Second Order Risk

Peter Shepard

MSCI Barra[*]

August 17, 2009

Abstract

Managing a portfolio to a risk model can tilt the portfolio toward weaknesses of the model. As a result, the optimized portfolio acquires downside exposure to uncertainty in the model itself, what we call "second order risk." We propose a risk measure that accounts for this bias. Studies of real portfolios, in asset-by-asset and factor model contexts, demonstrate that second order risk contributes significantly to realized volatility, and that the proposed measure accurately forecasts the out-of-sample behavior of optimized portfolios.

[*]The author is vice president and senior researcher at MSCI Barra; 2100 Milvia Street; Berkeley, CA 94704; peter.shepard@mscibarra.com.

1 Introduction

Classical finance assumes the markets to be like a game of chance: Although future events are uncertain, the distribution of these events is known. We cannot predict how the dice will land, but we can calculate the odds of any given outcome with certainty. We can expect to roll snake-eyes on average one time in 36, and the rules of the game do not change without warning.

Unfortunately, real financial markets do not behave like a game of chance: Market volatility is itself volatile; hot industries come and go; new companies are listed and others merge or go bankrupt. Under even the most generous assumptions, our estimates of financial risk are uncertain, based on limited historical observation, extrapolated forward.

For a passively invested portfolio, the effect of such uncertainty is as likely to be good or bad. The total risk may be overforecast or underforecast, but taken on average these errors tend to wash out. On the contrary, an optimized portfolio is more likely to be hurt by uncertainty than helped by it. Constructing portfolios to minimize risk can make them safer, but at the cost of introducing an asymmetric exposure to "second order risk."

In this paper, we explore a framework to quantify and forecast second order risk. Exploring only its mildest sources, we demonstrate that the act of optimizing a portfolio to a risk measure can render that measure systematically inaccurate. However, rather than abandon risk measurement or ignore its uncertainties, the framework shows that we may begin to account for second order risk as we do more familiar sources of uncertainty.

To quote a former US Secretary of Defense:

"There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know."

-Donald Rumsfeld, February 12, 2002 [1]

Our aim is to bring some of the latter into the category of "known unknowns." Correcting for these uncertainties in general leads to a more conservative view of risk. However, there will always be weaknesses in our models and much we cannot anticipate [2]. Perhaps a fourth category of "unknown knowns" is the most dangerous: things we think we know, but don't.

1.1   A Toy Model

To see the cause of second order risk, and how it can be forecast, consider the following toy example: Between two assets with the same expected return, an active manager aims to minimize risk by investing in the asset with the smaller standard deviation. In this example, investors are constrained to hold a single asset. After observing the returns of the assets, the manager finds Asset 1 to have a standard deviation of 8%, while Asset 2 has a standard deviation of 11%. Placing a bet on Asset 1, the active manager believes the portfolio to have a risk of 8%.

Although the manager doesn't know it, the returns of both assets are drawn from the same distribution, with standard deviation of 10%. The true risk is 10%, regardless of which asset was chosen, but *the active manager's strategy is more likely to make investments whose risk happens to be underforecast.* Meanwhile, passive investors are just as likely to hold either asset. Looking at the same data, a passive investor holding Asset 1 would underforecast risk, while an investor holding Asset 2 would overforecast risk, but with no bias toward either outcome.

Figure 1 shows the result of repeating this experiment many times. In each trial, two time series are drawn from the same distribution and risk estimates are made. The active manager bets on the asset with lower risk forecast, while the passive investor always holds Asset 1. Noise diminishes the accuracy of both investors' risk forecasts, but it systematically biases only the active manager, whose average forecast is 8.7%, less than the true 10%. The wise active manager would correct risk forecasts upward, to compensate for the bias introduced by active management. Although the active manager does not know the true distribution of returns, we will see that it is possible to compensate for this bias.

An intriguing implication is that the best risk forecast depends not just on the portfolio holdings, but also on the *strategy.* In the simulation, the two managers hold identical portfolios in half of the trials, and forecast risk based on identical returns. Nonetheless, because of differences of strategy, they have reason to make different risk forecasts, even when their portfolios exactly coincide.

2   Model Uncertainty

The example above is a case of aiming to maximize a utility function $U(\mathbf{w})$ for which we have only an approximate model $\hat{U}(\mathbf{w})$. With perfect information, we would choose the variables $\mathbf{w} = \mathbf{w}^*$ to maximize $U(\mathbf{w})$, but instead we must choose some other $\hat{\mathbf{w}}$, the best guess given what is known. The difference between $U(\mathbf{w})$ and $\hat{U}(\mathbf{w})$ leads to some discrepancy between the true best $\mathbf{w}^*$, and the best guess $\hat{\mathbf{w}}$ given the available information.
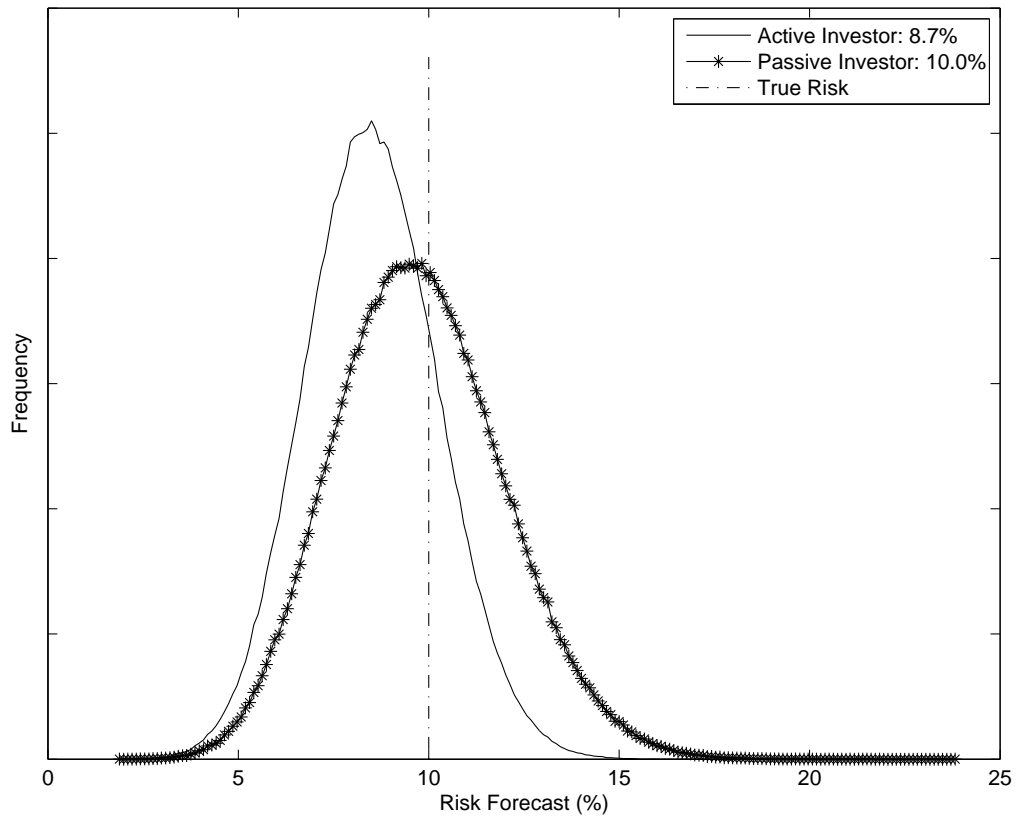
Figure 1: The distribution of risk forecasts of a toy model of active and passive investors. After observing the returns of two assets for ten periods, the active manager selects the asset with lower sample standard deviation, while a passive investor is equally likely to hold either asset. Although the true risk is 10% in all cases, the active manager consistently underforecasts risk.

