February 6, 2024

This document discusses optimal measurement of volatility in the Global Multi-Asset Model 3.0 (GMAM 3.0 or G3). It is a companion the the general model document.

1 General considerations

In general, an analyst may be interested in the volatility of the reported returns of fund i, given in terms of population moments as:

$$V(y_{it}|\Pi) = E\left[(y_{it} - E[y_{it}|\Pi])^2 |\Pi\right]$$

where y_t represents the output of the return reporting process at time t and Π represents the set of observed quantities. Without loss of generality, assume the analyst is most interested in unannualized returns. For notational brevity, future notational will not include the i subscript. Also it will be assumed that all expectations are over the posterior distribution and therefore implicitly conditioned on the available data.

The analyst may also be interested in the true economic volatility:

$$V(x_t|\Pi) = E\left[(x_t - E[x_t|\Pi])^2 |\Pi\right]$$

where x_t represents the log difference in economic value of the underlying fund. The below discussion considers several estimators for each of these quantities, both inclusive and exclusive of volatility from parameter uncertainty.

2 Naive Measurements

Naive estimates of volatility consist of direct application of the standard formula for calculating sample variance $\frac{1}{T}\sum_{t\in 1:T}(r_t - \overline{r})^2$. Ignore the asymptotically equivalent small sample correction of using $T-1$ in the denominator.

2.1 Historical volatility

The standard way of measuring historical volatility remains a valid approach, provided the analyst is most interested in the volatility of the reported returns. An appropriate estimator is given as:

$$\hat{\sigma}_y^2 = \frac{1}{T_h}\sum_{t\in 1:T_h}(y_t - \overline{y})^2$$

where T_h is the number of periods and $\overline{y}$ is the simple average of returns. This estimator is perfectly reasonable provided the historical time series is long enough. It is often not the most efficient choice, however, it is a consistent estimator in a statistical sense.

2.2 Back-casted volatility

In contrast, naive back-casted volatility is not appropriate in almost any circumstances. To see why, consider a fund with a market beta of 1.0 and no other factor exposures. For simplicity, assume a zero interest rate. Let f_t represent the return on the market. The expected backcasted return is then:

$$\begin{aligned} E\left[x_t|\Pi\right] =& E\left[f_t + \varepsilon_t|\Pi\right] \\ =& f_t \end{aligned}$$

The last quantity follows from conditioning on the observed factor returns and interest rates. The variance of this quantity is then:

$$\begin{aligned} V\left(E\left[x_t|\Pi\right]\right) =& V\left[f_t|\Pi\right] \\ =& \sigma_m^2 \end{aligned}$$

This quantity represents the variances of the prediction as opposed to a variance of expected returns. It will underestimate volatility by the residual variance σ_ε^2. It perversely increases as model fit declines, implying that risk measured in this manner will understate true volatility more in scenarios where the model has a poor "understanding" of the underlying funds.

3 Bayesian expected volatility

3.1 Interpolated de-smoothed returns

Over the window of historical returns, G3 estimates the de-smoothed returns x_t by conditioning on both the factor data and the historical returns. The calculation is slightly more complex due to x_t being a parameter, and therefore a random variable within the model. Specifically, the expected de-smoothed volatility using this series is given by:

$$\begin{aligned} E\left[\left(x_t - \overline{x}\right)^2|\Pi\right] =& \int_\Theta E\left[\left(x_t - E\left[x|\Theta,\Pi\right]\right)^2|\Theta,\Pi\right] p\left(\Theta|\cdot\right) d\Theta \\ \hat{\sigma}_x^2 =& \frac{1}{J}\sum_{j\in 1:J}\frac{1}{T}\sum_{t\in 1:T}\left(x_{tj} - \overline{x}_j\right)^2 \end{aligned}$$

where Θ is the set of parameters and j indexes the J draws from the MCMC, and Π represents all observed quantities. The last equation follows the standard methodology for estimating the expectation over a posterior distribution.

