\begin{aligned} e_t &= \sqrt{h_t}z_t, \quad z_t \sim \text{i.i.d. } (0, 1) \\ h_t &= \omega + \alpha e_{t-1}^2 + \beta h_{t-1} \end{aligned} \quad (4)$$

Bai, Russell and Tiao (2001b) assume  $E(x_t) = 0$  and show the variance of  $s^2$  can be expressed in terms of the excess kurtosis and the GARCH parameters. Following a similar approach, the variance of  $s^2$  can be estimated by,

$$\widehat{\text{var}}(s^2) = \frac{2s^4}{T} \left( 1 + \frac{\widehat{\gamma}}{2} \right) \left( 1 + \frac{2\widehat{\rho}}{1 - \widehat{\alpha} - \widehat{\beta}} \right) \quad (5)$$

where  $\gamma$  denotes the excess kurtosis of  $e_t$ ,  $\gamma = E(e_t^4)/(E(e_t^2)^2) - 3$ , and  $\rho$  denotes the lag 1 autocorrelation of  $e_t^2$ . See the Appendix for derivation. Eq. (5) lends insight into how the features of asset return can affect the variance of sample variance. The first term in the right hand side of (5) is identical to (1), which assumes an i.i.d. normal distribution. The second term measures the effect of kurtosis. As one would expect, the larger the kurtosis is, the more variability in the sample variance. The last term shows the effect of the time varying volatility, where the volatility persistence in a GARCH(1,1) process is measured by the sum of  $\alpha$  and  $\beta$ . Therefore, the more persistence in volatility, the less accurate the variance estimate becomes.

## 2.2 Parametric Estimation of $\text{var}(s^2)$

To implement Eq. (5), one needs to estimate the excess kurtosis ( $\gamma$ ), the lag 1 autocorrelation ( $\rho$ ) and the GARCH parameters ( $\alpha, \beta$ ). It is straightforward to estimate the GARCH

---

<sup>4</sup>Theoretically, deviation cannot be a GARCH series unless the mean is known. For an unknown mean, the sum of deviations must add up to zero, which implies there is one less degree of freedom. We assume the sample size is large enough so that degrees of freedom is not an issue.

parameters,  $\alpha$  and  $\beta$ , by a QMLE technique. Then the lag 1 autocorrelation of  $e_t^2$ ,  $\rho$ , can be estimated by the GARCH estimates.

$$\hat{\rho} = \hat{\alpha} \left( 1 + \frac{\hat{\alpha}\hat{\beta}}{1 - \hat{\beta}^2 - 2\hat{\alpha}\hat{\beta}} \right) \quad (6)$$

Estimating the excess kurtosis deserves explanation. One can simply estimate it by the sample excess kurtosis. Cho (2002), however, argues that the sample kurtosis can perform poorly in finite samples. Specifically, sample kurtosis will be smaller than the true kurtosis and it will result in a smaller variance of the sample variance. In other words, if one blindly uses sample excess kurtosis, then the variance of sample variance tends to be smaller than the true value.

Bai, Russell and Tiao (2002) derive an analytical expression for the excess kurtosis in a GARCH(1,1) model. Suppose  $e_t$  follows the GARCH(1,1) process in Eq. (4). Then the excess kurtosis of  $e_t$  is given by,

$$\gamma = \frac{\gamma^{(g)} + \gamma^{(z)} + \frac{5}{6}\gamma^{(g)}\gamma^{(z)}}{1 - \frac{1}{6}\gamma^{(g)}\gamma^{(z)}}, \quad (7)$$

$$\text{where } \gamma^{(g)} = \frac{6\alpha^2}{1 - 3\alpha^2 - 2\alpha\beta - \beta^2} \quad (8)$$

$$\gamma^{(z)} = E(z_t^4) - 3. \quad (9)$$

Assuming a normal distribution for the innovation, then  $\gamma$  is equal to  $\gamma^{(g)}$ , which is larger than 0. It can generate kurtosis larger than an i.i.d. normal distribution, however, it does not generate large enough kurtosis to explain the behavior of a return series. Instead, one can assume a standardized t distribution or a mixture of normal distributions for the innovation in the GARCH model and estimate the parameters by the MLE method. Cho (2002) argues that even though the kurtosis implied by these ML estimates can match the true excess kurtosis better than the sample excess kurtosis, there are two drawbacks to this method. First, high persistence in the GARCH process yields unstable estimates of the kurtosis. This is expressed in a multiplier effect in Eq. (7). Specifically, if  $(\gamma^{(z)} + 3)\alpha^2 + 2\alpha\beta + \beta^2$  gets close to 1, then one can show that the denominator in (7) gets close to 0, which implies that a small change in parameter estimates can cause wide dispersion in excess kurtosis. Second, ML estimation requires a correct specification for the distribution and hence it is vulnerable to misspecification. Unless one is certain about the correct model for the return series, the kurtosis estimate can be misleading.

One possible solution to solve the misspecification problem is to estimate the excess kurtosis semiparametrically. Even though the sample kurtosis of  $e_t$  can be downward biased

in small samples, the sample kurtosis of standardized residual,  $\hat{z}_t = e_t/\text{sqrt}(\hat{h}_t)$ , is accurately measured. Since only the consistent estimate of  $\hat{h}_t$  is required, one can even use the QMLE estimation assuming a normal distribution for the innovation. Then one can compute the excess kurtosis in Eq. (7) by the GARCH estimates and the sample excess kurtosis of  $\hat{z}_t$ . Therefore, this method is robust to the distribution assumption for the innovation.

In short, if the deviations follow a GARCH(1,1) model, then the variance of  $s^2$  estimated by Eq. (5), imposing a parametric structure, will perform the best. In addition, one can calculate (5) parametrically using the GARCH estimates. Nevertheless, if one is uncertain about the correct specification, then one should estimate the variance of  $s^2$  nonparametrically by the GMM method.

## 2.3 Empirical Results

In this subsection, I measure the uncertainty in sample variance for four data sets. The first data set is the CRSP value weighted monthly excess returns (VW) from 1926:07 to 2001:12. I also investigate the value weighted returns during a subsample period starting from 1963:07, a common starting period for empirical finance studies. The third data set is daily returns of IBM from 1990 to 2000. The last data is half hour foreign exchange rate returns obtained from Olsen and Associates. I analyze one complete year (1996) of half hour exchange rate on Dollar/Deutsche Mark(DD). The DD series is MA(1) adjusted to take out the serial correlation in the first moment. See Bai, Russell and Tiao (2001a) for more details. It is also standardized for convenience because the scale is too small. In this analysis, I am not estimating the variance but estimating the variance of the variance, and hence using the standardized return is not an issue.

Table 1 reports the summary statistics of the four data sets. The CRSP value weighted returns (VW) in the entire sample period show a large variance and a large kurtosis. On the other hand, the CRSP value weighted return (VW2) during the subsample looks more like a normal distribution. The sample excess kurtosis is 2.1822, much smaller than 7.7754 using the entire sample. In addition, the volatility is much lower than the entire sample period. The main reason for the more volatile return in the entire sample is due to the Great Depression. Excluding the data prior to 1940, the variance becomes 18.28 and the kurtosis becomes 2.5235, which are similar to the results of the subsample. Both the IBM and DD series show large kurtosis, which is an evidence of non-normality behavior.



	VW(%)	VW2(%)	IBM(%)	DD
sample period	1926-2001	1963-2001	1990-2000	1996
frequency	monthly	monthly	daily	half hour
sample size	904	460	2,780	7,488
mean	0.6485	0.4524	0.0759	0.0000
variance	30.4531	19.9296	4.1718	1.0000
skewness	0.2245	-0.5022	0.1652	-0.1344
excess kurtosis	7.7754	2.1822	6.5175	8.3416

Table 1: Summary statistics

The table reports the summary statistics for four data sets. VW is monthly excess return of the CRSP value weighted portfolio from 1926:07 to 2001:12. VW2 is also the CRSP value weighted return but using a subsample period from 1963:07 to 2001:12. IBM is daily return from the CRSP database from 1990 to 2000. DD is half hour return on Dollar/Deutsche Mark foreign exchange rate from Olsen and Associates. DD is MA(1) adjusted to take out the serial correlation in the return and also standardized.

	VW	VW2	IBM	DD
$r_t - \bar{r} = e_t, e_t = \sqrt{h_t} z_t, h_t = \omega + \alpha e_{t-1}^2 + \beta h_{t-1}$				
$\hat{\omega}$	0.7116 (0.3521)	0.9064 (0.6711)	0.0309 (0.0156)	0.1668 (0.0262)
$\hat{\alpha}$	0.1208 (0.0287)	0.0850 (0.0303)	0.0378 (0.0086)	0.1470 (0.0232)
$\hat{\beta}$	0.8588 (0.0336)	0.8766 (0.0556)	0.9570 (0.0096)	0.6876 (0.0396)

Table 2: GARCH parameter estimates

The table reports the GARCH parameter estimates using a standard QMLE technique. The robust standard errors are reported in the parentheses.

	VW	VW2	IBM	DD
Panel A: i.i.d. normal distribution				
$\widehat{\text{var}}(s^2)$	2.0517	1.7269	0.0125	0.0003
Panel B: GMM				
$\widehat{\text{var}}(s^2)$	97.4664	6.2590	0.7497	0.0087
Panel C: GARCH(1,1)				
$\widehat{\alpha} + \widehat{\beta}$	0.9796	0.9616	0.9949	0.8345
$1 + \frac{\widehat{\gamma}}{2}$	4.8877	3.8594	4.2588	13.6427
$1 + \frac{\widehat{\gamma}}{1 - \widehat{\alpha} - \widehat{\beta}}$	35.1468	9.4231	61.3237	3.3283
$\widehat{\text{var}}(s^2)$	352.4654	62.8207	3.2700	0.0121

Table 3: Variance of sample variance

The table reports the variance of sample variance. Panel A assumes an i.i.d. normal distribution and estimates the variance of sample variance by Eq. (1). Panel B reports GMM estimates by Eq. (3). Prespecified truncation lags used for GMM estimates are 120 for VW, 60 for VW2, 1,000 for IBM and DD. Panel C assumes a GARCH(1,1) process for the return series and estimates the variances by Eq. (5). Panel C also reports the components used in Eq. (5).

