arXiv:1505.00829v2 [q-fin.ST] 22 Jun 2015

Inference on the Sharpe ratio via the upsilon distribution

Steven E. Pav *

June 24, 2015

Abstract

The upsilon distribution, the sum of independent chi random variates and a normal, is introduced. As a special case, the upsilon distribution includes Lecoutre's lambda prime distribution. [22] The upsilon distribution finds application in Frequentist inference on the Sharpe ratio, including hypothesis tests on independent samples, confidence intervals, and prediction intervals, as well as their Bayesian counterparts. These tests are extended to the case of factor models of returns.

1 Introduction

For $k \geq 1$, given a k-vector $[t_1, t_2, \ldots, t_k]^\top$ and k-vector of positive reals $[\nu_1, \nu_2, \ldots, \nu_k]^\top$, if

$$y = \sum_{j=1}^{k} t_i \sqrt{\frac{\chi_i^2}{\nu_i}} + Z, \tag{1}$$

where $Z \sim \mathcal{N}(0, 1)$ independently of $\chi_i^2 \sim \chi^2(\nu_i)$, which are independent, then we say y follows an upsilon distribution[1] with coefficient $\boldsymbol{t} = [t_1, t_2, \ldots, t_k]^\top$, and degrees of freedom $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_k]^\top$. Let this be written as $y \sim \Upsilon(\boldsymbol{t}, \boldsymbol{\nu})$. The summands are reminiscent of Nakagami variates, but the latter may only take positive weights by definition, whereas the elements of $\boldsymbol{t}$ are unrestricted. [14]

Lecoutre's *lambda prime* distribution appears as a specific case of the upsilon, with $k = 1$. [15, 22] It should be stressed that the use of the lambda prime for inference with the t-statistic was first described independently by Akahira and Lecoutre and the results here should be viewed as extensions of that work. [1, 2, 16] Lecoutre does not consider the Sharpe ratio, which is simply the t statistic rescaled, nor does he consider the multiple independent samples case, which is first described here.

*spav@alumni.cmu.edu

[1] I have chosen Υ since it vaguely looks like a 'y', half way between the 'x' of the chi summands and the 'z' of the normal. The '$\boldsymbol{t}$' is used for the coefficient since in practice the elements of this will typically be t-statistics.

1.1 Computing the upsilon distribution

It is customary, when introducing a new probability distribution, to quote its density, and describe cases in which it arises in real settings, how to interpret its parameters, *etc.* The upsilon is somewhat idiosyncratic in that it is not a distribution of primary interest, however, and only useful as a kind of 'dual' to the distribution of the Sharpe ratio. [27] As such, even though one could write down the density, as the convolution of the densities of the normal and the chi distributions, doing so would not much help in practical computation of the density. Moreover, inference on the parameters $\boldsymbol{t}$ is not typically required.

However, computation of the cumulative distribution and the quantile of the upsilon are of interest in performing inference on the Sharpe ratio. These are complicated to write down: the CDF is the integral of the ugly convolution alluded to above, and the quantile requires inverting that integral. It would not seem a fruitful pursuit to find highly accurate methods of computing these functions, since the tests described herein for inference on the Sharpe ratio are only exact in the case of Gaussian asset returns, which is a poor approximation for most real assets. Instead, methods for approximating the distribution and quantile function are likely to be sufficient for most users.

Approximating these functions via the Edgeworth and Cornish-Fisher expansions is fairly simple, and relies only on knowing the cumulants of the upsilon distribution. [3, 17, 11] These are readily available since the cumulants commute with addition for the case of independent random variables. This means one only needs to know the cumulants of the normal distribution[2] and the chi distribution. Users of the **R** language have access to the approximate density, distribution, and quantile functions via the **sadists** package available from CRAN. [23, 21]