3

As depicted in Figure 2, the effect of such a discrepancy is generically a loss: any departure $\Delta\mathbf{w}$ from $\mathbf{w}^*$ reduces utility. For small errors $\Delta\mathbf{w}$, the utility of $\hat{\mathbf{w}}$ can be approximated

$$U[\hat{\mathbf{w}}] = U[\mathbf{w}^*] + \Delta U \simeq U[\mathbf{w}^*] + \Delta\mathbf{w}'\mathbf{H}\Delta\mathbf{w}, \tag{1}$$

where $\mathbf{H}$ is the Hessian of $U(\mathbf{w})$ at $\mathbf{w}^*$, the matrix of second derivatives. Simply because any function is concave at its maximum, $\mathbf{H}$ is a negative-definite matrix, and the correction $\Delta\mathbf{w}'\mathbf{H}\Delta\mathbf{w}$ is negative for any $\Delta\mathbf{w} \neq 0$.
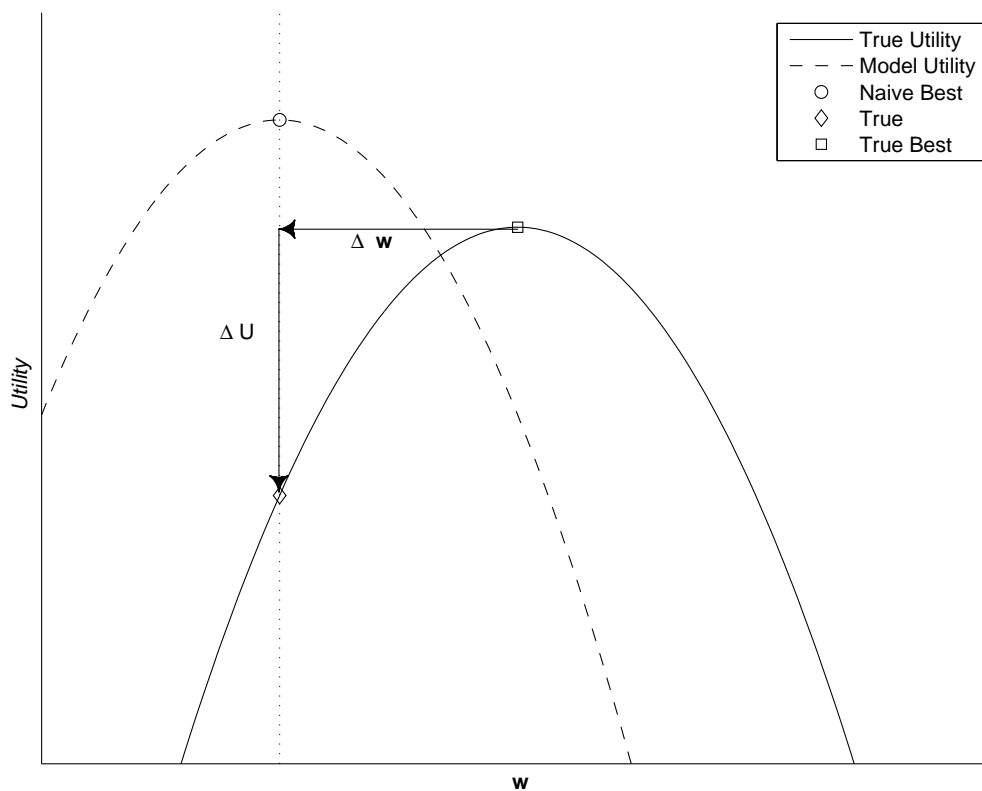

Figure 2: The schematic effect of errors in the model utility function. A model utility function usually has its maximum at least slightly removed from the true maximum, $\Delta\hat{\mathbf{w}}$. Though the realized $\mathbf{w}$ appears optimal to the model, it incurs a penalty ΔU under the true utility function.

Using the model $\hat{U}(\mathbf{w})$ to forecast the utility of $\hat{\mathbf{w}}$ misses the penalty $\Delta\mathbf{w}'\mathbf{H}\Delta\mathbf{w}$ that is the

4

inevitable side-effect of having only an approximation. If we can calculate a distribution of $\Delta\mathbf{w}$, we can account for the average loss $\sim E(\Delta\mathbf{w}'\mathbf{H}\Delta\mathbf{w})$ that arises due to uncertainty in $\mathbf{w}^*$.

For the utility functions of finance, these errors are compounded by a tendency for models $\hat{U}(\mathbf{w})$ to appear intrinsically better than the true $U(\mathbf{w})$.[1] As a result, the point labeled Naive Best in Figure 2 is above even the True Best, attainable with perfect information utility U, and well above the True utility of $\hat{\mathbf{w}}$.

2.1 Uncertainty and Active Management

Classical portfolio theory [3] instructs the portfolio manager to build the optimal portfolio $\mathbf{w}$ from the covariance matrix $\mathbf{\Omega}$ and vector α of expected excess returns. A variety of utility functions may be used, among them the Sharpe ratio

$$U(\mathbf{w}) = \frac{\mathbf{w}'\alpha}{\sqrt{\mathbf{w}'\mathbf{\Omega}\mathbf{w}}}. \tag{2}$$

In the absence of constraints, the portfolio maximizing (2) has weights proportional to

$$\mathbf{w}^* = \mathbf{\Omega}^{-1}\alpha. \tag{3}$$

However, even assuming the markets to be stationary and Gaussian, the covariance matrix $\mathbf{\Omega}$ must be estimated from observation of historical behavior, which introduces noise.[2]

Even if this noise level can be made relatively small, so that each element of $\mathbf{\Omega}$ is known with relative certainty, optimization tends to align the portfolio with the noise [4], compounding many small errors into a large effect. As the number of observations T increases, the noise tends to be reduced by $\sim 1/T$, and a good estimator can insure that these errors average to zero.

The impact of this small amount of noise is nonetheless significant. For $\mathbf{\Omega}$ estimated directly from N assets, we will see that the effect of noise on the optimized portfolio does not average to zero, but yields corrections of order

$$\frac{1}{(1 - N/T)},$$

---

[1]Physicists may recognize a relation to the tendency of quantum mechanical perturbations to systematically lower the ground state energy.

[2]Another source of noise, uncertainty in α, may also be significant. This subjective uncertainty could be incorporated into this framework, but we concentrate upon uncertainty in $\mathbf{\Omega}$, taking α to be known to the investor.

growing without bound as the number of assets approaches the number of observations. Since T is limited by changing dynamics and market microstructure, this can lead to significantly inaccurate risk forecasts, and diminished out-of-sample performance.

For a factor model of risk [5], N/T is replaced by the milder K/T, where K is the number of systematic risk factors. This makes portfolio optimization among many assets more robust to estimation errors, but may leave significant corrections to risk forecasts.

2.2   A Second Order Risk Measure

The denominator of the Sharpe ratio (2) is the standard deviation Σ of future portfolio returns, a common measure of portfolio risk:

$$\Sigma^2 = E_{\mathbf{r}}\left((\mathbf{w'r} - \mathbf{w'}\bar{\mathbf{r}})^2|\mathbf{\Omega}\right) \;\; = \;\; \mathbf{w'\Omega w}. \tag{4}$$

Here $E_x(f(x)|y)$ denotes an average over the variable x, conditional on y.[3] If the true covariance matrix $\mathbf{\Omega}$ were known, (4) would be a good measure of uncertainty, but in practice we must make do with an estimate $\hat{\mathbf{\Omega}}$ of the true distribution, based on observation.

Relative to the hypothetical true covariance matrix, the estimate $\hat{\mathbf{\Omega}}$ is a random variable. With $\hat{\mathbf{\Omega}}$ used in place of $\mathbf{\Omega}$ in Equation (2), the optimized portfolio

$$\hat{\mathbf{w}}(\hat{\mathbf{\Omega}}) = \hat{\mathbf{\Omega}}^{-1}\alpha, \tag{5}$$

is also a random variable, distributed about the true optimal portfolio $\mathbf{w}^*$.