Table 2 reports the GARCH parameter estimates by the QMLE method. The table also reports the robust standard error accounting for non-normality. The persistence in volatility, which is measured by the sum of two GARCH parameters,  $\alpha$  and  $\beta$ , is very close to 1. This indicates very strong volatility persistence in the data.

Table 3 reports the estimated variance of the sample variance under the three methods: (1), (3) and (5). If one blindly assumes an i.i.d. normal distribution, then the variance of sample variance is much smaller than those under other assumptions, as shown in Panel A. Since most statistical software use this assumption to calculate the confidence interval for the second moment parameters, the small numbers reported by the software can be misleading. The GMM estimates are calculated from Eq. (3), where the prespecified truncation lags are 120 for VW, 60 for VW2, 1,000 for IBM and 1,000 for DD.<sup>5</sup> The GMM estimates show how small  $\widehat{\text{var}}(s^2)$  is under the i.i.d. normal assumption. Specifically, the GMM estimates are 4 to 60 times larger than those under the i.i.d. normality case.

If one can correctly specify the parametric structure, one may be able to obtain a better

<sup>5</sup>Andrews (1991) and Newey and West (1992) offered some guidance for choosing an optimal truncation lag.

	VW	VW2	IBM
Panel A: $SCV(\bar{r})$			
	0.0801	0.2117	0.2608
Panel B: $SCV(s^2)$			
i.i.d. normality	0.0022	0.0043	0.0007
GMM	0.1051	0.0158	0.0431
GARCH(1,1)	0.3801	0.1581	0.1879

Table 4: Squared coefficient of variation of the sample moments

The table reports the squared coefficient of variation (SCV) of the sample mean and the sample variance.  $SCV(\bar{r})$  is defined as  $\frac{\widehat{\text{var}}(\bar{r})}{\bar{r}^2}$  and  $SCV(s^2)$  is  $\frac{\widehat{\text{var}}(s^2)}{s^4}$ . SCV of the sample variance is computed under three different methods. Assuming an i.i.d. normal distribution, the SCV of the sample variance is  $\frac{2}{T}$ . The GMM estimator is calculated using the Newey and West method as in (3). For the GARCH(1,1) model,  $\widehat{\text{var}}(s^2)$  is estimated by Eq. (5).

estimate for the variance of sample variance. Assuming the GARCH(1,1) model is the correct model, Eq. (5) can be used for calculating  $\widehat{\text{var}}(s^2)$ . The variance of sample variance under the GARCH(1,1) specification is much larger than that under the i.i.d. normality assumption, moreover, they are even larger than the GMM estimates. For example, the GARCH variance for VW2 results in ten times larger than the GMM estimate. The reason for the small GMM estimate is due to the fact that asset returns during the subsample period behave more like a normal distribution and hence the second moment can be more accurately estimated. At this point, I cannot conclude which method is preferable. Since the GARCH(1,1) estimates are vulnerable to misspecification, one cannot take the estimates given by the GARCH model for granted. Nevertheless, no matter which method one uses, GMM or GARCH, it is clear that the precision of the second moments is much worse than that under the i.i.d. normality assumption.

Motivated by the findings that the variance is less accurately estimated, I would like to compare the precision of the sample variance with that of the sample mean return. It is commonly believed that the expected return is harder to estimate than the variance of return. This is true in most of the cases under the i.i.d. normality assumption. The departures from the ideal distribution, however, can make it more interesting. Since the first and second moments differ in units and size, one needs to standardize them to make any meaningful comparison.

The squared coefficient of variation (SCV) of a random variable is defined as the ratio

between the variance of the random variable and the squared mean of the random variable.<sup>6</sup> Since SCV is standardized to take out the units, one can make meaningful comparisons across different variables. Table 4 reports the SCV of the first two sample moments. First, I argue that the second moment is less accurately measured than the first moment for VW. The SCV of the sample mean is 0.0801 but the SCV of the sample variance is 0.1051 using the GMM estimate, and 0.3801 assuming the GARCH(1,1) model.<sup>7</sup>

The more recent data set from 1963:07 to 2001:12 results in more accurate sample variance than using the entire sample. If VW2 has the same distribution as VW, then SCV should double because the sample size of VW2 is roughly half of that of VW. The SCV of the average return is 0.2117, which implies that the mean return becomes less accurate than the case using the entire sample. On the other hand, the SCV of the sample variance is 0.0043 assuming i.i.d., 0.0158 by the GMM estimate, and 0.1581 assuming the GARCH(1,1). The main reason for a more accurate sample variance is due to a smaller kurtosis and less volatility persistence after 1963. The excess kurtosis in the subsample period is 2.18, which is much smaller than the excess kurtosis in the entire sample, 7.78. The sum of the two GARCH parameters,  $\alpha$  and  $\beta$ , which measures the volatility persistence, is 0.9616 in the subsample, compared with 0.9796 during the entire sample. These two features imply that VW2 behaves closer to a normal distribution than VW. Hence the accuracy of the sample variance improves during the subsample. On the other hand, if one assumes a GARCH(1,1) model, then the SCV of sample variance is 0.1581, not very different from the SCV of average return.

The SCV of IBM implies how hard it is to estimate the moments using high frequency data. Even though the sample size is 2,780, the variance is not accurately estimated, assuming that the GARCH(1,1) process is the correct model. This is due to a high persistence in volatility. The volatility persistent measurement,  $\hat{\alpha} + \hat{\beta}$ , is 0.9949, which implies any shock in the variance is persistent. In such a case, estimating the unconditional variance becomes extremely difficult.

To summarize, contrary to what people believe, the uncertainty in second moments is large. This is due to the large kurtosis and time varying volatility features, which are commonly found in returns data. Using the CRSP value weighted monthly excess returns for the period of 1926 to 2001, I have found that the sample variance is less precise than the

---

<sup>6</sup>Operations management is one of the academic areas which uses SCV often.

<sup>7</sup>The autocorrelation structure reveals that the market excess return series has positive lag 1 autocorrelation and negative lag 3 autocorrelation. Thus the simple calculation method of  $s^2/T$  could be misleading. Nevertheless, GMM estimates show  $s^2/T$  is a good approximation and hence we will use the conventional method.

sample mean return. Even though other data sets show that the variances are more precisely estimated than the expected returns, the variances are much less accurately estimated than under the i.i.d. normality assumption.

### 3 Portfolio Allocation

In Section 2, I have shown that the second moments are less accurately measured, when the i.i.d. normality assumption is violated. Motivated by this finding, I explore the economic significance of the uncertainty in second moments. Specifically, I ask whether the uncertainty in second moments affects the investor and compare the effect of the uncertainty in second moments with that of the uncertainty in first moments.

#### 3.1 Certainty Equivalent Loss under Parameter Uncertainty

Suppose an investor tries to maximize a mean-variance objective function. Define  $\mu$  as an  $N$  by 1 vector of expected excess returns and  $\Sigma$  as an  $N$  by  $N$  covariance matrix of returns. Denote  $w$  as an  $N$  by 1 vector of weights on  $N$  assets. Then the investor's portfolio allocation problem becomes,

$$\max_w w' \mu - \frac{1}{2} A w' \Sigma w, \quad (10)$$

subject to constraints in the weights, where  $A$  is the coefficient of absolute risk aversion. The unconstrained optimal portfolio is well known,

$$w_o = \frac{1}{A} \Sigma^{-1} \mu. \quad (11)$$

For simplicity, assume there is no margin requirement. Later, I will introduce a 50% margin requirement imposed by Regulation T, which applies to most investors and institutions. The objective function evaluated at the optimal portfolio can be regarded as the maximized certainty equivalent to the investor.<sup>8</sup>

$$CE_o = w'_o \mu - \frac{1}{2} A w'_o \Sigma w_o \quad (12)$$

A certainty equivalent is a fixed return, which gives the same expected utility for an investor who faces a random return, or a return distribution. Since the investor has no information on the true moments,  $\mu$  and  $\Sigma$ , however, he needs to replace the true moments in Eq. (11) with

---

<sup>8</sup>The maximized certainty equivalent can be simplified to  $\frac{1}{2} w'_o \mu$ .

the estimates given by the model he employs. He can use sample moments or the moments implied by any factor model. Alternatively, he can also employ a Bayesian approach and use the moments given by a predictive distribution.<sup>9</sup> No matter which method the investor uses, the portfolio using these estimates will be suboptimal.

$$\hat{w}_s = \frac{1}{A} \hat{\Sigma}^{-1} \hat{\mu} \quad (13)$$

Then the objective function evaluated at the suboptimal portfolio,  $\hat{w}_s$ , can be viewed as the suboptimal certainty equivalent.

$$CE_s = \hat{w}_s' \mu - \frac{1}{2} A \hat{w}_s' \Sigma \hat{w}_s \quad (14)$$

Note that even though the investors do not have full information about the true moments, I will evaluate the certainty equivalent of suboptimal portfolio using the true moments in order to reflect the true certainty equivalent. Thus the difference between the optimal certainty equivalent in (12) and the suboptimal certainty equivalent in (14) measures the certainty equivalent loss due to parameter uncertainty.

$$CE \text{ loss} = CE_o - E(CE_s) \approx \frac{1}{2} A \times tr [\text{cov}(\hat{w}_s) \Sigma] \quad (15)$$

See the Appendix for derivation. Equality holds, if  $\hat{w}_s$  is an unbiased estimator of  $w_o$ . The reason for taking the expected value of the certainty equivalent is that the certainty equivalent does not reflect all the uncertainties that the investor faces. There are two sources of uncertainties: one in the future return,  $r_{T+1}$ , and the other in the estimates, or the sampling distribution. The uncertainty in the future return is taken care by the certainty equivalent. On the other hand, the uncertainty in the estimates is not embedded in the certainty equivalent, and thus one needs to take the expectation of certainty equivalent over the sampling distribution in order to reflect all the uncertainties that the investor faces.

The certainty equivalent loss due to the uncertainty in the first two moments can be approximated by this simple method. Furthermore, Eq. (15) can be applied to a constrained portfolio problem because the derivation does not rely on the unconstrained optimal portfolio. The difference in certainty equivalent has been used in Frost and Savarino (1986), Kandel and Stambaugh (1996), Pastor and Stambaugh (2000). They compare the feasible optimal portfolio accounting for estimation risk with the suboptimal portfolio, and their approaches are different from here in that respect. I compute the certainty equivalent at the suboptimal

---

<sup>9</sup>Brown (1979), Frost and Savarino (1986) showed that the Bayesian approach results in higher expected utility than a non-Bayesian approach which uses sample moments.

and compare it with the certainty equivalent at the optimal assuming the parameters are known. Thus this approach measures the total loss due to the parameter uncertainty.