3.1.1 Estimate uncertainty and interpolated de-smoothed returns

An analyst may be interested in the volatility of the implied distribution of returns. This is different than the expectation of the volatility measurement because it accounts for parameter uncertainty. From an estimation standpoint, such a volatility measurement is the variance across paths.

To see why this can be important, suppose an analysis indicates that a fund is equally likely to have a β to the market between 0 and 2 with β independent of all other parameters. Assume idiosyncratic variance σ_ε^2 and the variance of the market σ_m^2 are known with certainty, and the risk-free rate is constant. Then fund's expected variance over a long time horizon is given by:

$$\begin{aligned}
\sigma_x^2 =& \frac{1}{2}\int_0^2 V(x|\beta)\, d\beta \\
=& \frac{1}{2}\int_0^2 \left(\beta^2\sigma_m^2 + \sigma_\varepsilon^2\right) d\beta \\
=& \frac{4}{3}\sigma_m^2 + \sigma_\varepsilon^2
\end{aligned}$$

However, the above does not represent the volatility from which the subsequent period's return is drawn, because it does not reflect the uncertainty in the estimate of β. To account for this uncertainty, consider a simple single factor model for an asset where the asset's exposure to the factor f_m is given by $\beta \sim Unif\,(0,2)$. Then the volatility of the next data point given the observed data is given by:

$$\begin{aligned}
\tilde{\sigma}_x^2 &= \frac{1}{2}\int_0^2 \left[\left(E[x|\beta] - \mu_x\right)^2\right] d\beta \\
&= \frac{1}{2}\int_0^2 \left[E\left[\left(\beta f_m + r_f + \varepsilon - \mu_x\right)^2 |\beta\right]\right] d\beta \\
&= \frac{1}{2}\int_0^2 \left[E\left[\left(\beta f_m + r_f + \varepsilon - E[\beta f_m + r_f + \varepsilon]\right)^2 |\beta\right]\right] d\beta \\
&= \frac{1}{2}\int_0^2 \left[E\left[\left(\beta f_m + \varepsilon - \mu_m\right)^2 |\beta\right]\right] d\beta \\
&= \frac{4\sigma_m^2 + \mu_m^2}{3} + \sigma_\varepsilon^2
\end{aligned}$$

Which answer is more appropriate depends on the question asked. An analyst might wish to know the volatility of the subsequent month's return given the available information, or an analyst may wish to know the long-run volatility they would realize from an investment. To compute the former in the context of G3, just compute the pooled grand variance:

$$\begin{aligned}
\varsigma_x^2 &= \int_\Theta E\left[\left(x_{tj} - E[x|\Theta,\Pi]\right)^2\right] p(\Theta|\Pi)\, d\Theta \\
\hat{\varsigma}_x^2 =& \frac{1}{J}\sum_{j\in 1:J}\sum_{t\in 1:T}\left(x_{tj} - E[x|\Pi]\right)^2
\end{aligned}$$

For estimation purposes, the only difference is the use of a pooled mean as an estimator for $E[x|\Pi]$.

3.2 Enhancements to de-smoothed volatility estimates

The previous estimator of de-smoothed volatility is statistically consistent. It can be improved, however, by disaggregating the de-smoothed volatility into systematic and idiosyncratic components. The the de-smoothed idiosyncratic variance is conditionally modeled as a non-central T distribution with

ν degrees of freedom.

$$\begin{aligned}
\varepsilon_{xt} \sim & Tdist\left(0, \frac{1}{\tau_x \tau_y}, \nu\right) \\
\implies Var\left(\varepsilon_{xt}|\Theta, \Pi\right) = & \frac{1}{\tau_x \tau_y}\left(\frac{\nu}{\nu-2}\right)
\end{aligned}$$

where ν, τ_x, τ_y are parameters within the posterior distribution. The estimator for this quantity is therefore:

$$\begin{aligned}
\sigma^2_{\varepsilon x} = & \int_\Theta Var\left(\varepsilon_{xt}|\Theta, \Pi\right) p\left(\Theta|\Pi\right) d\Theta \\
= & \int_\Theta \frac{1}{\tau_x \tau_y}\left(\frac{\nu}{\nu-2}\right) p\left(\Theta|\Pi\right) d\Theta \\
\hat{\sigma}^2_{\varepsilon x} = & \frac{1}{J}\sum_{j\in 1:J} \frac{1}{\tau_{xj}\tau_{yj}}\left(\frac{\nu_j}{\nu_j - 2}\right)
\end{aligned}$$