The risk of $\hat{\mathbf{w}}(\hat{\mathbf{\Omega}})$ therefore arises from two contributions. In addition to the usual uncertainty of future returns $\mathbf{r}$, there is a second risk associated with the randomness of the observation $\hat{\mathbf{\Omega}}$ about $\mathbf{\Omega}$, which is typically neglected.

To account for the latter uncertainty, we define a risk measure by extending the expectation value of Equation (4) to average over both ensembles:

$$\Sigma^2_{SO} \equiv E_{\hat{\mathbf{\Omega}},\mathbf{r}}\left((\hat{\mathbf{w}}'\mathbf{r} - \hat{\mathbf{w}}'\mathbf{r})^2\,\big|\,\mathbf{\Omega}\right). \tag{6}$$

Performing the average over the returns $\mathbf{r}$ given $\hat{\mathbf{\Omega}}$ we have

$$\Sigma^2_{SO} \;\; = \;\; E_{\hat{\mathbf{\Omega}}}\left(E_{\mathbf{r}}\left((\hat{\mathbf{w}}'\mathbf{r} - \hat{\mathbf{w}}'\bar{\mathbf{r}})^2\,\big|\,\hat{\mathbf{\Omega}}\right)\,\Big|\,\mathbf{\Omega}\right),$$

---

[3]If there is no ambiguity, this may be denoted $E(f(x))$ or $E(f)$, to avoid cluttering the notation.

or

$$\Sigma_{SO}^2 \;=\; E_{\hat{\boldsymbol{\Omega}}}\left(\hat{\mathbf{w}}'\boldsymbol{\Omega}\hat{\mathbf{w}}\middle|\boldsymbol{\Omega}\right). \tag{7}$$

The final expression accounts for both the risk present in a given distribution and the additional risk due to distributional uncertainty. Although similar in appearance to (4), it differs significantly in that it depends not on the portfolio holdings, but on the strategy that led to them, through $\hat{\mathbf{w}}(\hat{\boldsymbol{\Omega}})$. It is our aim to reliably estimate it.

What is typically used to forecast risk, the "naive estimator"

$$\hat{\Sigma}_{naive}^2 = \hat{\mathbf{w}}'\hat{\boldsymbol{\Omega}}\hat{\mathbf{w}}, \tag{8}$$

may be significantly biased, even if the covariance matrix estimator $\hat{\boldsymbol{\Omega}}$ is unbiased, $E_{\hat{\boldsymbol{\Omega}}}(\hat{\boldsymbol{\Omega}}|\boldsymbol{\Omega}) = \boldsymbol{\Omega}$. Active management induces a functional dependence $\hat{\mathbf{w}}(\hat{\boldsymbol{\Omega}})$, a correlation between the portfolio and the estimation error in $\hat{\boldsymbol{\Omega}}$, so that

$$E_{\hat{\boldsymbol{\Omega}}}\left(\hat{\mathbf{w}}'\hat{\boldsymbol{\Omega}}\hat{\mathbf{w}}\middle|\boldsymbol{\Omega}\right) \;\neq\; E_{\hat{\boldsymbol{\Omega}}}\left(\hat{\mathbf{w}}'E_{\hat{\boldsymbol{\Omega}}}(\hat{\boldsymbol{\Omega}}|\boldsymbol{\Omega})\hat{\mathbf{w}}\middle|\boldsymbol{\Omega}\right), \tag{9}$$

or

$$E_{\hat{\boldsymbol{\Omega}}}(\hat{\Sigma}_{naive}^2|\boldsymbol{\Omega}) \neq \Sigma_{SO}^2. \tag{10}$$

The naive estimate of portfolio risk is typically lower than the true risk, and lower even than the optimal risk attainable with perfect knowledge of $\boldsymbol{\Omega}$. Intuitively, the optimized portfolio tends to overweight assets with underforecast risk, and to underweight assets whose risk $\hat{\boldsymbol{\Omega}}$ overestimates.

The degree of this bias grows with the uncertainty in $\hat{\boldsymbol{\Omega}}$ and the sensitivity of the portfolio to $\hat{\boldsymbol{\Omega}}$, via $\hat{\mathbf{w}}(\hat{\boldsymbol{\Omega}})$. For a portfolio constructed independent of $\hat{\boldsymbol{\Omega}}$, such as a passive index fund, the left and right of (10) are equal.

We compare Σ_{SO}^2 to the risk of the true, unknown optimal portfolio, $\mathbf{w}^*$. Any portfolio on the efficient frontier is the minimum risk portfolio under a fixed return constraint. For a minimum risk portfolio $\hat{\mathbf{w}}$ subject to continuous constraints, we may formally expand the risk about $\mathbf{w}^*(\boldsymbol{\Omega})$ as $\hat{\mathbf{w}} = \mathbf{w}^* + \Delta\hat{\mathbf{w}}$:

$$
\begin{aligned}
E(\hat{\mathbf{w}}'\boldsymbol{\Omega}\hat{\mathbf{w}}) &= E\left((\mathbf{w}^* + \Delta\hat{\mathbf{w}})'\boldsymbol{\Omega}(\mathbf{w}^* + \Delta\hat{\mathbf{w}})\right) \\
&= \mathbf{w}^{*'}\boldsymbol{\Omega}\mathbf{w}^* + 2\mathbf{w}^{*'}\boldsymbol{\Omega}E(\Delta\hat{\mathbf{w}}) + E\left(\Delta\hat{\mathbf{w}}'\boldsymbol{\Omega}\Delta\hat{\mathbf{w}}\right) \\
&= \mathbf{w}^{*'}\boldsymbol{\Omega}\mathbf{w}^* + E\left(\Delta\hat{\mathbf{w}}'\boldsymbol{\Omega}\Delta\hat{\mathbf{w}}\right).
\end{aligned} \tag{11}
$$

The cross term $\mathbf{w}^{*'}\mathbf{\Omega}\Delta\hat{\mathbf{w}}$ vanishes not just in expectation but for any $\hat{\mathbf{w}}$ satisfying the constraints, by the optimality condition on $\mathbf{w}^*$. The final expression (11) gives an intuitive decomposition of risk as that attainable with perfect knowledge of $\mathbf{\Omega}$ plus the cost of uncertainty.

Note that $E\left(\Delta\hat{\mathbf{w}}'\mathbf{\Omega}\Delta\hat{\mathbf{w}}\right)$ is positive, so the effect of the uncertainty is a risk penalty. Considering the portfolio $\hat{\mathbf{w}}$ to be an estimator of the true optimal portfolio $\mathbf{w}^*$, Equation (11) quantifies the risk cost of estimation error.[4]

Although we focus on the uncertainty due to estimation errors in $\hat{\mathbf{\Omega}}$, the expected value in Equation (7) may also be extended to other sources of uncertainty, such as stochastic time-dependence in $\mathbf{\Omega}$ and α. To quantify this behavior requires additional modeling assumptions, resulting in greater subjectivity, but the result is qualitatively the same: optimization produces an asymmetric downside exposure to model uncertainty.

3 Asset Covariance Matrix

We first explore second order risk in the context of the covariance matrix estimated directly from asset returns:

$$\hat{\mathbf{\Omega}} \equiv \frac{1}{T}\mathbf{r}\mathbf{r}'. \tag{12}$$

Here $\mathbf{r}$ is the $N \times T$ matrix of de-meaned[5] returns of N assets over T observation periods. Assuming Gaussian returns, $\hat{\mathbf{\Omega}}$ follows a Wishart distribution [6].