For concreteness, let κ_i be the i^{th} raw cumulant of the upsilon distribution with coefficient $\left[t_1, t_2, \ldots, t_k\right]^{\top}$, and degrees of freedom $\left[\nu_1, \nu_2, \ldots, \nu_k\right]^{\top}$. Note that κ_1 is the expected value, and κ_2 is the variance of this distribution. Then

$$\begin{aligned}
\kappa_1 &= \sqrt{2}\sum_{1\le j\le k}\frac{t_j}{\sqrt{\nu_j}}\frac{\Gamma\left(\frac{\nu_j+1}{2}\right)}{\Gamma\left(\frac{\nu_j}{2}\right)},\\
\kappa_2 &= 1 + \sum_{1\le j\le k}\frac{{t_j}^2}{\nu_j}\left(\nu_j - 2\left(\frac{\Gamma\left(\frac{\nu_j+1}{2}\right)}{\Gamma\left(\frac{\nu_j}{2}\right)}\right)^2\right),\\
\kappa_3 &= \sum_{1\le j\le k}\frac{{t_j}^3}{{\nu_j}^{3/2}}\sqrt{2}\frac{\Gamma\left(\frac{\nu_j+1}{2}\right)}{\Gamma\left(\frac{\nu_j}{2}\right)}\left(1 - 2\nu_j + 4\left(\frac{\Gamma\left(\frac{\nu_j+1}{2}\right)}{\Gamma\left(\frac{\nu_j}{2}\right)}\right)^2\right),\ldots
\end{aligned} \tag{2}$$

In practice, it is easier to first compute the raw moments of the chi distribution with ν_j degrees of freedom, since the i^{th} raw moment is simply

$$2^{i/2}\frac{\Gamma\left(\frac{\nu_j+1}{2}\right)}{\Gamma\left(\frac{\nu_j}{2}\right)}.$$

[2] The only distribution with a finite number of non-zero cumulants.

One then can translate the raw moments of the chi variables to raw cumulants. The cumulants of the upsilon can then be computed using properties of the cumulants (invariance, additivity and homogeneity).

2 The Sharpe ratio

The Sharpe ratio is the most commonly used metric of the *historical* performance of financial assets. t is defined as

$$\hat{\zeta} =_{\mathrm{df}} \frac{\hat{\mu} - r_0}{\hat{\sigma}}, \tag{3}$$

where $\hat{\mu}$ is the historical, or sample, mean return of the mutual fund, $\hat{\sigma}$ is the sample standard deviation of returns, and r_0 is some fixed *risk-free* or *disastrous* rate of return. Under Sharpe's original definition, r_0 was equal to zero. [26] One typically uses the vanilla sample mean and the Bessel-corrected sample standard deviation in computing the ratio,

$$\hat{\mu} =_{\mathrm{df}} \sum_{1 \le t \le n} x_t, \quad \hat{\sigma} =_{\mathrm{df}} \sqrt{\frac{\sum_{1 \le i \le n} \left(x_t - \hat{\mu}\right)^2}{n-1}}, \tag{4}$$

where x_t are the returns of the asset at time t. Herein, *Sharpe ratio* will refer to this quantity, computed from sample statistics, whereas *signal-noise ratio* will refer to the analagously defined population parameter,

$$\zeta =_{\mathrm{df}} \frac{\mu - r_0}{\sigma}. \tag{5}$$

In general, hats will be placed over quantities to denote population estimates.

Interpretation of the Sharpe ratio relies on a very simple model of returns: identically distributed, homoskedastic, and unconditional on any state variables. A more realistic model of real assets allows attribution of returns to contemporaneous observable *factors*. In the general case one attributes the returns of the asset in question as the linear combination of l factors, one of which is typically the constant one:

$$x_t = \beta_0 1 + \sum_{i}^{l-1} \beta_i f_{i,t} + \epsilon_t, \tag{6}$$

where $f_{i,t}$ is the value of some i^{th} 'factor' at time t, and the innovations, ϵ, are assumed to be zero mean, and have standard deviation σ. Here we have forced the zeroth factor to be the constant one, $f_{0,t} = 1$. [25]