Because factors are taken as a given in the model, an estimate of their covariance is necessary in order to compute a disaggregated systematic volatility. Let $\tilde{F}$ represent the matrix of factor returns augmented with an extra column containing the risk-free return. Similarly, let $\tilde{\beta}$ represent the exposures of the asset augmented by a 1. To estimate $\Sigma_{\tilde{F}}$, use the standard covariance estimator for F:

$$\hat{\Sigma}_{\tilde{F}} = \frac{1}{T}\left(\tilde{F}'\tilde{F} - \hat{\mu}_{\tilde{F}}\hat{\mu}'_{\tilde{F}}\right)$$

Note that the length of the factor history does not need to coincide with the length of the fund's track record. Particularly for short history funds, using the longer history will lead to an improved estimate (see appendix). Let $\hat{\Sigma}^{LR}_{\tilde{F}}$ indicate the long-run covariance matrix. Then the systematic variance is given as:

$$\begin{aligned}
\sigma^2_{sys} = & \int_\Theta V\left(\tilde{f}'_t\beta|\Theta, \Pi\right) p\left(\Theta|\Pi\right) d\Theta \\
= & \int_\Theta \tilde{\beta}'\hat{\Sigma}^{LR}_{\tilde{F}}\tilde{\beta} p\left(\Theta|\Pi\right) d\Theta \\
\hat{\sigma}^2_{sys} = & \frac{1}{J}\sum_{j\in 1:J} \tilde{\beta}'_j\hat{\Sigma}^{LR}_{\tilde{F}}\tilde{\beta}_j
\end{aligned}$$

The return reporting process includes idiosyncratic variation not reflected in the above, denoted as contemporaneous reporting variance. Because the variance is irreducible and iid, analysts may choose to include the reporting variance in the estimate of de-smoothed volatility. An alternative interpretation is the reporting variance represents perfectly known idiosyncratic changes to the valuation orthogonal to the smoothed idiosyncratic variation. Either way, the reporting variance is given as:

$$
\begin{aligned}
\sigma^2_{\varepsilon y} =& \int_\Theta V\left(\varepsilon_{ty}|\Theta,\Pi\right) p\left(\Theta|\Pi\right) d\Theta \\
=& \int_\Theta \frac{1}{\tau_y} p\left(\Theta|\Pi\right) d\Theta \\
\hat{\sigma}^2_{\varepsilon y} =& \frac{1}{J}\sum_{j\in 1:J}\frac{1}{\tau_{yj}}
\end{aligned}
$$

Note that the reporting variance is of frequency determined by the historical fund data, e.g. monthly or quarterly, while the other variance quantities are of the frequency of the factor data, e.g. monthly. Let Δt represent the number of factor observations per observation of the fund historical data. If the fund historical data is quarterly and the factor data is monthly, $\Delta t = 3$. If both sets of observations are of the same frequency, $\Delta t = 1$. Including the correction gives a model implied estimate of de-smoothed volatility is:

$$
\begin{aligned}
\sigma^2_x =& \int_\Theta \left[V\left(x_t + \frac{\varepsilon_y}{\sqrt{\Delta t}}|\Theta,\Pi\right)\right] p\left(\Theta|\Pi\right) d\Theta \\
=& \int_\Theta \left[\tilde{\beta}'\hat{\Sigma}^{LR}_{\tilde{F}}\tilde{\beta} + \frac{1}{\tau_x\tau_y}\left(\frac{\nu}{\nu-2}\right) + \frac{1}{\Delta t\tau_y}\right] p\left(\Theta|\Pi\right) d\Theta \\
\hat{\sigma}^2_x =& \frac{1}{J}\sum_{j\in 1:J}\left[\tilde{\beta}'_j\hat{\Sigma}^{LR}_{\tilde{F}}\tilde{\beta}_j + \frac{1}{\tau_{xj}\tau_{yj}}\left(\frac{\nu_j}{\nu_j-2}\right) + \frac{1}{\Delta t\tau_{yj}}\right]
\end{aligned}
$$

This represents the expectation of the long-run variance of the de-smoothed returns.