For the simple portfolio of Equation (5), the risk of (7) can be calculated explicitly. In terms of the observable $\hat{\mathbf{w}}'\hat{\mathbf{\Omega}}\hat{\mathbf{w}}$, we find

$$\Sigma^2_{SO} = E(\hat{\mathbf{w}}'\mathbf{\Omega}\hat{\mathbf{w}}) \quad \simeq \quad E(\hat{\mathbf{w}}'\hat{\mathbf{\Omega}}\hat{\mathbf{w}})\left(1 - \frac{N}{T}\right)^{-2}. \tag{13}$$

Details of the calculation are given in the Appendix.

The significance of Equation (13) is twofold: it demonstrates the scale of the bias, and immediately suggests how to correct it. Equation (13) implies

$$\hat{\Sigma}^2_{SO} \equiv \hat{\mathbf{w}}'\hat{\mathbf{\Omega}}\hat{\mathbf{w}}\left(1 - \frac{N}{T}\right)^{-2} \tag{14}$$

---

[4]bounded below by the Cramer-Rao bound of statistics.

[5]For our purposes, neglecting the $\sim 1/T$ estimation error of ex-post mean returns is a harmless simplifying assumption, unrelated to the difficult question of quantifying uncertainty in the forecast α.

is an unbiased estimator of the risk of the optimized portfolio:

$$E_{\hat{\mathbf{\Omega}}}(\hat{\Sigma}_{SO}^2|\mathbf{\Omega}) = \Sigma_{SO}^2.$$

Crucially, the correction for second order risk is a function of N and T only, which are known to the investor without additional information about $\mathbf{\Omega}$, making it possible to forecast second order risk. For an investment universe of 500 assets, and an asset covariance matrix estimated from 4 years of daily returns, Equation (14) doubles the predicted standard deviation of portfolio returns.

3.1 Empirical Results

The simplicity of the Sharpe ratio optimized portfolio (5) aided in deriving the simple second order risk correction in Equation (14), but the inflation factor $\left(1 - \frac{N}{T}\right)^{-2}$ can be a sufficient approximation to the correction needed for other utility functions.

Figure 3 shows the results of a Monte Carlo simulation for the minimum risk portfolio, constrained to be fully invested and to have fixed expected return $\hat{\mathbf{w}}'\alpha = R$ with respect to a randomly chosen α and fixed covariance matrix $\mathbf{\Omega}$. For each value of R, a new $\hat{\mathbf{\Omega}}$ is estimated from $T = 100$ observations of $N = 50$ returns.

The curve labeled "True Frontier" is the efficient frontier that could be achieved if $\mathbf{\Omega}$ were known with certainty, corresponding to the True Best point in Figure 2. Risk along the true frontier is given by $\sqrt{\mathbf{w}^{*\prime}\mathbf{\Omega}\mathbf{w}^*}$.

The points labeled "Realized" show the actual risk $\sqrt{\hat{\mathbf{w}}'\mathbf{\Omega}\hat{\mathbf{w}}}$ of the optimized portfolios, which correspond to True in Figure 2. This risk is well above the optimal risk, showing that estimation error degrades performance by preventing the optimal hedging of risk.

The "Naive Forecast" risk, $\sqrt{\hat{\mathbf{w}}'\hat{\mathbf{\Omega}}\hat{\mathbf{w}}}$, is seen to be significantly over-optimistic, on average by a factor of two. Its location to the right of the true frontier is in correspondence with the position of the Naive Best point in Figure 2, overestimating not only the utility attainable with a model, but also what would be attainable with perfect information.

In contrast, the "Corrected Forecast" $\sqrt{\hat{\mathbf{w}}'\hat{\mathbf{\Omega}}\hat{\mathbf{w}}}\left(1 - \frac{N}{T}\right)^{-1}$ accurately captures the risk of the optimized portfolio. Although its efficiency is diminished by noise, the corrected forecast provides unbiased estimates.

Testing the methodology with real market data is complicated by the fact that the "true" covariance matrix is not known, so $\hat{\mathbf{w}}'\mathbf{\Omega}\hat{\mathbf{w}}$ must be estimated by observing realized volatility. In the context of portfolios, forecasts, and market conditions that are changing in time, the Bias Statistic is a useful tool for testing the accuracy of risk forecasts.
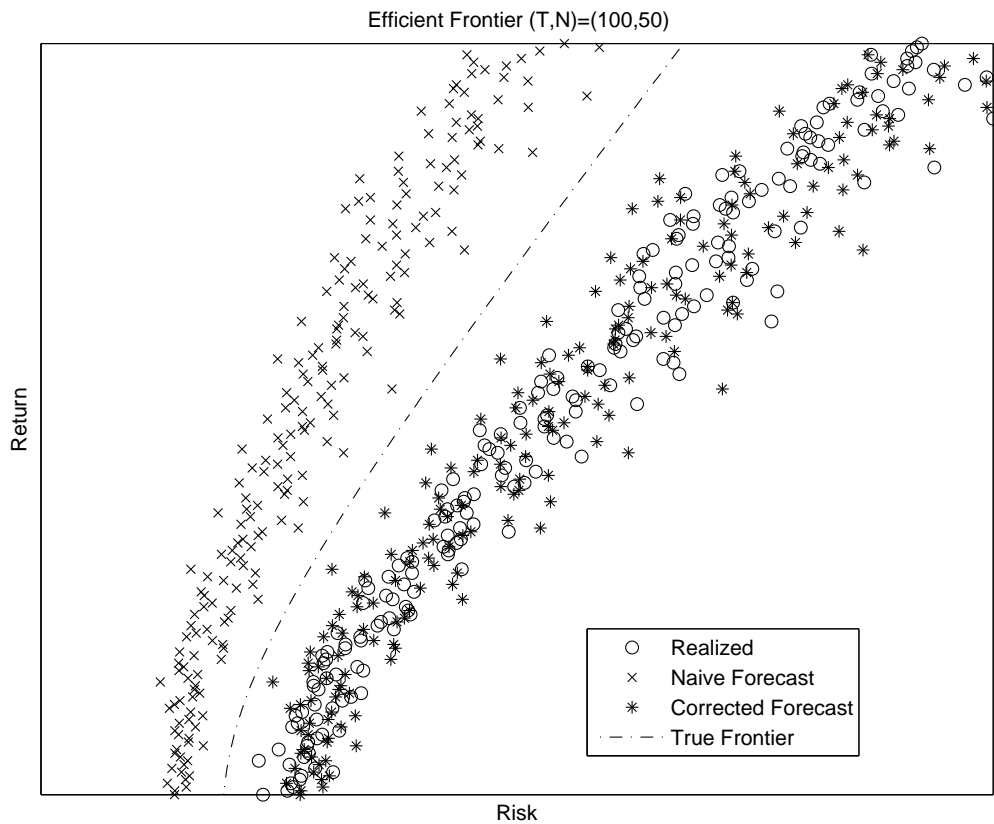
Figure 3: Monte Carlo simulation. The average Corrected Forecast is the same as the average Realized risk, below the true efficient frontier accessible only with perfect information about $\mathbf{\Omega}$. The Naive Forecast appears better than even the true efficient frontier.
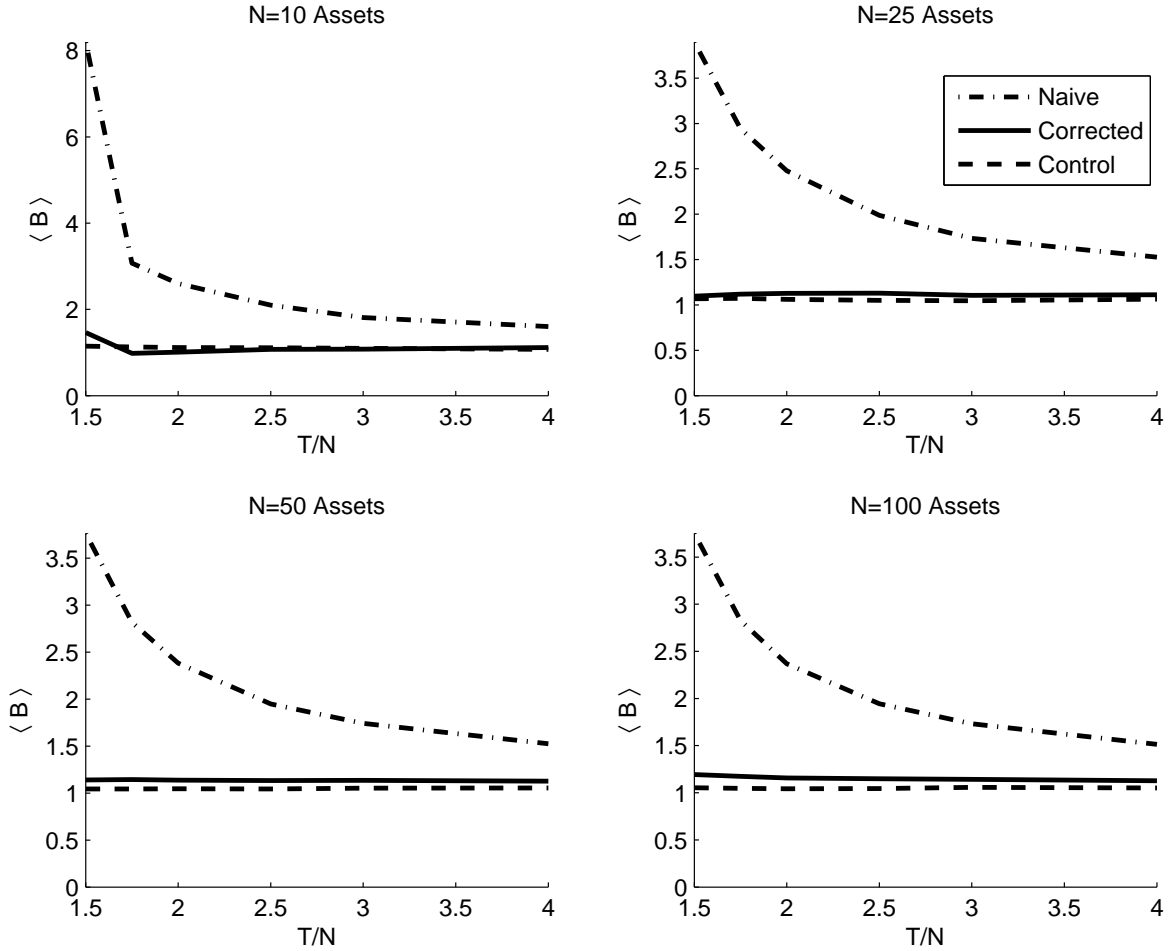
10