Given n observations, let F be the $n \times l$ matrix whose rows are the observations of the factors (including a column that is the constant 1), and let $\boldsymbol{x}$ be the n length column vector of returns; then the multiple linear regression estimates are

$$\hat{\boldsymbol{\beta}} =_{\mathrm{df}} \left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1} \mathsf{F}^{\top} \boldsymbol{x}, \qquad \hat{\sigma} =_{\mathrm{df}} \sqrt{\frac{\left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)^{\top} \left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)}{n-l}}. \tag{7}$$

We can then define a Sharpe ratio for factor models as follows: let $\boldsymbol{v}$ be some non-zero vector, and let r_0 be some risk-free, or disastrous, rate of return. Then define

$$\hat{\zeta}_g =_{\mathrm{df}} \frac{\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{v} - r_0}{\hat{\sigma}}. \tag{8}$$

Typically in the study of equity strategy returns, one chooses $\boldsymbol{v} = \boldsymbol{e}_0$, the vector of all zeros except a one corresponding to the intercept term.

There are numerous candidates for the factors, and their choice should depend on the return series being modeled. For example, one would choose different factors when modeling the returns of a single company versus those of a broad-market mutual fund versus those of a market-neutral hedge fund, *etc.* Moreover, the choice of factors might depend on the type of analysis being performed. For example, one might be trying to 'explain away' the returns of one investment as the returns of another investment (presumably one with smaller fees) plus noise. Alternatively, one might be trying to establish that a given investment has idiosyncratic 'alpha' (*i.e.*, β_0) without significant exposure to other factors, either because those other factors are some kind of benchmark, or because one believes they have zero expectation in the future. [6, 13]

2.1 The Sharpe ratio under Gaussian returns

While normality of returns is a *terrible* model for most market instruments [7], it is a terribly convenient model. We will adhere to this terrible model and assume returns are unconditionally *i.i.d.* Gaussian, *i.e.*, $x_t \sim \mathcal{N}\left(\mu, \sigma^2\right)$. The equivalent assumption for the factor model, Equation 6, is that the innovations are *i.i.d.* Gaussian, *i.e.*, $\epsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)$.

Under Gaussian returns, the Sharpe ratio follows a t distribution up to scaling. That is

$$\sqrt{n}\hat{\zeta} \sim \frac{Z + \sqrt{n}\zeta}{\sqrt{\chi^2_{n-1}/(n-1)}}, \tag{9}$$

where Z is a standard normal independent of the chi-square χ^2_{n-1} which has $n-1$ degrees of freedom. The Sharpe ratio under the factor model, defined in Equation 8, has a considerably more complicated distribution:

$$\left(\boldsymbol{v}^{\top}\left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\boldsymbol{v}\right)^{-1/2}\hat{\zeta}_g \sim \frac{Z + \left(\boldsymbol{v}^{\top}\left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\boldsymbol{v}\right)^{-1/2}\zeta_g}{\sqrt{\chi^2_{n-l}/(n-l)}}, \tag{10}$$

where $\zeta_g = \frac{\boldsymbol{\beta}^{\top}\boldsymbol{v} - r_0}{\sigma}$ is the population analogue of $\hat{\zeta}_g$, and again Z is a standard normal independent of χ^2_{n-l}, a chi-square with $n-l$ degrees of freedom. For the exact results described in this paper, one must make further unrealistic assumptions regarding the factor model, namely that the factor returns are essentially deterministic.

3 Frequentist inference on the Sharpe ratio

Assuming normal returns, the connection between the Sharpe ratio and the upsilon distribution becomes apparent when one rearranges the terms. Conditional on observing $\hat{\zeta}$,

$$\sqrt{n}\zeta = Z + \sqrt{n}\hat{\zeta}\sqrt{\chi^2_{n-1}/(n-1)}.$$

(Since Z is an unobserved standard normal random variable, we will often flip its sign without renaming it.) Thus conditional on $\hat{\zeta}$, $\sqrt{n}\zeta$ takes an upsilon distribution with $k = 1$, $\boldsymbol{t} = \left[\sqrt{n}\hat{\zeta}\right]$ and degrees of freedom $\boldsymbol{\nu} = [n - 1]$. The same rearrangement can be performed for multiple independent samples, leading to tests of linear combinations of the population parameters.