3.2.1 Accounting for estimate uncertainty in enhanced de-smoothed volatility estimates

As noted before, the analyst may be interested in the variance of the subsequent return, specifically accounting for uncertainty in the parameter estimates. The discussed approaches apply with a few modifications. The strategy will be to estimate the non-central second moment and expected return separately, then compute $E\left[r^2\right] - E\left[r\right]^2$. The second non-central moment can then be calculated conditional on parameter values and integrated to the unconditional moment.

First, note the plug-in estimator for the mean factor return:

$$
\hat{\mu}_f = \frac{1}{T}\sum_{t\in 1:T}\tilde{f}_t
$$

As before, let $\hat{\mu}^{LR}_{\tilde{f}}$ correspond with the long-run estimate.

Next compute the expectation of the non-central second moment:

$$\begin{aligned}
E\left[x_t^2\right] &= \int_{\Theta} \left(V\left(x_t|\Theta,\cdot\right) + \left[\left(\hat{\mu}_{\tilde{f}}^{LR}\right)'\tilde{\beta}\right]^2 \right) p\left(\Theta|\Pi\right) d\Theta \\
&= \int_{\Theta} \left(\tilde{\beta}'\hat{\Sigma}_{\tilde{F}}\tilde{\beta} + \frac{1}{\tau_x\tau_y}\left(\frac{\nu}{\nu-2}\right) + \frac{1}{\Delta t\tau_y} + \left[\left(\hat{\mu}_{\tilde{f}}^{LR}\right)'\tilde{\beta}\right]^2 \right) p\left(\Theta|\Pi\right) d\Theta \\
\hat{\mu}_{x^2} &= \frac{1}{J}\sum_{j\in 1:J}\left[\tilde{\beta}_j'\hat{\Sigma}_{\tilde{F}}\tilde{\beta}_j + \frac{1}{\tau_{xj}\tau_{yj}}\left(\frac{\nu_j}{\nu_j-2}\right) + \frac{1}{\Delta t\tau_{yj}} + \left[\left(\hat{\mu}_{\tilde{f}}^{LR}\right)'\tilde{\beta}_j\right]^2 \right]
\end{aligned}$$

The long-run unconditional expected return is given as:

$$\begin{aligned}
E\left[x_t|\cdot\right] &= E\left[x_t|\Theta,\Pi\right] \\
&= \int_{\Theta} \left(\hat{\mu}_{\tilde{f}}^{LR}\right)'\tilde{\beta} p\left(\Theta|\Pi\right) d\Theta \\
\hat{\mu}_x &= \frac{1}{J}\sum_{j\in 1:J}\left(\hat{\mu}_{\tilde{f}}^{LR}\right)'\tilde{\beta}_j
\end{aligned}$$

Then the variance of the subsequent month's return given all available model information is given by:

$$\hat{\varsigma}_x^2 = \hat{\mu}_{x^2} - \hat{\mu}_x^2$$

3.3 Reporting variance

The reporting variance must account for the smoothing properties of the reporting process. Let $\tilde{\phi}$ represent the $P+\Delta t$ terms of the smoothing process inclusive of any imposed restricts (see the model doc). Note that the reporting variance is already de-smoothed. Then because the de-smoothed returns are unconditionally iid, the variance using the above enhanced methodology is given by:

$$\begin{aligned}
\sigma_y^2 &= \int_{\Theta} V\left(y_t|\Theta,\Pi\right) p\left(\Theta|\Pi\right) d\Theta \\
&= \int_{\Theta} \left(\tilde{\phi}'\tilde{\phi}\left[\tilde{\beta}'\hat{\Sigma}_{\tilde{F}}\tilde{\beta} + \frac{1}{\tau_x\tau_y}\left(\frac{\nu}{\nu-2}\right)\right] + \frac{1}{\tau_y} \right) p\left(\Theta|\Pi\right) d\Theta \\
\hat{\sigma}_y^2 &= \frac{1}{J}\sum_{j\in 1:J}\left(\tilde{\phi}_j'\tilde{\phi}_j\left[\tilde{\beta}_j'\hat{\Sigma}_{\tilde{F}}\tilde{\beta}_j + \frac{1}{\tau_{xj}\tau_{yj}}\left(\frac{\nu_j}{\nu_j-2}\right)\right] + \frac{1}{\tau_{yj}} \right)
\end{aligned}$$

The above represents the preferred metric for calculating the expected long-run variance of the returns. Note that the idiosyncratic portion is given by:

$$\hat{\sigma}_{\varepsilon}^2 = \frac{1}{J}\sum_{j\in 1:J}\left(\tilde{\phi}_j'\tilde{\phi}_j\frac{1}{\tau_{xj}\tau_{yj}}\left(\frac{\nu_j}{\nu_j-2}\right) + \frac{1}{\tau_{yj}} \right)$$

3.4 Accounting for estimate uncertainty when calculating the reporting variance

Analysts may seek the variance of the subsequent period's return as opposed to the expected long-run variance of the series. As with the de-smoothed returns, this strategy computes the non-central second moment.

Let $t[s]$ represent the mapping of s to t. The expectation of a moving average process is given by:

$$\begin{aligned} E[y_s|\Theta] &= \sum_{p\in 1:(P+\Delta t)} \tilde{\phi}_p \mu'_{\tilde{f}}\beta \\ &= \Delta t \mu'_{\tilde{f}}\beta \end{aligned}$$

Similarly, the long-run unconditional expected returns are given by:

$$\begin{aligned} E[y_s|\Pi] &= \int_\Theta E[y_s|\Theta,\Pi]\, p(\Theta|\Pi)\, d\Theta \\ &= \Delta t \int_\Theta \left(\hat{\mu}_{\tilde{f}}^{LR}\right)' \tilde{\beta} p(\Theta|\Pi)\, d\Theta \\ \hat{\mu}_x &= \frac{1}{J}\sum_{j\in 1:J} \left(\hat{\mu}_{\tilde{f}}^{LR}\right)' \tilde{\beta}_j \end{aligned}$$

The last equation follows from the restriction that $\sum \tilde{\phi}_p = \Delta t$. Then the non-central second moment is given by:

$$\begin{aligned} E\left[y_s^2|\Pi\right] &= \int_\Theta \left(V(y_s|\Theta,\Pi) + \left[(\Delta t)\left(\hat{\mu}_{\tilde{f}}^{LR}\right)' \tilde{\beta}\right]^2 \right) p(\Theta|\Pi)\, d\Theta \\ &= \int_\Theta \left(\tilde{\phi}'\tilde{\phi}\left[\tilde{\beta}'\hat{\Sigma}_{\tilde{F}}\tilde{\beta} + \frac{1}{\tau_x\tau_y}\left(\frac{\nu}{\nu-2}\right)\right] + \frac{1}{\tau_y} + \left[(\Delta t)\left(\hat{\mu}_{\tilde{f}}^{LR}\right)' \tilde{\beta}\right]^2 \right) p(\Theta|\cdot)\, d\Theta \\ \hat{\mu}_{y^2} &= \frac{1}{J}\sum_{j\in 1:J} \left(\tilde{\phi}_j'\tilde{\phi}_j\left[\tilde{\beta}_j'\hat{\Sigma}_{\tilde{F}}\tilde{\beta}_j + \frac{1}{\tau_{xj}\tau_{yj}}\left(\frac{\nu_j}{\nu_j-2}\right)\right] + \frac{1}{\tau_{yj}} + \left[(\Delta t)\left(\hat{\mu}_{\tilde{f}}^{LR}\right)' \tilde{\beta}_j\right]^2 \right) \end{aligned}$$

Then the variance of the next return observation is given as:

$$\hat{\varsigma}_y^2 = \hat{\mu}_{y^2} - \hat{\mu}_y^2$$

4 Appendix

4.1 Use of long-run factor history to enhance volatility estimates

The following analysis is an excerpt of an earlier memo, and considers two approaches to computing the historical volatility of an asset. The first method uses the standard sample variance. The second method computes the factor loading via an OLS regression, and then uses this loading to compute the variance of the asset from the historical factor returns.