Figure 4: Average bias statistics of real portfolios with varying number of assets N and length of observation window T. A bias statistic greater (less) than 1 indicates under- (over-) forecast risk. "Naive" denotes the conventional risk forecast for the optimized portfolio, Equation (5), with random alpha vector among N assets. The "Corrected" forecasts are for the same portfolios as Naive, but with forecasts corrected for second order risk. The bias statistics of random portfolios among the same assets are shown for comparison.

For a given time series of portfolios $\hat{\mathbf{w}}_t$, the bias statistic is constructed from the forecast standard deviations $\hat{\Sigma}_t$ and out-of-sample realized returns R_{t+1} as

$$B \equiv std\left(\frac{R_{t+1}}{\hat{\Sigma}_t}\right). \tag{15}$$

An underforecast $\hat{\Sigma}_t$ produces bias statistics greater than 1, while overforecasts lead to $B < 1$.

We construct portfolios from a universe of the largest stocks in the United States, studying daily returns for the ten years through April 2009. To reduce the effects of extreme events, asset returns are trimmed to -50% and +80%. Assets with an incomplete history are excluded from the sample, leaving about 1800 stocks.

For each trial, we construct portfolios among N stocks selected at random. Optimized portfolios are constructed according to Equation (5), with $\hat{\Omega}$ estimated from a rolling window of T days, and a random α vector. Each day, risk forecasts are constructed with the naive estimator (8) and corrected estimator (14), from which bias statistics are calculated. As a control, we also construct random portfolios from the same stocks, which are not subject to the bias of second order risk.

Figure 4 shows the average bias statistics over 50 trials, for portfolios of $N=10$, 25, 50 and 100 assets, and observation windows T with $T/N=1.5$, 1.75, 2, 2.5, 3 and 4. In all cases, the standard risk forecast of optimized portfolios significantly underforecasts realized volatility, as indicated by bias statistics greater than 1. Comparison with the control demonstrates that the underforecast risk is a result of optimization, not some other feature of the distribution.

Second order risk is therefore responsible for a significant portion of the out of sample portfolio volatility. In contrast, the corrected forecasts are nearly as accurate as the control, confirming the validity of the estimator $\hat{\Sigma}_{SO}^2$.

4 Factor Process

Section 3 assumed the returns generating process to be a multivariate Gaussian at the level of the assets, with no additional structure. As the resulting estimation error effects went like N/T, a large investment universe may require decades of data for robust optimization, far longer than the timescales over which market relationships are stable.

Perhaps due to experience with such effects, most practitioners do not build large optimized portfolios from covariance matrices estimated directly from the assets, instead using more robust factor models. Rather than estimate the correlation among every pair of assets, a smaller number of systematic factors is identified, such as Value, Momentum, or industry membership. The return

of each asset is decomposed into the $K \times T$ factor return $\hat{\mathbf{f}}$ and the asset-specific, or idiosyncratic, return $\hat{\mathbf{e}}$.

$$\mathbf{r} = \hat{\mathbf{X}}\hat{\mathbf{f}} + \hat{\mathbf{e}}.$$

The $N \times K$ matrix $\hat{\mathbf{X}}$ defines the model exposure of each asset to the K factors. It is assumed that all correlation among assets is due to their common exposure to the factors, so that the covariance matrix takes the form

$$\hat{\mathbf{\Omega}} = \hat{\mathbf{X}}\hat{\mathbf{F}}\hat{\mathbf{X}}' + \hat{\Delta}, \tag{16}$$

where the estimated factor covariance matrix is defined to be $\hat{\mathbf{F}} \equiv \hat{\mathbf{f}} \cdot \hat{\mathbf{f}}'/T$,[6] and the specific risk matrix $\hat{\Delta}$ is assumed to be diagonal. For a broadly diversified portfolio of N assets, the specific risk contribution is suppressed by a factor of $1/N$ relative to the factor risk, assuming the latter has not been hedged away. Because it is diagonal, estimation error in the specific risk matrix does not have the effects of covariance matrices, as an optimizer is not fooled into hedging out spurious correlations. For simplicity, we use the true Δ rather than an estimate, which omits corrections suppressed to order $1/N$ compared to the effects under study.

Rather than estimate all $N(N+1)/2$ elements of the asset covariance matrix, the smaller $K \times K$ factor covariance matrix can be estimated from a much shorter time history, allowing a better reflection of the current market conditions. Provided the factor model accurately captures the returns process, it yields a far more robust estimate of the covariance among assets.

4.1 Factor Modeling Errors

Unfortunately, factor models may be subject to a variety of errors, which can degrade their performance. Similar to the asset-level case above, the factor covariance matrix has estimation errors due to the finite number of observations from which it is estimated. These errors are of order K/T, typically far better than the N/T behavior of errors of the asset covariance matrix, but significant.

Unlike the asset-level case, factor models have an additional source of error in the factor exposures $\hat{\mathbf{X}}$, which is more difficult to quantify. In general, determining the exposures is an inexact science, requiring financial insight as much as straightforward econometric technique. Errors in the exposures, though difficult to measure, can skew the risk forecasts of portfolios, particularly those constructed to minimize exposure to systematic risk.