Surprisingly, Eq. (15) implies that the fourth moment may affect the certainty equivalent loss to an investor. The covariance matrix of  $\widehat{w}_s$  in (15) depends on the covariance matrix of  $\widehat{\Sigma}$  because  $\widehat{w}_s$  is a function of  $\widehat{\Sigma}$ . Furthermore, since the covariance of  $\widehat{\Sigma}$  depends on the kurtosis, the fourth moment will affect the certainty equivalent loss to a mean variance investor. Note that even though only the first two moments matter to a mean variance investor, the fourth moment has an effect on the investor, if he needs to estimate the two moments. To summarize, the uncertainty in second moments can affect the expected certainty equivalent loss to a mean variance investor facing parameter uncertainty, and thus it has economic significance to the portfolio allocation problem.

### 3.2 Single Risky Asset Case

Since  $\widehat{w}_s$  is nonlinear in moment estimates, it is difficult to obtain an analytical expression for  $\text{cov}(\widehat{w}_s)$  in Eq. (15). Nevertheless, if there is only one risky asset,<sup>10</sup> suppose the market portfolio, then the certainty equivalent loss can be expressed as,

$$CE \text{ loss} \approx \frac{1}{2}A \times \text{var}(\widehat{w}_s) \times \text{var}(r_M), \quad (16)$$

where  $r_M$  is the market excess return. In order to evaluate the certainty equivalent loss, one can replace the variances with the estimated variance of  $\widehat{w}_s$  and the estimated variance of  $r_M$ . First, one can estimate the variance of  $\widehat{w}_s$  by the Delta method.

$$\begin{aligned} \widehat{\text{var}}(\widehat{w}_s) &= \widehat{\text{var}}\left(\frac{1}{A}\widehat{\Sigma}^{-1}\widehat{\mu}\right) \\ &= \frac{1}{A^2}\widehat{\text{var}}\left(\frac{\bar{r}_M}{s_M^2}\right) \\ &\approx \frac{1}{A^2}\left(\frac{\bar{r}_M}{s_M^2}\right)^2\left[\frac{\widehat{\text{var}}(\bar{r}_M)}{\bar{r}_M^2} + \frac{\widehat{\text{var}}(s_M^2)}{s_M^4}\right] \end{aligned} \quad (17)$$

Next, one can estimate the variance of the market returns by the sample variance.

$$\widehat{\text{var}}(r_M) = s_M^2 \quad (18)$$

---

<sup>10</sup>1- $\widehat{w}_s$  will be invested in the risk free asset. Thus one can view this as a two asset case as well, one risky asset and one risk free asset.

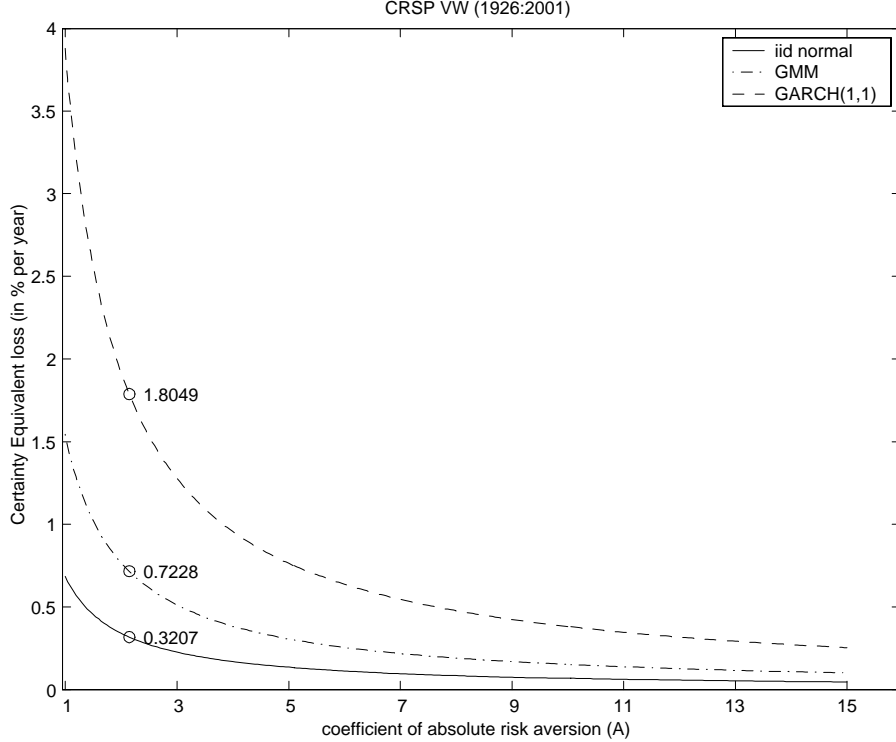


Figure 1: Certainty equivalent loss (1926-2001)

Then the certainty equivalent loss in the case of one risky asset becomes,

$$\begin{aligned}
 CE \text{ loss} &\approx \frac{1}{2A} \frac{\bar{r}_M^2}{s_M^2} \left[ \frac{\widehat{\text{var}}(\bar{r}_M)}{\bar{r}_M^2} + \frac{\widehat{\text{var}}(s_M^2)}{s_M^4} \right] \\
 &= \frac{1}{2A} \frac{\bar{r}_M^2}{s_M^2} [\text{SCV}(\bar{r}_M) + \text{SCV}(s_M^2)]. \tag{19}
 \end{aligned}$$

Eq. (19) implies that the loss can be decomposed into two components: uncertainty due to estimating the first moment, and that due to the second moment. Specifically, the size of the loss depends on the SCV of the first two sample moments. As discussed in subsection 2.3, the SCV of the sample variance is larger than that of the sample mean return using the CRSP value weighted returns during the entire sample period. A simple comparison between the SCVs shows the loss due to the uncertainty in the second moment is the larger of the two, if one uses the entire sample period to estimate the moments.

The coefficient of absolute risk aversion,  $A$ , is a free parameter denoting different investors. If one sets  $A$  equal to 2.13, then the optimal weight is 100%, which means he holds the market portfolio and leaves no wealth in the risk free asset. Table 5 reports the certainty equivalent and losses under three different assumptions. For an investor with risk aversion



	VW (1926:2001)			VW2 (1963:2001)		
	i.i.d. N	GMM	GARCH	i.i.d. N	GMM	GARCH
CE(optimal)	3.9604	3.9604	3.9604	2.7488	2.7488	2.7488
CE loss( $\bar{r}_M, s_M^2$ )	0.3207	0.7228	1.8049	0.5880	0.6191	1.0085
CE loss( $\bar{r}_M$ )	0.3120	0.3120	0.3120	0.5761	0.5761	0.5761
CE loss( $s_M^2$ )	0.0086	0.4096	1.4886	0.0118	0.0428	0.4301

Table 5: Certainty equivalent comparison in a single risky asset case

The table reports the maximized certainty equivalent (% per annum) at the optimal portfolio assuming the moment estimates given by the model are true parameters. The table also reports the certainty equivalent losses due to parameter uncertainty.

CE(optimal) is the maximized certainty equivalent assuming the sample moments are the true parameters. CE loss( $\bar{r}_M, s_M^2$ ) measures the expected certainty equivalent loss to an investor who has to estimate the first two moments to obtain his portfolio. CE loss( $\bar{r}_M$ ) measures the loss due to the uncertainty in first moments alone, assuming the investor has full information on the true second moments. On the other hand, CE loss( $s_M^2$ ) assumes full information on expected returns is available to the investor but that the true second moments are unknown. The only risky asset is the CRSP value weighted portfolio and the sample periods for VW and VW2 are 1926:2001 and 1963:2001, respectively.

equal to 2.13, the certainty equivalent at the optimal weight,  $w_o = 1$ , is equal to 3.9604% per annum.<sup>11</sup> Since the investor does not observe the true moments, he will incur some loss due to the estimation risk. The loss is 0.3207% per annum under the i.i.d. normal case, 0.7228% using the GMM estimates, and 1.8049% assuming GARCH(1,1). In other words, the investor loses up to a fixed return of 1.8049% per annum due to estimating the first two moments. The loss due to the uncertainty in first moment alone, assuming he has full information on the second moment, is 0.3120%, roughly one sixth of the entire loss. Therefore, the loss is mainly due to the second moment uncertainty. The result is surprising because the second moment is usually believed to be more accurately estimated than the first moment. Even if one uses the robust GMM estimate for the SCV of the second moment, the loss due to the second moment is 0.4096%, which is still larger than the loss due to the first moment. Note that if one assumes an i.i.d. normality, then the loss due to the variance is only 0.0086%, practically none. This shows how one can underestimate the loss due to uncertainty in variance by assuming an i.i.d. normal distribution. Figure 1 plots the loss (in % per annum) against  $A$ , the coefficient of absolute risk aversion, under 3 different assumptions. Circles in the figure denote an investor who holds 100% in the market portfolio. The certainty equivalent

<sup>11</sup>Since the certainty equivalent depends on the unknown first two moments, one cannot calculate the true certainty equivalent. Here we use the sample moments as the point estimate of the true moments in order to evaluate the certainty equivalent.