The below analysis shows that the factor augmented methodology is more efficient in most realistic scenarios.

4.1.1 Setup

Consider a DGP as follows:

$$\begin{aligned}
r_t =& \alpha + \beta f_t + \varepsilon_t \\
f_t \sim& N\left(\mu_f, \sigma_f^2\right) \\
\varepsilon_t \sim& N\left(0, \sigma_\varepsilon^2\right)
\end{aligned}$$

Here f_t is a factor return, ε_t is a white noise residual, r_t is an asset return, and α and β are coefficients. Suppose we have S observations of returns and T observations of factors.

4.1.2 Variance of the sample variance

Call this method the direct method for computing the sample volatility.These results are standard: the distribution of the sum of S squared standard normally distributed variables is χ_S^2 distributed. Then:

$$\begin{aligned}
S\left(\frac{1}{S}\sum_{s\in 1:S}\frac{\left(r_s-\mu_r\right)^2}{\beta^2\sigma_f^2+\sigma_\varepsilon^2}\right) =& S\frac{\hat{\sigma}^2\left(r_S\right)}{\beta^2\sigma_f^2+\sigma_\varepsilon^2} \\
\frac{S}{\beta^2\sigma_f^2+\sigma_\varepsilon^2}\hat{\sigma}^2\left(r_S\right) \sim& \chi_S^2 \\
\implies var\left(\hat{\sigma}^2\left(r_S\right)\right) =& \frac{2\left(\beta^2\sigma_f^2+\sigma_\varepsilon^2\right)^2}{S}
\end{aligned}$$

4.1.3 Factor-augmented methodology

Now consider the factor methodology. For this methodology, the variance is given by:

$$var\left(\hat{\beta}^2\hat{\sigma}_f^2\right)+\hat{\sigma}_\varepsilon^2$$

The general approach will be to compute the variance of the sample variance of each component. To account for the product of the two variables in the first term, approximate each as a normal distribution and apply the multivariate delta method.

The variance of the sample beta is given by the standard formula:

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma_{\varepsilon}^{2}}{S\sigma_{f}^{2}}\right)$$

Note if the normality assumption is relaxed, the above formula may still hold asymptotically. Using a non-central χ^2 distribution, we get:

$$\begin{aligned}\left(\frac{\sigma_{\varepsilon}^{2}}{S\sigma_{f}^{2}}\right)^{-1}\hat{\beta}^{2} &\sim NC\chi_{1}^{2}\left(\beta^{2}\left(\frac{\sigma_{\varepsilon}^{2}}{S\sigma_{f}^{2}}\right)^{-1}\right)\\ var\left(\hat{\beta}^{2}\right) &= 2\left(\frac{\sigma_{\varepsilon}^{2}}{S\sigma_{f}^{2}}\right)^{2}+4\beta^{2}\left(\frac{\sigma_{\varepsilon}^{2}}{S\sigma_{f}^{2}}\right)\end{aligned}$$

The following is a known property of the non-central χ^2 distribution:

$$\begin{gathered}\lim_{\lambda\to\infty} NC\chi_{k}\left(\lambda\right) \sim N\left(\mu, \sigma^{2}\right)\\ s.t.\\ \mu = k+\lambda\\ \sigma^{2} = 2\left(k+2\lambda\right)\end{gathered}$$