---

[6]We again neglect ex-post mean returns, for simplicity of exposition.

Two classes of exposures errors may be identified, which we call "coherent" and "incoherent." Incoherent errors are those that are uncorrelated with the portfolio, so that their aggregate effect is diversified. For example, if the true exposures $\mathbf{X}$ differ from the model exposures by uncorrelated random noise, $\mathbf{X} = \hat{\mathbf{X}} + \epsilon$, then the portfolio exposure error is approximately

$$\sum_i w_i \epsilon_i \sim \frac{1}{N}\sqrt{N} std(\epsilon)$$
$$\sim \frac{std(\epsilon)}{\sqrt{N}},$$

giving a contribution to variance $\sim 1/N$, behaving like diversifiable specific risk rather than systematic factor risk. The $\sqrt{N}$ suppression is due to the low likelihood that many terms contribute with the same sign.

A more dangerous class of errors, the coherent exposure errors, are those for which the portfolio is likely to accumulate a finite exposure. These errors can be generated by discrepancies between the alpha signal and the risk model, such as small differences in factor definitions. Similar to other errors we have seen, the optimized portfolio tends to align itself with these errors, resulting in an unsuppressed, non-diversifiable contribution

$$\sum_i w_i \epsilon_i \sim \frac{1}{N} N std(\epsilon)$$
$$\sim std(\epsilon).$$

For example, if a risk model defines a Value factor in terms of the Book-to-Price ratio, while the alpha Value factor uses Earnings-to-Price, the optimized portfolio may make large unintended bets [7] on the difference between these Value factor definitions. The contribution from this hidden exposure may be significant for a portfolio that has hedged away known factor exposures.

To see how these coherent exposure errors occur, consider the estimated maximum Sharpe ratio portfolio, Equation (8). Without loss of generality, we work in a basis of assets such that the specific risk is uniform, $\Delta = \sigma^2 \mathbf{1}_N$.[7] Furthermore, we choose a basis of model factors such that $\hat{\mathbf{X}}'\hat{\mathbf{X}} = N \mathbf{1}_K$. The factor of N captures the scaling behavior of $\hat{\mathbf{X}}'\hat{\mathbf{X}}$ with the number of assets.

For the factor model of Equation (16), it is useful to decompose α into components in the plane of the model exposures, and an orthogonal piece

$$\alpha = \hat{\mathbf{X}}\mathbf{a} + \alpha_\perp, \tag{17}$$

---

[7]This can be achieved by taking $\alpha_i \to \alpha_i \sigma / \sigma_i, r_i \to, r_i \sigma / \sigma_i, X_i \to X_i \sigma / \sigma_i$, which is the usual map from optimal regression weights proportional to σ_i^{-2} to an equivalent OLS regression.

where $\alpha_\perp \equiv (\mathbf{1} - \frac{\hat{\mathbf{X}}\hat{\mathbf{X}}'}{N})\alpha$ satisfies $\hat{\mathbf{X}}'\alpha_\perp = 0$, and $\mathbf{a} \equiv \frac{\hat{\mathbf{X}}'\alpha}{N}$. After some algebra, Equations (5), (16) and (17) yield

$$\hat{\mathbf{w}} = w\left(\alpha_\perp + \sigma^2 N^{-1}\hat{\mathbf{X}}\hat{\mathbf{F}}^{-1}\mathbf{a}\right) + \dots, \tag{18}$$

where w is an arbitrary constant. The $\dots$ represent terms $\mathcal{O}\left((\frac{\sigma^2}{N}\hat{\mathbf{F}}^{-1})^2\right)$, which are suppressed by a further factor of N in a large universe of assets.

Equation (18) shows the potential for distortions from a misalignment of the alpha factors and risk factors, with $\mathcal{O}(N^{-1})$ suppression for the factor component of α relative to the component $\alpha_\perp$ orthogonal to the plane of known factor risk. The model factor risk

$$\hat{\mathbf{w}}'\hat{\mathbf{X}}\hat{\mathbf{F}}\hat{\mathbf{X}}'\hat{\mathbf{w}} = \mathbf{a}'\hat{\mathbf{F}}^{-1}\mathbf{a} \tag{19}$$

sees no systematic risk in $\alpha_\perp$.

Unless the manager has extraordinary skill, $\alpha_\perp$ may represent noise rather than an arbitrage opportunity[8] (in the sense of the Arbitrage Pricing Theory), and optimization points the portfolio into a blind spot of the risk model.

These coherent exposure errors are avoided if the factor exposures contain the full alpha signal,

$$\alpha = \hat{\mathbf{X}}\mathbf{a}. \tag{20}$$

If the alpha signal contains a component orthogonal to the plane spanned by the factor exposures, this component can be used to estimate an additional factor [8]:

$$\hat{\mathbf{f}}^{(\alpha_\perp)} = \frac{\alpha_\perp'\mathbf{r}}{\alpha_\perp^2}. \tag{21}$$

Although the resulting exposures may still contain errors, they are of the safe, incoherent variety, as the portfolio is unlikely to align along them.

If it is not feasible to estimate an additional factor from the alpha signal, so that α is in the plane of model factor exposures, we do not know whether the orthogonal component $\alpha_\perp$ represents an arbitrage opportunity, or just a limitation of the risk model.

To see this, assume that there are true risk factors $\mathbf{X}$, which differ from the model factors $\hat{\mathbf{X}}$ by

$$\mathbf{X} = \hat{\mathbf{X}} + \epsilon. \tag{22}$$

---

[8]Even managers using a stock-picking strategy, rather than making explicit factor bets, may be judging assets on a handful of criteria that constitute risk factors.

We can assume without loss of generality that the noise is orthogonal to the model exposures[9],

$$\hat{\mathbf{X}}'\epsilon = 0. \tag{23}$$

A bit of algebra shows that the estimated factor returns $\hat{\mathbf{f}}$ associated with $\hat{\mathbf{X}}$ are related to the true factor returns $\mathbf{f}$ as

$$\hat{\mathbf{f}} = \mathbf{f} + \hat{\mathbf{X}}'\mathbf{e}/N. \tag{24}$$

Therefore, at large N and T, the estimated factor covariance matrix is accurate, $\hat{\mathbf{F}} \simeq \mathbf{F}$, and the true factor risk of (18) is related to the model, Equation (19), by

$$\hat{\mathbf{w}}'\mathbf{X}\mathbf{F}\mathbf{X}'\hat{\mathbf{w}} = \hat{\mathbf{w}}'\hat{\mathbf{X}}\hat{\mathbf{F}}\hat{\mathbf{X}}'\hat{\mathbf{w}} + w^2 \left(\alpha'_\perp \epsilon \mathbf{F} \epsilon' \alpha_\perp + 2\sigma^2 \mathbf{a}' \epsilon' \alpha_\perp \right). \tag{25}$$

The terms on the right represent a contribution to the factor risk that the model does not see, a form of second order risk. The noise ϵ is unknown, but the orientation of the portfolio to the plane of model exposures provides insight into its magnitude.

Assuming "no arbitrage", that α lies fully in the plane of the true factors $\mathbf{X}$

$$\alpha = \mathbf{X}\mathbf{a}, \tag{26}$$

then $\alpha_\perp = \epsilon \mathbf{a}$, and the remaining ϵ dependence is of the form $\epsilon'\epsilon$, which in a large universe is sensitive only to the statistics of ϵ, rather than its details. Imposing the further assumption that ϵ arises from white noise,

$$\epsilon'\epsilon \simeq N\rho^2 \mathbf{1}_K, \tag{27}$$

where ρ is an unknown noise parameter, we may solve for the total factor risk of (25)

$$\hat{\mathbf{w}}'\mathbf{X}\mathbf{F}\mathbf{X}'\hat{\mathbf{w}} = \hat{\mathbf{w}}'\hat{\mathbf{X}}\hat{\mathbf{F}}\hat{\mathbf{X}}'\hat{\mathbf{w}} + w^2 \left(\frac{\alpha_\perp^4 \mathbf{a}'\mathbf{F}\mathbf{a}}{\mathbf{a}^4} + 2\alpha_\perp^2 \sigma^2 \right). \tag{28}$$

Like Equation (14), the right-hand side forecasts second order risk using only information available to the investor. Though relying on the assumptions of Equations (26) and (27), these corrections warn of the susceptibility to second order risk for a portfolio tilted far out of the plane of the model exposures.