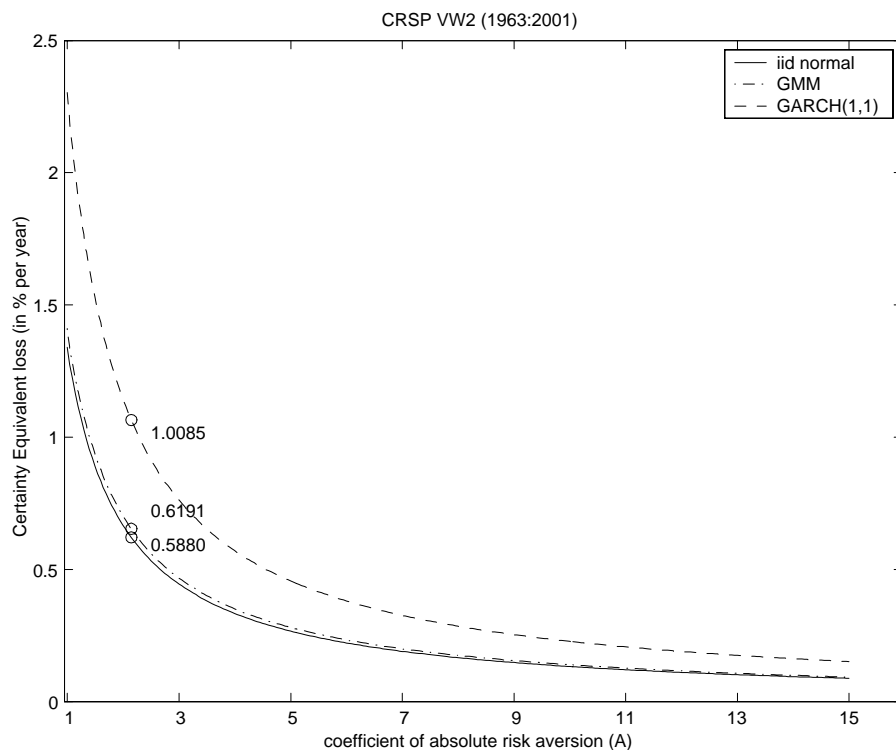


Figure 2: Certainty equivalent loss (1963-2001)

lent decreases as investors are more risk averse because a more risk averse investor will hold a smaller amount of the risky market portfolio. In addition, the differences between the losses under the three different assumptions become smaller for a more risk averse investor.

If the investor uses only the data after 1963 to estimate the moments, then the losses become smaller. The certainty equivalent to an investor with 100% weight in the market portfolio is 2.7488%.<sup>12</sup> Figure 2 plots the certainty equivalent loss against the risk aversion. The loss using the GARCH(1,1) model is 1.0085%, smaller than 1.8049% using the entire sample. The main reason is that the variance is more precisely estimated in the subsample period. If the first moment is the only source of uncertainty, then the loss is 0.5761%, more than half of the loss. Thus the loss due to the second moment is less than half of the loss. This is contrary to the result using the entire sample period, where the loss is mainly due to the uncertainty in the second moment. Instead of assuming GARCH(1,1), if one uses GMM estimates to compute the loss, the total loss is 0.6191%. The loss due to the second moment estimation is only 0.0428% compared with the loss of 0.5761% due to the first moment. The reason for the smaller loss due to the uncertainty in the variance is that the asset returns

<sup>12</sup>Coefficient of absolute risk aversion corresponds to 2.27.

behave more like an i.i.d. normal distribution during the subsample period. In a single risky asset case, in short, the loss due to estimating the variance is negligible, if one uses the data from 1963 to 2001. The results, however, do not extend to a multiple asset case.

### 3.3 Multiple Risky Asset Case

If there are  $N$  risky assets, the covariance matrix of the weight will be harder to calculate. In such a case, one can generate random draws of the first two moments from a multivariate distribution. From the randomly drawn moments, the weights can be calculated by Eq. (13). Then one can estimate the covariance matrix of weights from the simulated weights, which can be plugged into the formula in (15) to calculate the loss.

Fama and French (1993) introduced a three-factor model to explain anomalies such as the size effect and the value effect. This is a parsimonious model which can fit the empirical evidence. The three factors are market excess returns, excess returns on a portfolio of small stocks over big stocks and excess returns on a portfolio of high BE/ME over low BE/ME stocks. The model that the investor uses to estimate the moments will include the sampling model and the Fama French 3 factor model as well.<sup>13</sup>

Table 6 reports the first (%) and second (% square) moment estimates of 25 size, BE/ME sorted portfolio excess returns given by the sample moments and the Fama French (FF) 3 factor model. It also reports the robust standard error of the estimates using the GMM method. The sample period is from 1963:07 to 2001:12. I use 1963:07 as the starting date because that is when COMPUSTAT data becomes reliable. This sample period matches the subsample period of the single asset case, VW2, in subsection 3.2, where I have found the loss due to the second moment is small. In a single risky asset case, I have found that the loss due to the uncertainty in the first and second moments depends on the size of the SCVs. Table 7 reports the SCVs of the first and second moments of 25 portfolios. The average SCVs of the first moments are 0.2983 for the sampling model, and 0.1913 for the FF 3 factor model, which implies that one can reduce the uncertainty in the first moment by the FF 3 factor model. On the other hand, both the average SCVs of the variances across 25 portfolios are 0.0187. Due to the variance decomposition rule, both models result in almost identical numbers for the variances. Similar to the CRSP VW returns, the SCVs of the second moments are much smaller than those of the first moments, whichever model one uses. If one employs the FF

---

<sup>13</sup>This paper, however, does not consider the uncertainty in the asset pricing model. We present the results of both models independently, assuming the moments given by each model are the correct moments.

Book to Market Equity (BE/ME) Quintiles										
Size	Low	2	3	4	High	Low	2	3	4	High
Panel A: Sampling model: $\hat{\mu}_i = \bar{r}_i, \hat{\sigma}_i^2 = s_i^2$										
	$\hat{\mu}$					$SE(\hat{\mu})$				
Small	0.235	0.733	0.815	0.998	1.075	0.384	0.329	0.286	0.267	0.277
2	0.359	0.622	0.842	0.912	0.966	0.349	0.285	0.252	0.240	0.264
3	0.380	0.665	0.673	0.818	0.953	0.322	0.257	0.231	0.219	0.247
4	0.506	0.433	0.651	0.820	0.880	0.286	0.243	0.226	0.215	0.249
Big	0.442	0.441	0.493	0.603	0.583	0.225	0.214	0.202	0.200	0.216
	$\hat{\sigma}^2$					$SE(\hat{\sigma}^2)$				
Small	68.00	49.90	37.70	32.77	35.32	12.55	8.48	5.29	4.63	5.29
2	56.16	37.45	29.27	26.57	32.03	7.33	4.86	3.94	3.67	4.45
3	47.72	30.37	24.62	22.19	28.06	6.81	3.64	3.13	2.58	3.87
4	37.78	27.23	23.57	21.36	28.59	6.07	3.79	2.90	2.74	3.83
Big	23.43	21.05	18.87	18.40	21.57	3.09	2.74	2.48	2.19	2.06
Panel B: Fama French 3 factor model: $r_i = b_{1i}MKT + b_{2i}SMB + b_{3i}HML + e_i$										
	$\hat{\mu}$					$SE(\hat{\mu})$				
Small	0.588	0.726	0.769	0.824	0.952	0.369	0.320	0.279	0.261	0.271
2	0.520	0.708	0.778	0.835	0.965	0.340	0.276	0.244	0.233	0.257
3	0.448	0.676	0.760	0.811	0.945	0.314	0.244	0.218	0.208	0.235
4	0.363	0.641	0.731	0.767	0.931	0.278	0.227	0.212	0.203	0.231
Big	0.236	0.482	0.513	0.681	0.798	0.218	0.202	0.186	0.186	0.192
	$\hat{\sigma}^2$					$SE(\hat{\sigma}^2)$				
Small	68.13	49.90	37.70	32.80	35.34	12.56	8.48	5.28	4.61	5.28
2	56.18	37.46	29.27	26.57	32.03	7.37	4.88	3.93	3.65	4.45
3	47.73	30.37	24.63	22.19	28.06	6.83	3.64	3.15	2.58	3.87
4	37.80	27.27	23.58	21.36	28.59	6.05	3.83	2.91	2.74	3.83
Big	23.48	21.05	18.87	18.40	21.62	3.07	2.74	2.49	2.20	2.09

Table 6: Expected excess return and variance of return estimates and standard errors of 25 sorted portfolios

The table reports the first and second moment estimates of 25 size and BE/ME sorted portfolios. The moments are given by sample moments in Panel A and by the Fama French 3 factor model in Panel B. The sample period is from 1963:07 to 2001:12. Covariances are not reported for simplicity. Standard errors of the second moments in both Panels are estimated by the GMM estimator with truncation lag equal to 30. In Panel A, standard error of the sample mean is calculated by sample standard deviation divided by square root of number of observations. In Panel B, standard error of the first moment given by the FF3 model is calculated by  $cov(\hat{\mu}_i, \hat{\mu}_j) = b_i' cov(\bar{f}) b_j + tr[cov(\bar{f}) cov(b_i, b_j)]$ .

Book to Market Equity (BE/ME) Quintiles										
Size	Low	2	3	4	High	Low	2	3	4	High
Panel A: Sampling model: $\hat{\mu}_i = \bar{r}_i, \hat{\sigma}_i^2 = s_i^2$										
	SCV( $\hat{\mu}$ )					SCV( $\hat{\sigma}^2$ )				
Small	2.668	0.202	0.123	0.071	0.066	0.034	0.029	0.020	0.020	0.022
2	0.946	0.210	0.090	0.069	0.074	0.017	0.017	0.018	0.019	0.019
3	0.718	0.149	0.118	0.072	0.067	0.020	0.014	0.016	0.014	0.019
4	0.320	0.314	0.121	0.069	0.080	0.026	0.019	0.015	0.016	0.018
Big	0.260	0.235	0.168	0.110	0.138	0.017	0.017	0.017	0.014	0.009
Panel B: Fama French 3 factor model: $r_i = b_{1i}MKT + b_{2i}SMB + b_{3i}HML + e_i$										
	SCV( $\hat{\mu}$ )					SCV( $\hat{\sigma}^2$ )				
Small	0.394	0.194	0.132	0.100	0.081	0.034	0.029	0.020	0.020	0.022
2	0.429	0.152	0.099	0.078	0.071	0.017	0.017	0.018	0.019	0.019
3	0.491	0.131	0.082	0.066	0.062	0.020	0.014	0.016	0.014	0.019
4	0.585	0.126	0.084	0.070	0.061	0.026	0.020	0.015	0.016	0.018
Big	0.857	0.177	0.131	0.075	0.058	0.017	0.017	0.017	0.014	0.009

Table 7: Squared coefficient of variation of 25 sorted portfolios

The table reports the squared coefficient of variation of the first two moments for 25 size and BE/ME sorted portfolios. Squared coefficient of variation is defined as the variance divided by squared mean. SCV is unit less and thus makes it possible to compare across different variables. The sample period is from 1963:07 to 2001:12. SCVs of the second moments in both Panels are estimated by the GMM estimator with truncation lag equal to 30. In Panel A, SCV of the sample mean is calculated by sample standard deviation divided by square root of number of observations. In Panel B, SCV of the first moment given by the FF3 model is calculated by  $\text{cov}(\hat{\mu}_i, \hat{\mu}_j) = b_i' \text{cov}(\bar{f}) b_j + \text{tr}[\text{cov}(\bar{f}) \text{cov}(b_i, b_j)]$ .