Hence

$$\begin{aligned}\hat{\beta}^{2} &\approx N\left(\mu_{\beta^{2}}, \sigma_{\beta^{2}}^{2}\right)\\ &s.t.\\ \mu_{\beta^{2}} &\equiv \left(\frac{\sigma_{\varepsilon}^{2}}{S\sigma_{f}^{2}}\right)+\beta^{2}\\ \sigma_{\beta^{2}}^{2} &\equiv 2\left(\frac{\sigma_{\varepsilon}^{2}}{S\sigma_{f}^{2}}\right)^{2}+4\beta^{2}\left(\frac{\sigma_{\varepsilon}^{2}}{S\sigma_{f}^{2}}\right)\end{aligned}$$

Note that the approximation is likely to be more accurate for larger values of S, β, and σ_f^2, and smaller values of σ_ε^2. The variance of the factor sample variance is given by:

$$\begin{aligned}\frac{T}{\sigma_{f}^{2}}\sigma^{2}\left(f_{T}\right) &\sim \chi_{T}^{2}\\ \implies var\left(\hat{\sigma}_{f}^{2}\right) &= \frac{2\left(\sigma_{f}^{2}\right)^{2}}{T}\\ \hat{\sigma}_{f}^{2} &\rightsquigarrow N\left(\sigma_{f}^{2}, \sigma_{vf}^{2}\right)\\ &s.t.\\ \sigma_{vf}^{2} &\equiv \frac{2\left(\sigma_{f}^{2}\right)^{2}}{T}\end{aligned}$$

Note that both $\hat{\sigma}_f^2$ and $\hat{\beta}^2$ converge to finite means. Assume that the difference in history between f_T and f_S is enough that $\hat{\beta}^2$ is approximately uncorrelated with $\hat{\sigma}_f^2$. Then apply the multivariate delta method, slightly modified to account for the difference in sample sizes:

$$\begin{aligned} g\left(\hat{\beta}^2, \hat{\sigma}_f^2\right) =& \hat{\beta}^2 \hat{\sigma}_f^2 \\ \nabla g\left(\mu_{\beta^2}, \sigma_f^2\right) =& \begin{bmatrix} \sigma_f^2 \\ \mu_{\beta^2} \end{bmatrix} \\ \Sigma_{\beta^2 \sigma_f^2} =& \begin{bmatrix} \sigma_{\beta^2}^2 & \cdot \\ \cdot & \sigma_{vf}^2 \end{bmatrix} \\ \implies \hat{\beta}^2 \hat{\sigma}_f^2 \rightsquigarrow & N\left(\sigma_f^2 \mu_{\beta^2}, \nabla g' \Sigma_{\beta^2 \sigma_f^2} \nabla g\right) \\ =& N\left(\sigma_f^2 \mu_{\beta^2}, \left(\sigma_f^2\right)^2 \sigma_{\beta^2}^2 + \mu_{\beta^2}^2 \sigma_{vf}^2\right) \end{aligned}$$

Without correcting for the generated regressor property, the variance of the sample variance of the residuals is given by:

$$\frac{S}{\sigma_\varepsilon^2} \hat{\sigma}^2(\varepsilon) \sim \chi_S^2$$

Hence:

$$var\left(\hat{\sigma}_\varepsilon^2\right) = \frac{2\left(\sigma_\varepsilon^2\right)^2}{S}$$

Then combining the parts

$$\begin{aligned} var\left[\hat{\sigma}\left(\hat{\beta} f_T\right) + \hat{\sigma}_\varepsilon^2\right] &\approx \left(\sigma_f^2\right)^2 \sigma_{\beta^2}^2 + \sigma_{vf}^2 \mu_{\beta^2}^2 + \frac{2\left(\sigma_\varepsilon^2\right)^2}{S} \\ &= \left(\sigma_f^2\right)^2 \sigma_{\beta^2}^2 + \sigma_{vf}^2 \mu_{\beta^2}^2 + \frac{2\left(\sigma_\varepsilon^2\right)^2}{S} \end{aligned}$$