---

[9]Since factor processes are invariant under an arbitrary change of basis in the space of factors, $\mathbf{X} \rightarrow \mathbf{X}\mathbf{M}$, if $\hat{\mathbf{X}}'\epsilon \neq 0$, we can redefine $\mathbf{X} \rightarrow \mathbf{X} \left(\hat{\mathbf{X}}'\mathbf{X} \right)^{-1} \hat{\mathbf{X}}'\hat{\mathbf{X}}$, so that $\hat{\mathbf{X}}'\epsilon = \hat{\mathbf{X}}'(\mathbf{X} - \hat{\mathbf{X}}) = 0$.

4.2 Factor Model Estimation Error

If the model exposures contain α, as in Equation (20), then the effects of modeling errors discussed in Section 4.1 are replaced by the milder effects of estimation error.

We continue to assume as in Equations (22) and (23) that there is unknown noise in the exposures. Equation (24) implies $\hat{\mathbf{f}} = \mathbf{f}$, at large N. Unlike in Section 4.1, we do not assume the large T limit, so there is estimation error in the model factor covariance matrix $\hat{\mathbf{F}} = \frac{\mathbf{f} \cdot \mathbf{f}'}{T}$. With the assumption $\alpha_\perp = 0$, the true factor risk of the portfolio (18) is

$$
\begin{aligned}
\hat{\mathbf{w}}'\mathbf{XFX}'\hat{\mathbf{w}} &= N^{-2}\mathbf{a}'\hat{\mathbf{F}}^{-1}\hat{\mathbf{X}}'\mathbf{XFX}'\hat{\mathbf{X}}\hat{\mathbf{F}}^{-1}\mathbf{a} \\
&= \mathbf{a}'\hat{\mathbf{F}}^{-1}\mathbf{F}\hat{\mathbf{F}}^{-1}\mathbf{a},
\end{aligned}
\tag{29}
$$

while the naive estimate is

$$
\hat{\mathbf{w}}'\hat{\mathbf{X}}\hat{\mathbf{F}}\hat{\mathbf{X}}'\hat{\mathbf{w}} = \mathbf{a}'\hat{\mathbf{F}}^{-1}\mathbf{a}.
\tag{30}
$$

In close analogy with the asset-level case, taking expected values of Equations (29) and (30) yields

$$
\Sigma^2_{SO,f} \equiv E_{\hat{\mathbf{F}}}\left(\hat{\mathbf{w}}'\mathbf{XFX}'\hat{\mathbf{w}}\,\middle|\,\mathbf{F}\right) = \left(1 - \frac{K}{T}\right)^{-2} E_{\hat{\mathbf{F}}}\left(\hat{\mathbf{w}}'\hat{\mathbf{X}}\hat{\mathbf{F}}\hat{\mathbf{X}}'\hat{\mathbf{w}}\,\middle|\,\mathbf{F}\right).
\tag{31}
$$

Therefore, the estimator

$$
\hat{\Sigma}^2_{SO,f} \equiv \left(1 - \frac{K}{T}\right)^{-2} \hat{\mathbf{w}}'\hat{\mathbf{X}}\hat{\mathbf{F}}\hat{\mathbf{X}}'\hat{\mathbf{w}}
\tag{32}
$$

provides unbiased forecasts, without additional knowledge of the true factors. More complicated corrections hold beyond the $N \to \infty$ limit. Despite the true factor risk depending on both the unknown exposures $\mathbf{X}$ and factor covariance matrix $\mathbf{F}$, Equation (32) forecasts second order risk using only observable quantities.

While the $\frac{N}{T}$ effects of Section 3 could overwhelm a large portfolio, these $\frac{K}{T}$ are more likely to be under control. However, for a typical factor model of $K \sim 50$ factors and effective observation window[10] $T \sim 200$, the $\sim 30\%$ boost to forecast volatility is an important correction.

4.3 Empirical Results

To study the effects of factor model estimation errors, we consider the Barra Global Equity Model (GEM2) [9]. The model estimates a World factor, 34 industry factors, 55 country factors and 8

---

[10]See the discussion of effective observation window in Section 5.

style factors from an estimation universe based on the MSCI All Country World Index, consisting of about 8000 stocks.

We consider an ensemble of 500 portfolios constructed from the estimation universe, optimized using (18) relative to $\alpha = \hat{\mathbf{X}}\mathbf{a}$, with random vector $\mathbf{a}$.[11] To make contact with Equation (32) we consider $\hat{\boldsymbol{\Omega}}$ given by Equation (16), with $\hat{\mathbf{F}}$ and $\hat{\Delta}$ estimated with equally weighted covariance estimators over a rolling window of $T = 156$ weekly returns. We discuss the corrections necessary for the exponentially weighted, Newey-West estimator used in GEM2 in Section 5. To reduce the effect of extreme events, asset returns outside of $(-80\%, 400\%)$ or exceeding ten cross-sectional standard deviations are dropped.

As a control, we also construct 500 portfolios of the form $\hat{\mathbf{w}} = \hat{\mathbf{X}}\mathbf{b}$, with $\mathbf{b}$ a random vector. Though exposed to factor risk, these portfolios are not subject to the biased second order risk of optimized portfolios.
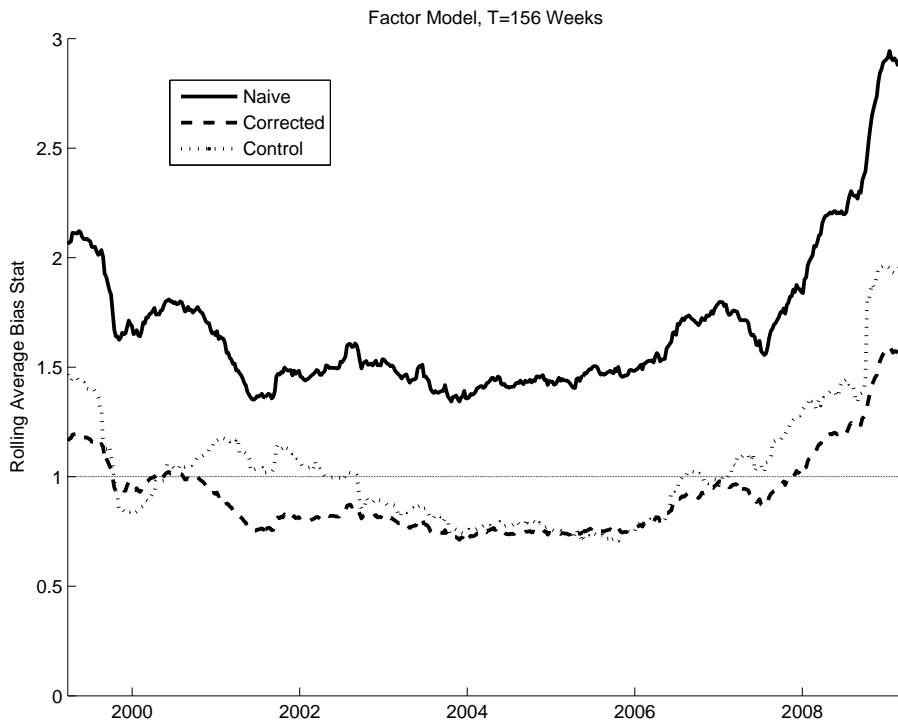

Figure 5: Average trailing bias statistics for ensembles of 500 optimized and random portfolios.