Book to Market Equity (BE/ME) Quintiles										
Size	Low	2	3	4	High	Low	2	3	4	High
Panel A: No Margin Requirement										
	Sampling model					FF 3 factor model				
Small	-680.56	152.47	57.61	715.13	185.80	-53.28	15.18	11.53	50.55	64.80
2	-70.12	-113.49	268.07	4.92	-87.20	-57.05	-20.98	11.00	50.44	48.78
3	-189.48	-9.53	-221.22	-31.59	126.22	-32.05	-21.68	-1.86	14.40	44.40
4	555.06	-474.59	-124.96	221.51	-44.74	28.37	-29.38	-16.49	39.19	19.45
Big	129.80	-52.32	183.51	-21.10	-116.63	-71.82	-25.22	17.18	50.96	46.46
Cash			-262.57					-82.88		
Panel B: 50% Margin Requirement										
	Sampling model					FF 3 factor model				
Small	-51.03	0.06	-0.04	16.16	124.56	-0.03	0.00	0.00	0.00	21.19
2	0.02	0.00	0.00	0.44	0.09	-7.49	0.00	0.00	0.01	37.33
3	-0.15	0.12	0.07	0.35	6.04	-25.43	0.00	0.00	0.01	48.32
4	0.03	0.01	0.14	0.14	0.40	-0.30	0.00	0.00	0.00	23.03
Big	0.02	0.01	0.07	0.00	0.04	-0.02	0.00	0.00	0.03	36.82
Cash			2.43					-33.46		

Table 8: Optimal portfolio allocations

The table reports the optimal allocations per \$100 of wealth for a mean variance investor with coefficient of absolute risk aversion equal to 2.27. The optimal portfolio is calculated assuming the moments given by the model are true parameters. The optimal portfolio without margin requirement is  $w = \frac{1}{2.27} \Sigma^{-1} \mu$ . With 50% margin requirement, the optimal portfolio is calculated by solving a nonlinear constrained maximizing problem. The constraint used is  $\sum_{i=1}^{25} |w_i| \leq 2$ . Sample period is from 1963:07 to 2001:12.

	No Margin Requirement		50% Margin Requirement	
	Sample	FF 3 factor	Sample	FF 3 factor
CE(optimal)	80.61	15.90	14.38	11.06
CE loss( $\hat{\mu}, \hat{\Sigma}$ )	22.53	4.02	1.07	0.99
CE loss( $\hat{\mu}$ )	15.29	1.79	0.99	0.85
CE loss( $\hat{\Sigma}$ )	4.78	1.64	0.12	0.13

Table 9: Certainty equivalent comparison in the 25 portfolios case

The table reports the maximized certainty equivalent (in % per annum) at the optimal portfolio and the certainty equivalent losses due to parameter uncertainty under a no margin requirement and a 50% margin requirement. The risk aversion coefficient corresponds to 2.27. Sample assumes sample moments are the true parameters. The FF 3 factor model assumes the moments implied by the FF 3 factor model are the true parameters. CE(optimal) is the maximized certainty equivalent assuming the moment estimates given by each model are the true parameters. CE loss( $\hat{\mu}, \hat{\Sigma}$ ) measures the expected certainty equivalent loss due to uncertainty in the first two moments. CE loss( $\hat{\mu}$ ) measures the loss due to the uncertainty in the first moments alone, assuming the second moments are known. CE loss( $\hat{\Sigma}$ ) measures the loss due to the uncertainty in the second moments alone. 10,000 simulations are used to obtain the results for the no margin requirement case and 5,000 simulations are used for the 50% margin requirement case.

3 factor model, then the SCVs of the first moments are about 10 times larger than those of the second moments. Thus the variances are more accurately measured than the expected return. Based on the SCV comparison, one might predict that the certainty equivalent loss will be mainly due to the estimation risk in the expected return. Since the  $N$  asset case involves lots of interactions between the assets through covariances, however, comparing the SCVs of the means with those of the variances, without the covariances, will be misleading.

Table 8 reports the optimal allocations per \$100 of wealth for a mean variance investor with a coefficient of absolute risk aversion equal to 2.27. I consider two scenarios on the weight constraints: a no margin requirement and a 50% margin requirement. First, if there is no margin requirement, then the optimal portfolio includes large long and short positions. Table 9 reports the certainty equivalent and the losses due to estimating the first two moments. If the sample moments are the true parameters, then the maximum certainty equivalent is equal to 80.61% per annum. In order to calculate the losses, I obtain a distribution of weights from 10,000 simulations. The loss due to parameter uncertainty is 22.53%, roughly one fourth of the optimal certainty equivalent.<sup>14</sup> The loss due to the first moment alone is

<sup>14</sup>In this analysis the investor uses the sample moments to calculate the portfolio which is inferior to a Bayesian approach. In subsection 3.4, the Bayesian approach will be used to account for estimation risk.

15.29%.<sup>15</sup> From Table 7 Panel A, the average SCV of the sample means is sixteen times larger than that of the sample variances. A simple comparison of SCV between the first two moments would suggest the loss due to the uncertainty in second moments should be sixteen times smaller than that due to the first moments. Nevertheless, the loss due to the second moments alone is 4.78%, roughly a quarter of the total loss. Even though the variances are sixteen times more accurately estimated than the expected returns, the actual loss due to the sample covariance matrix is significant.

Results are more significant for the FF 3 factor model. If the moments given by the FF 3 factor model are the true moments, then the expected return and variance of returns are given in Table 6 Panel B. The optimal portfolio in Table 8 Panel B under a no margin requirement scenario also involves long and short positions but it is not as severe as the sampling model. The maximum certainty equivalent at the optimal portfolio is equal to 15.90%. The entire loss is 4.02% and it breaks more or less evenly between the loss due to the first moments and the loss due to the second moments. Contrary to the sampling model, the loss due to the second moment is as large as the loss due to the first moment. Even though the second moments are, on average, ten times more accurately estimated than the first moments, the actual losses are about the same, implying that a simple SCV comparison underestimates the significance of the uncertainty in the second moments.

The reason for the significant loss due to the uncertainty in the second moments is as follows. In a multiple asset case, the uncertainties in the second moments can be compounded to have a large aggregate effect on the loss. Intuitively, the first moment is a vector of 25 expected excess returns but the second moment is a 25 by 25 matrix of covariances. Thus even though the covariances are, individually, estimated more accurately than the expected returns, they can have a compounding effect on the loss. One evidence for this argument can be found in a constrained problem, where the optimal portfolio consists of only a few number of assets.

The Regulation T imposes almost all individual and institutional investors to require a 50% margin. In other words, the ratio between the total position size and the invested capital cannot be greater than two. The unconstrained optimal portfolio using the sample moments takes position of \$4,838, which is 48 times the invested capital. Thus such a portfolio is unrealistic under the 50% margin requirement. Instead, the investor must solve

---

<sup>15</sup>The sum of the losses due to the first and second moments does not sum up to the total loss. The decomposition in Eq. (19) was an approximation and the exact equality may not hold for the multiple asset case.



the following constrained maximization problem

$$\begin{aligned} \max_w \quad & w' \mu - \frac{1}{2} A w' \Sigma w \\ \text{s.t.} \quad & \sum_{i=1}^{25} |w_i| \leq 2 \end{aligned} \tag{20}$$

The optimal portfolio under the sampling model takes a long position in the small value portfolio and shorts the small growth portfolio. Basically, the optimal portfolio takes a long position in the highest Sharpe ratio portfolio and a short position in the lowest Sharpe ratio portfolio. From Table 9, the maximized certainty equivalent at the optimum is equal to 14.38%. For the constrained problem, the losses are calculated from 5,000 simulations. The loss is only 1.07%, only one thirteenth of the maximum certainty equivalent. The reduced loss under the margin requirement is well documented in Pastor and Stambaugh (2000). With a 50% margin requirement, the loss is mostly due to the uncertainty in first moments, which is 0.99%, roughly eight times larger than the loss due to the second moment. The FF 3 factor model gives a similar result for the loss comparison. The optimal portfolio under the FF 3 factor model takes long positions in five value portfolios. The maximum certainty equivalent is 11.06% and the loss due to estimation is 0.99%. The loss is divided by 0.85% and 0.13% among the first and the second moment uncertainties, respectively. This decomposition is similar to what a simple SCV comparison would suggest. The weighted average of SCVs of the first moments is 0.1340, where the weighted average of SCVs of the second moments is 0.0176.<sup>16</sup> A simple comparison would suggest that the loss due to the first moments should be 7.59 times larger than the loss by the second moments. In fact, the loss due to the first moment is 6.40 times larger than the loss due to the second moment. In the 50% margin requirement scenario, the SCV comparison correctly predicts the losses.

The reason why the SCV comparison predicts well for the case with a 50% margin requirement is due to the lack of interactions between the 25 portfolios. Since the optimal portfolio takes positions in only a few number of assets, the compounding effect caused by the covariances is limited. For example, in the sampling model, the investor takes positions in only two portfolios and thus one can regard it as a 2 risky asset case. Thus a simple comparison between the SCVs of the first two moments will be informative to gauge the losses due to the first two moments. In short, the smaller number of assets the investor holds, the smaller loss he suffers from the uncertainty in second moments.

The above analysis has surprising implications for investors. If an investor takes positions in many stocks, then the uncertainty in second moments can be compounded and results in

---

<sup>16</sup>The weights used for the calculation are  $\frac{|w_i|}{2}$ .

a large certainty equivalent loss to the investor as in the no margin requirement scenario. Assuming the FF 3 factor model, the optimal portfolio takes positions in all 25 portfolios. Even though the accuracy of second moments is 10 times greater than that of the first moments, the aggregate loss due to the second moments is as large as the loss due to the first moments. Instead, if the investor only takes a position in a small number of portfolios, then the loss due to the second moments is negligible as in the 50% margin requirement scenario. The loss due to the second moments decreases as the number of assets in the portfolio gets smaller.