4.1.4 Comparison of convergence properties

Plugging in for the auxiliary variables, factor method has uncertainty of:

$$var\left[\hat{\sigma}\left(\hat{\beta} f_T\right) + \hat{\sigma}_\varepsilon^2\right] \approx \left(\sigma_f^2\right)^2 \left[2\left(\frac{\sigma_\varepsilon^2}{S\sigma_f^2}\right)^2 + 4\beta^2\left(\frac{\sigma_\varepsilon^2}{S\sigma_f^2}\right) + \frac{2}{T}\left(\frac{\sigma_\varepsilon^2}{S\sigma_f^2} + \beta^2\right)^2\right] + \frac{2\left(\sigma_\varepsilon^2\right)^2}{S}$$

For comparison:

$$var\left[\hat{\sigma}^2\left(r_S\right)\right] = \frac{2\left(\beta^2 \sigma_f^2\right)^2 + 4\sigma_\varepsilon^2 \beta^2 \sigma_f^2 + 2\left(\sigma_\varepsilon^2\right)^2}{S}$$

Subtracting the terms provides intuition:

$$var\left[\hat{\sigma}\left(\hat{\beta}f_T\right)+\hat{\sigma}_\varepsilon^2\right]-var\left[\hat{\sigma}^2\left(r_S\right)\right]\approx\left(\sigma_f^2\right)^2\left[2\left(\frac{\sigma_\varepsilon^2}{S\sigma_f^2}\right)^2+\frac{2}{T}\left(\frac{\sigma_\varepsilon^2}{S\sigma_f^2}+\beta^2\right)^2\right]-\frac{2\left(\beta^2\sigma_f^2\right)^2}{S}$$

If we drop all higher order terms, the difference is:

$$var\left[\hat{\sigma}\left(\hat{\beta}f_T\right)+\hat{\sigma}_\varepsilon^2\right]-var\left[\hat{\sigma}^2\left(r_S\right)\right]\approx\frac{2\left(\beta^2\sigma_f^2\right)^2}{T}-\frac{2\left(\beta^2\sigma_f^2\right)^2}{S}$$

The asymptotics generally indicate that the factor calculation methodology, augmented by the historical factor returns, will be more precise than the direct method for computing the sample variance. The relationship is not guaranteed due to the higher order terms, use of asymptotics, and other approximations, but it seems likely to hold in most cases where $T \gg S$.

What is the cost of the improved efficiency? The mean estimate of the variance using the factor method is given by:

$$\begin{aligned}E\left[\hat{\sigma}\left(\hat{\beta}f_T\right)+\hat{\sigma}_\varepsilon^2\right]\approx&\sigma_f^2\left(\frac{\sigma_\varepsilon^2}{S\sigma_f^2}+\beta^2\right)+\sigma_\varepsilon^2\\=&\frac{(S+1)}{S}\sigma_\varepsilon^2+\sigma_f^2\beta^2\end{aligned}$$

The factor estimator is therefore slightly biased. The bias seems unlikely to be material.

4.1.5 Monte-carlo analysis

The above statistics are easy to simulate. Each parameter set is computed over 100,000 simulations with $T = 1000$, and $\alpha = 0$. Other parameters are scenario dependent. The *std* (direct) and *std* (augmented) columns give the direct and factor-augmented standard deviation of the variance estimates. The results indicate the following conclusions:

1. The asymptotic approximation seems to do well at capturing the dynamics of the monte-carlo simulations, validating the above analytical exercise.

2. The augmented method can dramatically improve the estimates of asset variance.

3. In cases where the augmented method does not improve the estimates, any loss in efficiency seems minor.

S	β	σ_f^2	σ_ε^2	*std* (direct)	*std* (augmented)	asympt. *std* (direct)	asympt. *std* (augmented)
12	1	2	1	1.3	1.0	1.2	0.9
12	5	5	1	53.8	9.3	51.4	8.6
12	1	1	10	4.7	5.0	4.5	4.6
24	1	2	1	0.9	0.7	0.9	0.7
24	5	5	1	37.1	7.4	36.4	7.2
24	1	1	10	3.2	3.3	3.2	3.2