3.1 Hypothesis tests

The classical one-sample test for the signal-noise ratio could be interpreted in terms of the upsilon, although it is far more typical to rely on the t-distribution instead. The typical way to test

$$H_0 : \zeta = \zeta_0 \quad \text{versus} \quad H_1 : \zeta > \zeta_0$$

at the α level is to reject if the statistic $t = \sqrt{n}\hat{\zeta}$ is greater than $t_{1-\alpha}\left(\delta_0, n-1\right)$, the $1-\alpha$ quantile of the non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter $\delta_0 = \sqrt{n}\zeta_0$. Alternatively one can perform this test by rejecting at the α level if $\sqrt{n}\zeta_0 < \Upsilon_\alpha\left(\sqrt{n}\hat{\zeta}, n-1\right)$, the α quantile of the upsilon with $\boldsymbol{t} = \left[\sqrt{n}\hat{\zeta}\right]$ and $\boldsymbol{\nu} = [n - 1]$.

For the one sample case, use of the upsilon is superfluous[3], so consider the case of testing on multiple independent samples. A two-sample test for equality of signal-noise ratio, given independent observations, appears, at first glance, to be related to the Behrens-Fisher problem, which has no known solution. [5] However, this hypothesis can be tested exactly. For $i = 1, 2$, given n_i *i.i.d.* draws from Gaussian returns from two assets with signal-noise ratios ζ_i, to test

$$H_0 : \zeta_1 = \zeta_2 \quad \text{versus} \quad H_1 : \zeta_1 > \zeta_2,$$

compute the sample Sharpe ratios, $\hat{\zeta}_i$. Again, rearrange to get

$$\zeta_i = \hat{\zeta}_i\sqrt{\frac{\chi_i^2}{n_i - 1}} + \frac{1}{\sqrt{n_i}}Z_i,$$

where the $Z_i \sim \mathcal{N}\left(0, 1\right)$ independently and independent of the chi-square random variables $\chi_i^2 \sim \chi^2\left(n_i - 1\right)$, which are independent.

Then under the null,

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\left(\zeta_1 - \zeta_2\right) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\left[\hat{\zeta}_1\sqrt{\frac{\chi_1^2}{n_1 - 1}} - \hat{\zeta}_2\sqrt{\frac{\chi_2^2}{n_2 - 1}}\right] + Z,$$

where $Z \sim \mathcal{N}\left(0, 1\right)$ independently of the $\chi_i^2 \sim \chi^2\left(n_i - 1\right)$. On the right hand side is an upsilon random variable. To perform the hypothesis test at the α level, compute the α quantile of the upsilon random variable with coefficient

$$\boldsymbol{t} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\left[\hat{\zeta}_1, -\hat{\zeta}_2\right]^\top,$$

[3] And coincides with the lambda prime distribution.

and degrees of freedom $\boldsymbol{\nu} = [n_1 - 1, n_2 - 1]^{\top}$. If this quantile is bigger than zero, reject H_0 in favor of H_1.

Note that this two sample test can be further generalized to a k independent sample test for a single equation involving signal-noise ratios. For $i = 1, 2, \ldots, k$, given n_i independent draws from Gaussian returns from k assets with signal-noise ratios ζ_i, to test, for fixed $a_1, a_2, \ldots, a_k, b$, to test the hypothesis

$$H_0 : \sum_i a_i \zeta_i = b \quad \text{versus} \quad H_1 : \sum_i a_i \zeta_i > b,$$

compute the sample Sharpe ratios, $\hat{\zeta}_i$. As noted previously,

$$\zeta_i = \hat{\zeta}_i \sqrt{\frac{\chi_i^2}{n_i - 1}} + \frac{1}{\sqrt{n_i}} Z_i,$$

and so, under the null,

$$\left(\sum_{i=1}^{k} \frac{a_i^2}{n_i}\right)^{-1/2} b = \left(\sum_{i=1}^{k} \frac{a_i^2}{n_i}\right)^{-1/2} \sum_{i=1}^{k} a_i \hat{\zeta}_i \sqrt{\frac{\chi_i^2}{n_i - 1}} + Z.$$

As previously, this random variable is an upsilon under the null, and so one rejects at the α level if the left hand side exceeds the α quantile of the appropriate upsilon distribution. Note that this formulation subsumes the $k = 1$ case.