Traditionally, it has been advocated to diversify the portfolio to reduce the risk. This analysis, however, suggests that an adverse effect of diversification can occur. Intuitively, as the number of assets increases, there are more parameters the investor needs to estimate to calculate the portfolio weights and hence he is more susceptible to estimation risk. This can be one possible explanation to why investors tend to hold a small number of stocks. Furthermore, they may try to reduce the estimation risk by investing in the assets that they are familiar with. Specifically, people tend to invest in a business that they are associated with or a company that is geographically close to them. This phenomenon has been known as the "home bias puzzle" and attempts to explain this consider transaction costs or hedging needs.<sup>17</sup> The above analysis considering the estimation risk in second moments provides another explanation to the home bias puzzle.

On the other hand, the loss due to the first moments is also affected by the number of portfolios, but the effect is small. Specifically, the loss due to the first moment in the no margin requirement scenario is 1.79% versus 0.85% in the 50% margin requirement scenario. The small reduction can be explained by the smaller investment. With no margin requirement, the investor takes a total position of \$842.51, but with a 50% requirement, the total size of the position is \$200. Since he invests less under the 50% margin requirement scenario, it is natural to expect the loss to decrease. I argue that the compounding effect on the loss is more significant in the second moment than in the first moment.

The uncertainty in the second moments in the above portfolio analysis is measured by the GMM estimator, which is robust to misspecification. In subsection 2.3, I find the GMM estimates are smaller than the GARCH(1,1) estimates given by Eq. (5). If one uses the variance given by Eq. (5) to generate the second moments, then one can expect the variability in the weight to increase. This implies that the loss due to the second moments will become

---

<sup>17</sup>For more on home bias, see Kang and Stulz (1997), Coval and Moskowitz (1999), Grinblatt and Keloharju (2001), and Huberman (2001).

larger. If one replaces the diagonal terms in the covariance matrix of variances with the values estimated by Eq. (5) using the FF 3 factor model, the loss due to parameter uncertainty becomes 1.57% under the 50% margin requirement. The loss due to first moment stays the same at 0.85%, but the loss due to the second moment increases to 0.82%. If one imposes a GARCH(1,1) structure on returns, then the loss due to second moment becomes as large as the loss due to the first moment even in the case of a 50% margin requirement.

### 3.4 Bayesian Approach

Subsection 3.3 uses the classical method to compute the moments. For example, the sample moments are used to compute the portfolio under the sampling model. Instead, one can use a Bayesian approach to estimate the moments to compute the portfolio. I will show that the Bayesian approach performs better than the classical approach, nevertheless, the losses due to the parameter uncertainty are still significant, and the general conclusions remain the same as in subsection 3.3.

Ever since Zellner and Chetty (1965), Klein and Bawa (1976) and Brown (1979) introduced a Bayesian approach to deal with the portfolio choice under parameter uncertainty, it has been used as the main framework. Brown (1979) shows the Bayesian approach with a noninformative prior performs better than the classical approach using the sample moments.<sup>18</sup> Specifically, if one assumes noninformative priors for the first two moments and assumes returns are drawn from a multivariate normal distribution, then the posterior distribution for the first moment is a normal distribution and the second moment follows an inverse Wishart distribution. Hence the moments from the predictive distribution can be solved analytically. Let  $\mu^*$  and  $\Sigma^*$  denote the first and second moments implied by the predictive distribution,  $p(r_{T+1}|r)$ . Brown (1979) showed that

$$\mu^* = \hat{\mu}, \quad \Sigma^* = \frac{(T-1)(T+1)}{T(T-N-2)} \hat{\Sigma}, \quad (21)$$

where  $\hat{\mu}$  and  $\hat{\Sigma}$  are the sample mean return and sample covariance matrix, respectively and  $N$  is the number of assets. See the Appendix for derivation. Then the investor can form a portfolio using  $\mu^*$  and  $\Sigma^*$  in (21).

$$\hat{w}_s^* = \frac{1}{A} \Sigma^{*-1} \hat{\mu}^* = \kappa \times \hat{w}_s, \quad (22)$$

$$\text{where } \kappa = \frac{T(T-N-2)}{(T-1)(T+1)} < 1 \quad (23)$$

---

<sup>18</sup>Frost and Savarino (1986) argues that using an informative prior will shrink the estimates toward the grand mean and the performance dominates that using a noninformative prior.

This implies the relative weights among  $N$  risky assets are not affected but the total amount to be invested in the risky assets will be reduced by a factor of  $\kappa = \frac{T(T-N-2)}{(T-1)(T+1)} = 0.9413$ . The investor using the Bayesian approach, will invest 5.87% less than the classical investor who uses the sample moments to compute his portfolio. Since Eq. (15) is derived by the asymptotic distribution theory, asymptotically, the certainty equivalent loss using the Bayesian estimates will be identical to the case using sample moments. In a finite sample, however, the certainty equivalent loss will be different. It can be shown that the difference in the loss is

$$CE_{s^*} - CE_s = (1 - \kappa) \left[ (1 + \kappa) \frac{1}{2} A \widehat{w}'_s \Sigma \widehat{w}_s - \widehat{w}'_s \mu \right]. \quad (24)$$

Following a similar approach in Pastor and Stambaugh (2000), one can apply the Bayesian approach for the Fama French 3 factor model as well.<sup>19</sup> The first moments implied by the predictive distribution is

$$\mu^* = \widehat{\mu} = \widehat{B}' \bar{f}, \quad (25)$$

where  $B$  is a 3 by 25 matrix of stacked betas on FF 3 factors and  $\bar{f}$  is a 3 by 1 vector of sample mean return for the three factors. Thus the first moments are identical to those given by the classical approach. The second moments, however, differ and the  $i$ th and  $j$ th element of  $\Sigma^*$  can be expressed as

$$\begin{aligned} \Sigma_{(i,j)}^* &= b'_i \Omega^* b_j + tr \left[ \Omega^* \text{cov}(b_i, b'_j | r) \right] + \tilde{\sigma}_{\varepsilon_i, \varepsilon_j} + \bar{f}' \text{cov}(b_i, b'_j | r) \bar{f}, \\ \Omega^* &= \text{cov}^*(f) = \frac{(T-1)(T+1)}{T(T-5)} \widehat{\text{cov}}(f), \\ \tilde{\sigma}_{\varepsilon_j, \varepsilon_j} &= \frac{T-3}{T-25-3} s_{\varepsilon_i, \varepsilon_j}, \end{aligned} \quad (26)$$

where  $\widehat{\text{cov}}(f)$  is the sample covariance of the three factors.<sup>20</sup>

Table 10 reports the certainty equivalent losses using the Bayesian approach.<sup>21,22</sup> Since the Bayesian approach results in the same expected return but increased covariance, the investor will be more cautious to invest in risky assets. In the no margin requirement scenario, the certainty equivalent loss is 19.39% compared with 22.53% when the investor

<sup>19</sup>The difference from Pastor and Stambaugh (2000) is that we force the intercept to be zero because this analysis doesn't involve the uncertainty in the correct asset pricing model.

<sup>20</sup>We take the limiting case where the degrees of freedom of the prior distribution of  $\Sigma$  go to zero and the scale matrix goes to infinite so that the prior of the second moment is essentially noninformative.

<sup>21</sup>The loss is not calculated from Eq. (15). Otherwise, it will always be smaller than before because the Bayesian weights are smaller than the classical weights. We adjust the loss in Table 9 by taking the difference between  $CE_s$  and  $CE_{s^*}$ .

<sup>22</sup>For a known mean,  $\Sigma^*$  is given by  $\frac{T-1}{T-N-2} \widehat{\Sigma}$  and for a known variance,  $\Sigma^*$  is given by  $\frac{T+1}{T} \widehat{\Sigma}$ .

	No Margin Requirement		50% Margin Requirement	
	Sample	FF 3 factor	Sample	FF 3 factor
CE(optimal)	80.61	15.90	14.38	11.06
CE loss( $\mu^*, \Sigma^*$ )	19.39	3.75	0.95	0.98
CE loss( $\mu^*$ )	15.20	1.78	0.89	0.85
CE loss( $\Sigma^*$ )	3.99	1.45	0.12	0.13

Table 10: Certainty equivalent using the Bayesian approach

The table reports the maximized certainty equivalent (in % per annum) at the optimal portfolio and the certainty equivalent losses due to parameter uncertainty under no margin requirement and a 50% margin requirement using the Bayesian approach. The risk aversion coefficient corresponds to 2.27. Sample assumes sample moments are the true parameters. FF 3 factor model assumes the moments implied by FF 3 factor model are the true parameters. CE(optimal) is the maximized certainty equivalent assuming the moment estimates given by each model are the true parameters. CE loss( $\hat{\mu}, \hat{\Sigma}$ ) measures the expected certainty equivalent loss due to uncertainty in the first two moments. CE loss( $\hat{\mu}$ ) measures the loss due to the uncertainty in the first moments alone, assuming the second moments are known. CE loss( $\hat{\Sigma}$ ) measures the loss due to the uncertainty in the second moments alone. 10,000 simulations are used to obtain the results for the no margin requirement case and 5,000 simulations are used for the 50% margin requirement case.

uses sample moments to update the portfolio. As suggested by Brown (1979), the Bayesian loss is smaller than before. Using a higher covariance given by the Bayesian approach, the investor maintains the same relative weights in 25 portfolios but reduces the total amount to be invested in the risky stocks. There are two opposite effects when the weight becomes lower. First, the expected return from the portfolio becomes lower due to a smaller investment. On the other hand, it will reduce the variability in the weights so the loss due to the variability decreases.<sup>23</sup> The results show that the second effect offsets the first effect so that the Bayesian investor incurs a smaller loss than a non-Bayesian investor, or a classical investor. The same results apply to the FF 3 factor model. Eq. (26) implies the covariance will be larger than the sample covariance. So the Bayesian investor will invest less in risky assets than otherwise. The positive effect again outweighs the negative effect to reduce the certainty equivalent loss. The loss due to the second moment for a Bayesian investor is 1.45%, which is 0.19% smaller than the loss for a classical investor, but still is economically significant.

With a 50% margin requirement, the Bayesian approach improves a little over the classical method. Nevertheless, the improvement is very small compared with the no margin

<sup>23</sup>If one uses Eq. (15), it will only consider the second effect and will report a smaller loss than the classical method.

requirement case. For example, in the FF 3 factor model, the Bayesian investor can reduce the entire loss by 0.01%, or 1 basis point per annum. This small improvement by the Bayesian method is due to the small investment restricted by the 50% margin requirement.