---

[11]The two-fold exact multicollinearity of GEM2 is resolved by projecting to a smaller subspace of factors.

Each week, the risk is forecast using the standard risk estimator and the corrected estimator (31). Bias Statistics (15) are calculated with a standard deviation over a trailing 52 week window, over the decade ending March 2009.

The results in Figure 5 are easily interpreted. The control has bias statistics greater than one during periods of upward-trending volatility, prior to 2000 and since late 2007. Similarly, the control has bias statistics consistently below one between 2001 and early 2007, during which volatility trended downward and the trailing window overforecast volatility of future returns. The deviation from one is exacerbated by the use of the OLS estimator for this study, which is less responsive to changing market conditions.

The naive risk forecasts of the optimized portfolios have bias statistics significantly greater than one for the whole of the study, demonstrating that volatility is underforecast when second order risk is neglected. Since the risk forecasts are so much worse than those of the control, the error must be due to optimization, rather than the underlying factor model. An average bias statistic of about two indicates that the naive forecasts only capture about half the true risk. The risk due to distributional uncertainty therefore contributes fully half of the risk of these portfolios.

In comparison, the corrected risk forecasts match the accuracy of the control for the length of the study. By accounting for second order risk in this way, it is possible to more accurately forecast the volatility of optimized portfolios.

5 Discussion

The techniques developed here begin to account for the costs of distributional uncertainty; however, they are not fully general, having made a number of simplifying assumptions to calculate the corrections to the risk forecast.

Some of these assumptions may be relaxed without much difficulty. Rather than the equal-weighted covariance matrix estimators we have considered, it is common to use an exponentially weighted estimator that puts greater weight on recent events. For a half-life τ we may account for these effects to leading order in K/τ by replacing the number of observations T with an effective time window

$$T \to 2\tau/ln(2). \tag{33}$$

Similarly, the Newey-West estimator [10] accounting for $n > 1$ lags of serial correlation can be

approximated with

$$T \to 3T/2(n+1),$$

and for a fat-tailed process with uniform kurtosis k, we may take

$$T \to 2T/(k-1). \tag{34}$$

Note that Equation (33) accounts for the use of the EWMA estimator, but not the non-stationarity for which it is adopted. To account for the latter, the general expression (7) could be adapted to a model for the market process.

Similarly, while the kurtosis correction (34) approximately accounts for a uniformly kurtotic underlying process, it does not treat an optimization that attempts to minimize risk to extreme events with a utility function that penalizes assets according to their estimated tail behavior. We expect the effects of this paper to be amplified in the context of fat-tails optimization, due to the increased estimation error associated with rare, large events.

Another generalization is to the case of constrained optimization. We find the second order risk forecasts to be robust to the inclusion of a small number of constraints, and linear constraints may be accounted for easily. However, it is difficult to extend analytic results to more general constraints, and suggest a heuristic based on the transfer coefficient [11], which measures the discrepancy between the constrained and unconstrained optimal portfolios. A Monte Carlo approach may also be useful.

Other generalizations are more difficult. A portfolio manager might reduce a position based on a large forecast marginal contribution to risk, creating the correlations between the portfolio weights and $\hat{\mathbf{\Omega}}$ that lead to biases in the standard risk forecasts, but such an investment strategy is difficult to quantify. More difficult still are the true "unknown unknowns", whose effects – by definition – cannot be anticipated.

On a larger scale, second order risk provides insight into the current financial crisis, in which a financial system optimized under models and assumptions of the economy finds itself with far more risk than it had accounted for. While our focus has been the risk of an optimized portfolio, there is likely a general principle at work. If something has been tuned to a particular measure, that measure is likely to become exaggerated. Similar distortions may be commonplace, with examples ranging from standardized test scores to the balance sheet information upon which executive compensation is based.

6 Acknowledgements

The author is grateful to Lisa Goldberg and Jose Menchero for helpful comments, and to MSCI Barra for support of this work. A sample of related work may be found in [12], [13] and [14].

A Mathematical Details

To establish (13), we use two identities of the Wishart distribution:

$$E(\hat{\boldsymbol{\Omega}}^{-1}) = \frac{\boldsymbol{\Omega}^{-1}}{(1 - \frac{N-1}{T})},$$

$$E(\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Omega}\hat{\boldsymbol{\Omega}}^{-1}) = T^2(T-1)\left[(T-N)(T-N-1)(T-N-3)\right]^{-1}\boldsymbol{\Omega}^{-1}$$

$$\simeq \left(1 - \frac{N}{T}\right)^{-3}\boldsymbol{\Omega}^{-1}$$

$$\simeq \left(1 - \frac{N}{T}\right)^{-2}E(\hat{\boldsymbol{\Omega}}^{-1}).$$

The approximations drop $\mathcal{O}(1/N)$ and $\mathcal{O}(1/T)$ terms, for simplicity. For the portfolio

$$\hat{\mathbf{w}} = \hat{\boldsymbol{\Omega}}^{-1}\alpha$$

comparison of the average naive forecast

$$E(\hat{\mathbf{w}}'\hat{\boldsymbol{\Omega}}\hat{\mathbf{w}}) = \alpha'E(\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\Omega}}^{-1})\alpha$$

$$= \alpha'E(\hat{\boldsymbol{\Omega}}^{-1})\alpha,$$

with the average true risk

$$E(\hat{\mathbf{w}}'\boldsymbol{\Omega}\hat{\mathbf{w}}) = \alpha'E(\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Omega}\hat{\boldsymbol{\Omega}}^{-1})\alpha$$

$$= \left(1 - \frac{N}{T}\right)^{-2}\alpha'E(\hat{\boldsymbol{\Omega}}^{-1})\alpha,$$

yields Equation (13).

References

[1] http://www.defenselink.mil/transcripts/transcript.aspx?transcriptid=2636.

[2] Taleb, Nassim Nicholas (2004), *Fooled by Randomness*. New York: Random House.
   Taleb, Nassim Nicholas (2007), *The Black Swan*. New York: Random House.

[3] Markowitz, H. (1952), "Portfolio Selection". The Journal of Finance, Vol. 7, No. 1 , pp. 77-91

[4] Michaud. R. O. (1989), "The Markowitz Optimization Enigma: Is Optimized Optimal?" Financial Analysts Journal 45, 31-42.

[5] Rosenberg, B. and Guy, J. (1976), "Prediction of Beta from Investment Fundamentals" Financial Analysts Journal pp. 60-72

[6] http://en.wikipedia.org/wiki/Wishart_distribution

[7] Lee, J.H. and Stefek, D. (2008), "Do Risk Factors Eat Alpha?" Journal of Portfolio Management

[8] Menchero, J. and Poduri, V. (2008), "Custom Factor Attribution" Financial Analysts Journal Vol. 64; No. 2, pp. 81-92

[9] Menchero, J., Morozov, A. and Shepard, P. (2008), "The Global Equity Model GEM2" MSCI Barra Model Insights

[10] Newey, W. and West, K. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelationconsistent Covariance Matrix." Econometrica, Vol. 55, No. 3: 703-708, http://ssrn.com/abstract=225071

[11] R. Clark, de Silva, H. and Thorley, S. (2001), "Portfolio Constraints and the Fundamental Law of Active Management" Financial Analysts Journal, http://ssrn.com/abstract=290916

[12] Zhou, G. (2008), "On the Fundamental Law of Active Management: What Happens If Our Estimates Are Wrong?" Journal of Portfolio Management, Summer 2008, pp. 26-33.

[13] Briner, B. and Connor, G. (2008), "How Much Structure is Best?" Journal of Risk

[14] Ledoit, O. and Wolf, M. (2004), "Honey, I shrunk the sample covariance matrix" Journal of Portfolio Management, 30(4):110119.