The analagous tests for the case of *dependent* returns does not seem to be simply defined. Rather one should rely on the standard asymptotic computation of the Sharpe ratio based on the central limit theorem and delta method, then test based on asymptotic normality, as described by Leung and Wong. [12, 19, 18]

The k independent sample test can be applied to the case of Sharpe ratio under factor models as well.[4] For $i = 1, 2, \ldots, k$, given n_i independent draws from factor models with Gaussian errors on k assets, for fixed $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}, c$, to test the hypothesis

$$H_0 : \sum_i \frac{{\beta_i}^{\top} \boldsymbol{v_i}}{\sigma_i} = c \quad \text{versus} \quad H_1 : \sum_i \frac{{\beta_i}^{\top} \boldsymbol{v_i}}{\sigma_i} > c,$$

compute the sample factor model Sharpe ratios for each sample:

$$\hat{\zeta}_{g,i} = \frac{\hat{\beta}_i^{\top} \boldsymbol{v_i}}{\hat{\sigma}_i}.$$

Then check the probability that an upsilon distribution with coefficient

$$\boldsymbol{t} = \left(\sum_{i=1}^{k} \boldsymbol{v_i}^{\top} \left({\mathsf{F}_i}^{\top} \mathsf{F}_i\right)^{-1} \boldsymbol{v_i}\right)^{-1/2} \left[\hat{\zeta}_{g,1}, \hat{\zeta}_{g,2}, \ldots, \hat{\zeta}_{g,k}\right]^{\top},$$

[4] Again it must be stressed that the factor returns should be deterministic, as otherwise their variability would contribute to extra uncertainty in the test statistics.

and degrees of freedom $\boldsymbol{\nu} = [n_1 - l_1, n_2 - l_2, \ldots, n_k - l_k]^\top$ exceeds

$$c \left(\sum_{i=1}^{k} \boldsymbol{v_i}^\top \left(\mathsf{F}_i{}^\top \mathsf{F}_i \right)^{-1} \boldsymbol{v_i} \right)^{-1/2}.$$

3.2 Confidence and Prediction intervals

The one and two sample hypothesis tests can be rearranged to form confidence and prediction intervals. For example, a $1 - \alpha$ confidence interval on ζ has endpoints $[\zeta_l, \zeta_u]$ defined by quantiles of the upsilon distribution:

$$\begin{aligned} \sqrt{n}\zeta_l &= \Upsilon_{\alpha/2}\left(\left[\sqrt{n}\hat{\zeta}\right], [n-1]\right), \\ \sqrt{n}\zeta_u &= \Upsilon_{1-\alpha/2}\left(\left[\sqrt{n}\hat{\zeta}\right], [n-1]\right). \end{aligned} \tag{11}$$

By the same argument, an $\alpha/2$ confidence level for the factor model signal-noise ratio, ζ_g, is given by $[\zeta_{g,l}, \zeta_{g,u}]$ defined as

$$\begin{aligned} \left(\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}\right)^{-1/2} \zeta_{g,l} &= \Upsilon_{\alpha/2}\left(\left(\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}\right)^{-1/2} \hat{\zeta}_g, [n-l]\right), \\ \left(\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}\right)^{-1/2} \zeta_{g,u} &= \Upsilon_{1-\alpha/2}\left(\left(\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}\right)^{-1/2} \hat{\zeta}_g, [n-l]\right). \end{aligned} \tag{12}$$

Note that the confidence intervals use the $k = 1$ case of the upsilon, corresponding to the lambda prime distribution.

Suppose, based on a sample of size n_1, you observed $\hat{\zeta}_1$ for some asset stream. What can you expect of the Sharpe ratio for n_2 future observations? The Frequentist answers this question via a prediction interval, an interval which, conditional on $\hat{\zeta}_1$ and n_1, contains the Sharpe ratio of those future observations with some specified probability, under replication.[5] Prediction intervals are actually an application of the two independent sample hypothesis test of equality.