To summarize, as suggested by Brown (1979), a Bayesian investor, who accounts for estimation risk, achieves a higher certainty equivalent than a classical investor. Nevertheless, the losses due to the uncertainty in second moments are still economically significant as in the classical approach case.

## 4 Conclusion

This study shows that the uncertainty in second moments is larger than what people believe and also shows how this uncertainty can affect the certainty equivalent of a mean variance investor. A common misconception among finance theorists is that the second moments can be estimated accurately. Under the i.i.d. normality assumption, empirical returns data reveals that the sample variance is more accurate than the sample mean. The departures from an i.i.d. normality assumption, however, may turn the tables. As many empirical studies pointed out, asset returns do not follow an i.i.d. normality assumption. In particular, large kurtosis and time varying volatility are well known features that are too prominent to be ignored.

I have shown that the variance of a sample variance can be expressed in terms of the kurtosis and time varying volatility. Specifically, the excess kurtosis and the persistence in the volatility can be shown to have a direct effect on the precision of the sample variance. Empirically, second moments are less accurately measured than what one expect. Using the CRSP value weighted monthly excess returns from 1926 to 2001, I have shown that the sample variance is less accurately measured than the sample mean in terms of squared coefficient of variation (SCV), which allows one to compare the accuracy of the estimates in different units. Using more recent data from 1963 to 2001, the accuracy of sample variance improves. This is mainly due to a smaller kurtosis and smaller autocorrelations in the squared return after 1963. On the other hand, the accuracy of the first moment deteriorated in the subsample.

The imprecision of the second moments has economic significance in a portfolio allocation problem. I have presented a simple framework to compute the expected certainty equivalent loss to a mean variance investor and have shown that the loss depends on the uncertainty

in second moments. If the asset universe includes 25 size and BE/ME sorted portfolios, I have found that the loss due to the second moment is larger than what a simple SCV comparison would suggest. Assuming the Fama French 3 factor model is correct, the loss due to the uncertainty in second moments is as large as the loss due to the first moments, even though the second moments are 10 times more accurately estimated than the first moments. I argue that the interactions between the assets through covariances can explain this large aggregate effect of the uncertainty in second moments on the losses. One evidence that supports this argument can be found in a constrained portfolio problem, where the optimal portfolio consists of a small number of assets. In such a case, the loss due to the uncertainty in the second moments is negligible because interaction between the assets is limited. As the number of assets in the portfolio increases, the loss due to the uncertainty in second moments becomes larger. Hence in order to reduce the estimation risk, investors may not fully diversify the portfolio but own assets that they are familiar with. This offers an explanation to the home bias puzzle.

Asset returns in the last 20 years have been more volatile than in the 60's and 70's. Specifically, the kurtosis and volatility persistence are higher than before. This implies that the second moments will be less precisely estimated and hence the estimation risk in the second moments will increase. Thus if an investor uses the last 20 years of data to estimate the moments to compute a portfolio, then the loss due to the uncertainty in the second moments will become larger.

I have shown that the uncertainty in second moments is large, and hence it can reduce the certainty equivalent of an investor. As Merton (1980) suggested, if one uses higher frequency data to estimate the variance of monthly returns, one may be able to obtain a more accurate variance estimate and thus the loss will be smaller than when one uses the monthly return series to estimate the variance. Nevertheless, as Bai, Russell and Tiao (2001a) argued, the precision of variance estimates using higher frequency data is not as promising as Merton's ideal case. As high frequency data depart more from an i.i.d. normal distribution, the benefit of using more data can be negated by severe non-normality and volatility clustering features.

The portfolio allocation problems under parameter uncertainty have been extensively studied. Most literature on this issue focus on the uncertainty in first moments, such as return predictability and the correct asset pricing model. On the other hand, this paper focuses on the uncertainty in second moments and compares its effect on the losses with the effect of the first moments. The investor in question solves the portfolio problem in two stages. First, he assumes that he has full information on the parameters and solves for

the optimal portfolio to maximize his expected utility. Then he estimates the parameters from a sample and replaces the assumed parameters with the point estimates. In such a two stage problem, the uncertainty in the second moments may affect the certainty equivalent loss but it does not affect the portfolio decision rule because only the point estimates affect the portfolio. However, integrating the estimation problem with the portfolio problem will be a more realistic approach. Then the uncertainty in the estimates may have a direct effect on the decision rule. Brandt (1999) shows a way by developing a nonparametric approach to integrate the problem. Finally, time varying volatility models such as the GARCH model provide conditional variance forecasts. Nevertheless, the investor is assumed to use the unconditional variance to calculate the portfolio and he does not take advantage of the conditional variance. Naturally, predictability in volatility should be included in the portfolio allocation problem. Furthermore, there can be other interesting implications for a long horizon portfolio allocation problem under volatility predictability. These issues present directions for future research.



# Appendix

## A.1. Derivation of $\text{var}(s^2)$ in a GARCH(1,1) series

The derivation is from Bai, Russell and Tiao (2001b). Suppose  $e_t$  is weakly stationary with no autocorrelation so that  $\text{corr}(e_t, e_{t+l}) = 0$  for all  $l$ . Assume  $e_t^2$  is also weakly stationary but autocorrelation  $\rho_l = \text{corr}(e_t^2, e_{t+l}^2)$  exists for all  $l$ . Let  $y$  denote a vector of squared deviations,  $(e_1^2, e_2^2, \dots, e_T^2)'$ . Then sample variance  $s^2$  is  $\frac{1}{T-1}i_T'y$ .

The covariance matrix of  $y$  is

$$\text{cov}(y) = \text{var}(e_t^2) \begin{bmatrix} 1 & \rho_1 & \cdot & \cdot & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdot & \cdot \\ \cdot & \rho_1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \rho_1 \\ \rho_{T-1} & \cdot & \cdot & \rho_1 & 1 \end{bmatrix}. \quad (27)$$

Variance of  $s^2$  can be expressed as,

$$\begin{aligned} \text{var}(s^2) &= \frac{1}{(T-1)^2} i_T' \text{cov}(y) i_T \\ &= \frac{1}{(T-1)^2} [\text{var}(e_t^2)] [T + 2(T-1)\rho_1 + 2(T-2)\rho_2 + \dots + 2\rho_{T-1}] \\ &= \frac{T}{(T-1)^2} [\text{var}(e_t^2)] \left[ 1 + 2 \sum_{j=1}^T \left(1 - \frac{j}{T}\right) \rho_j \right] \\ &= \frac{T}{(T-1)^2} [E(e_t^2)^2 - (E(e_t^2))^2] \left[ 1 + 2 \sum_{j=1}^T \left(1 - \frac{j}{T}\right) \rho_j \right] \\ &= \frac{T}{(T-1)^2} (E(e_t^2))^2 \left[ \frac{E(e_t^4)}{(E(e_t^2))^2} - 1 \right] \left[ 1 + 2 \sum_{j=1}^T \left(1 - \frac{j}{T}\right) \rho_j \right] \\ &= \frac{T}{(T-1)^2} \left( \frac{T-1}{T} \sigma^2 \right)^2 [\gamma_e + 2] \left[ 1 + 2 \sum_{j=1}^T \left(1 - \frac{j}{T}\right) \rho_j \right] \end{aligned} \quad (28)$$

$$= \frac{2\sigma^4}{T} \left[ 1 + \frac{\gamma_e}{2} \right] \left[ 1 + 2 \sum_{j=1}^T \left(1 - \frac{j}{T}\right) \rho_j \right]. \quad (29)$$

The last term can be simplified, if one assumes  $e_t$  is a GARCH(1,1) series.

$$\begin{aligned} e_t &= \sqrt{h_t} z_t, \quad z_t \sim \text{i.i.d. } (0, 1) \\ h_t &= \omega + \alpha e_{t-1}^2 + \beta h_{t-1} \end{aligned} \quad (30)$$

The autocorrelations of  $e_t^2$  are well known.

$$\begin{aligned}\rho_j &= \rho(\alpha + \beta)^{j-1}, \quad j > 1 \\ \rho &= \alpha \left( 1 + \frac{\alpha\beta}{1 - \beta^2 - 2\alpha\beta} \right)\end{aligned}\quad (31)$$

Then the third term in Eq. (29) can be written as

$$\begin{aligned}1 + 2 \sum_{j=1}^T \left( 1 - \frac{j}{T} \right) \rho_j &= 1 + 2\rho \sum_{j=1}^T \left( 1 - \frac{j}{T} \right) (\alpha + \beta)^{j-1} \\ &= 1 + 2\rho \left( \frac{1 - (\alpha + \beta)^T}{1 - (\alpha + \beta)} - \frac{1 - (\alpha + \beta)^T - T(\alpha + \beta)^{T-1}(1 - \alpha - \beta)}{T(1 - \alpha - \beta)^2} \right).\end{aligned}\quad (32)$$

Since  $\alpha + \beta < 1$ , as  $T$  approaches infinity, it converges to,

$$\lim_{T \rightarrow \infty} 1 + 2 \sum_{j=1}^T \left( 1 - \frac{j}{T} \right) \rho_j = 1 + \frac{2\rho}{1 - \alpha - \beta}\quad (33)$$

Finally, the variance of  $s^2$  for the GARCH(1,1) series is

$$\text{var}(s^2) = \frac{2\sigma^4}{T} \left( 1 + \frac{\gamma}{2} \right) \left( 1 + \frac{2\rho}{1 - \alpha - \beta} \right).\quad (34)$$

## A.2. Derivation of Certainty Equivalent Loss

Here I derive the expected certainty equivalent loss to a mean variance investor who is facing parameter uncertainty. The mean variance investor solves the following portfolio problem.

$$\begin{aligned}\max_w \quad & w' \mu - \frac{1}{2} A w' \Sigma w, \\ \text{s.t.} \quad & g(w) \leq 0\end{aligned}\quad (35)$$

The objective function evaluated at the optimal portfolio can be regarded as the maximized certainty equivalent to the investor.

$$CE_o = w'_o \mu - \frac{1}{2} A w'_o \Sigma w_o\quad (36)$$

The suboptimal certainty equivalent is the objective function evaluated at the suboptimal portfolio computed by the estimates.

$$CE_s = \hat{w}'_s \mu - \frac{1}{2} A \hat{w}'_s \Sigma \hat{w}_s.\quad (37)$$

Assuming  $\widehat{w}_s$  is asymptotically unbiased, then

$$\begin{aligned}
E(CE_s) &= E\left[\widehat{w}'_s\mu - \frac{1}{2}A\widehat{w}'_s\Sigma\widehat{w}_s\right] \\
&= E(\widehat{w}_s)'\mu - \frac{1}{2}AE(\widehat{w}'_s\Sigma\widehat{w}_s) \\
&\approx w'_o\mu - \frac{1}{2}AE(\widehat{w}'_s\Sigma\widehat{w}_s).
\end{aligned} \tag{38}$$

Now one can subtract Eq. (38) from Eq. (36) to calculate the expected certainty equivalent loss.

$$\begin{aligned}
CE_o - E(CE_s) &\approx \frac{1}{2}AE(\widehat{w}'_s\Sigma\widehat{w}_s) - \frac{1}{2}Aw'_o\Sigma w_o \\
&= \frac{1}{2}AE[\text{tr}(\widehat{w}'_s\Sigma\widehat{w}_s)] - \frac{1}{2}Aw'_o\Sigma w_o \\
&= \frac{1}{2}AE[\text{tr}(\widehat{w}_s\widehat{w}'_s\Sigma)] - \frac{1}{2}Aw'_o\Sigma w_o \\
&= \frac{1}{2}A \times \text{tr}[E(\widehat{w}_s\widehat{w}'_s)\Sigma] - \frac{1}{2}Aw'_o\Sigma w_o \\
&= \frac{1}{2}A \times \text{tr}[\text{cov}(\widehat{w}_s)\Sigma] + \frac{1}{2}A \times \text{tr}[E(\widehat{w}_s)E(\widehat{w}_s)'\Sigma] - \frac{1}{2}Aw'_o\Sigma w_o \\
&\approx \frac{1}{2}A \times \text{tr}[\text{cov}(\widehat{w}_s)\Sigma] + \frac{1}{2}A \times \text{tr}[w'_o\Sigma w_o] - \frac{1}{2}Aw'_o\Sigma w_o \\
&= \frac{1}{2}A \times \text{tr}[\text{cov}(\widehat{w}_s)\Sigma]
\end{aligned} \tag{39}$$

### A.3. First Two Moments given by the Bayesian Approach

The derivation is from Brown (1979). Assume that  $r_t$  is an  $N$  by 1 vector of excess return drawn from a multivariate normal distribution with unknown parameters  $\mu$  and  $\Sigma$ . Let  $p(\mu, \Sigma)$  denote the prior distribution of the first two moments. Then the investor forms a posterior distribution,  $p(\mu, \Sigma|r)$  and forms a predictive distribution for  $r_{T+1}$ ,  $p(r_{T+1}|r)$ .

$$p(r_{T+1}|r) = \int_{\mu} \int_{\Sigma} p(r_{T+1}|r, \mu, \Sigma)p(\mu, \Sigma|r)d\mu d\Sigma \tag{40}$$

The investor replaces  $\mu$  and  $\Sigma$  in Eq. (11) with the moments given by Eq. (40). Let  $\widehat{\mu}$  and  $\widehat{\Sigma}$  denote the sample mean and sample covariance of returns. The likelihood function is

$$p(r|\mu, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\text{tr}\left\{\Sigma^{-1}(r - i_T\mu)'\right\}\right]. \tag{41}$$

If one uses the following multivariate Jeffreys prior density,

$$p(\mu, \Sigma) \propto |\Sigma|^{-\frac{N+1}{2}}, \tag{42}$$

then the posterior distribution of the first two moments is

$$\begin{aligned} p(\mu, \Sigma|r) &\propto p(r|\mu, \Sigma)p(\mu, \Sigma) \\ &\propto |\Sigma|^{-\frac{T+N+1}{2}} \exp \left[ -\frac{1}{2}tr \{ \Sigma^{-1} (r - i_T \mu')' (r - i_T \mu') \} \right]. \end{aligned} \quad (43)$$

It follows that

$$\Sigma|r \sim \text{Inv-Wishart}_{T-1}((T-1)\widehat{\Sigma}) \quad (44)$$

$$\mu|\Sigma, r \sim N(\widehat{\mu}, \frac{\Sigma}{T}). \quad (45)$$

The marginal posterior distribution of  $\mu$  can be shown to be a multivariate student  $t$  distribution.

$$\mu|r \sim t_{T-N} \left( \widehat{\mu}, \frac{(T-1)\widehat{\Sigma}}{T(T-N)} \right), \quad (46)$$

where  $\widehat{\mu}$ , is the location parameter and  $\frac{(T-1)\widehat{\Sigma}}{T(T-N)}$  is the scale parameter. Therefore, the posterior means of the first two moments, denoted by  $\widetilde{\mu}$  and  $\widetilde{\Sigma}$ , become

$$\begin{aligned} \widetilde{\mu} &= E(\mu|r) = \widehat{\mu}, \\ \widetilde{\Sigma} &= E(\Sigma|r) = \frac{(T-1)}{T-N-2} \widehat{\Sigma}. \end{aligned} \quad (47)$$

One also needs the posterior variance of the  $\mu$ .

$$\text{var}(\mu|r) = \frac{T-1}{T(T-N-2)} \widehat{\Sigma} \quad (48)$$

Finally, one can compute the first two moments implied by predictive distribution in Eq. (40). Let  $\mu^*$  and  $\Sigma^*$  be the mean and variance of the excess return from the predictive distribution.

$$\mu^* = E(r_{T+1}|r) = E[E(r_{T+1}|\mu, \Sigma, r)|r] = E(\mu|r) = \widetilde{\mu} = \widehat{\mu}, \quad (49)$$

$$\begin{aligned} \Sigma^* &= \text{var}(r_{T+1}|r) = E(\Sigma|r) + \text{var}(\mu|r) \\ &= \frac{(T-1)}{T-N-2} \widehat{\Sigma} + \frac{T-1}{T(T-N-2)} \widehat{\Sigma} = \frac{(T-1)(T+1)}{T(T-N-2)} \widehat{\Sigma} \end{aligned} \quad (50)$$

Law of iterated expectations and variance decomposition rule are applied in the last equation.

If the first moment is known, then

$$\mu^* = \widehat{\mu}, \Sigma^* = E(\Sigma|r) = \frac{(T-1)}{T-N-2} \widehat{\Sigma}. \quad (51)$$

If the second moment is known, then

$$\mu^* = \widehat{\mu}, \Sigma^* = E(\Sigma|r) + \text{var}(\mu|r) = \Sigma + \frac{1}{T} \Sigma = \frac{T+1}{T} \Sigma. \quad (52)$$

Note that even though the second moment is known,  $\Sigma^*$  is not equal to  $\Sigma$  because the estimation risk of the first moment is embedded in the second moment estimate.

## References

- Andrews, D. W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817-858.
- Bai, X., J. Russell, and G. C. Tiao, 2001a, Beyond Merton's utopia (i), Working paper, University of Chicago.
- Bai, X., J. Russell, and G. C. Tiao, 2001b, Effects of kurtosis and time varying volatility on the precision of variance estimates, Working paper, University of Chicago.
- Bai, X., J. Russell, and G. C. Tiao, 2002, Kurtosis of GARCH and stochastic volatility models with non-normal innovations, *Journal of Econometrics*, forthcoming.
- Barberis, N. C., 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225-264.
- Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307-327.
- Bollerslev, T., 1987, A conditionally heteroskedastic time series models for speculative prices and rates of return, *Review of Economics and Statistics* 69, 542-547.
- Black, F., E. Derman, and W. Toy, 1990, A one-factor model of interest rates and its application to treasure bond options, *Financial Analysts Journal* (January-February 1990): 33-39.
- Black, F., and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637-654.
- Brandt, M. W., 1999, Estimating portfolio and consumption choice: A conditional Euler equations approach, *Journal of Finance* 54, 1609-1645.
- Brown, S. J., 1979, Optimal portfolio choice under uncertainty: a Bayesian approach. In Bawa, V. S., S. J. Brown, and R. W. Klein (Eds.), *Estimation Risk and Optimal Portfolio Choice*, North Holland, Amsterdam.
- Cho, D. D., 2002, Sample properties of the fourth moment estimates, Working paper, University of Chicago.
- Cochrane, J., 2001, *Asset Pricing*, Princeton NJ, Princeton University Press.

- Coval, J. D., and T. J. Moskowitz, 1999, Home bias at home: Local equity preference in domestic portfolios, *Journal of Finance* 54, 1-39.
- Engle, R., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, 987-1008.
- Fama, E. F. 1965, The behavior of stock market prices, *Journal of Business* 38, 34-105.
- Fama, E. F., K. R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.
- Frost, P. A., J. E. Savarino, 1986, An empirical Bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* 21, 293-305.
- Grinblatt, M. and M. Keloharju, 2001, Distance, language, and culture bias: The role of investor, *Journal of Finance* 56, 1053-1073.
- Hansen, L. P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029-1954.
- Ho, T., and S. B. Lee, 1986, Term structure movements and pricing interest rate contingent claims, *Journal of Finance* 41, 1011-1029.
- Huberman, G., 2001, Familiarity breeds investment, *Review of Financial Studies* 14, 659-680.
- Kandel, S., R. F. Stambaugh, 1996, On the predictability of stock returns: an asset allocation perspective, *Journal of Finance* 51, 385-424.
- Kang, J., and R. Stulz, 1997, Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan, *Journal of Financial Economics* 46, 3-28.
- Klein, R. W. and V. S. Bawa, 1976, The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics* 3, 215-231.
- Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13-37.
- Mandelbrot, 1963, The variation of certain speculative prices, *Journal of Business* 36, 394-419.
- Merton, R. C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867-887.

- Merton, R. C., 1980, On estimating the expected return on the market, *Journal of Financial Economics* 8, 323-361.
- Newey, W. K., and K. D. West, 1987, A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.
- Newey, W. K., and K. D. West, 1992, Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* 61, 631-653.
- Pastor, L., R.F. Stambaugh, 2000. Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335-381.
- Ross, S. A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341-360.
- Sharpe, W. F., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425-442.
- Zellner, A., and V. K., Chetty, 1965, Prediction and decision problems in regression models from the Bayesian point of view, *Journal of the American Statistical Association* 60, 608-616.