Suppose you observe $\hat{\zeta}_1$ on n_1 observations of normally distributed *i.i.d.* returns, then observe $\hat{\zeta}_2$ on n_2 observations from the same returns stream. As in the case of the two independent samples test, write

$$\hat{\zeta}_1 \sqrt{\chi_1^2 / (n_1 - 1)} + Z_1/\sqrt{n_1} = \zeta = \hat{\zeta}_2 \sqrt{\chi_2^2 / (n_2 - 1)} + Z_2/\sqrt{n_2}, \tag{13}$$

where the $Z_i \sim \mathcal{N}(0, 1)$, and the $\chi_i^2 \sim \chi^2(n_i - 1)$ are independent. Then, the probability that $\hat{\zeta}_2$ is less than some value, say y, is the probability that an upsilon with coefficient

$$\boldsymbol{t} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left[\hat{\zeta}_1, -y\right]^\top,$$

and degrees of freedom $\boldsymbol{\nu} = [n_1 - 1, n_2 - 1]^\top$, is less than zero.

[5] Typically 'prediction interval' is reserved for an interval around a single future observation, while 'tolerance interval' is used for multiple future observations. Our application is somewhat between these two.

Thus the prediction intervals are given by $[\zeta_{2,l}, \zeta_{2,u}]$, defined implicitly by

$$\begin{aligned} 0 &= \Upsilon_{\alpha/2}\left(\sqrt{\frac{n_1 n_2}{n_1+n_2}}\left[\hat{\zeta}_1, -\zeta_{2,l}\right]^\top, [n_1-1, n_2-1]^\top\right), \\ 0 &= \Upsilon_{1-\alpha/2}\left(\sqrt{\frac{n_1 n_2}{n_1+n_2}}\left[\hat{\zeta}_1, -\zeta_{2,u}\right]^\top, [n_1-1, n_2-1]^\top\right). \end{aligned} \tag{14}$$

To find these prediction intervals, one must resort to a numerical root finder applied to the cumulative distribution function of the upsilon. It should be noted that a more direct computation could be performed if one had access to the quantile function of Lecoutre's K-prime distribution. [15, 21] Manipulating Equation 13, one has

$$\hat{\zeta}_2 = \frac{\hat{\zeta}_1\sqrt{\chi_1^2/\left(n_1-1\right)} + \sqrt{\frac{n_1+n_2}{n_1 n_2}}Z}{\sqrt{\chi_2^2/\left(n_2-1\right)}}, \tag{15}$$

where Z is a standard normal independent of the two chi-squares. In Lecoutre's terminology, conditional on observing $\hat{\zeta}_1$, $\hat{\zeta}_2$ is distributed as $K'_{n_1-1,n_2-1}\left(\hat{\zeta}_1, \sqrt{\frac{n_1+n_2}{n_1 n_2}}\right)$

4 Bayesian inference on the Sharpe ratio

The Frequentist tests considered previously all have similar Bayesian counterparts. In the traditional development of Bayesian inference on a Gaussian distribution with unknown parameters, prior and posterior distributions are considered on the mean and variance, μ and σ^2, or the mean and *precision*, the latter defined as σ^{-2}. [10, 20] It is a simple task to reformulate these in terms of the signal-noise ratio, ζ, and some transform of, say, the variance.

One commonly used conjugate prior is the 'Normal-Inverse-Gamma', under which one has an unconditional inverse gamma prior distribution on σ^2 (this is, up to scaling, one over a chi-square), and, conditional on σ, a normal prior on μ. [10, 3.3] These can be stated as

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_0/2, m_0\sigma_0^2/2\right), \\ \mu\left|\sigma^2\right. &\propto \mathcal{N}\left(\mu_0, \sigma^2/n_0\right), \end{aligned} \tag{16}$$

where σ_0^2, m_0, μ_0 and n_0 are the hyper-parameters. Under this formulation, an noninformative prior corresponds to $m_0 = 0 = n_0$.

After observing n *i.i.d.* draws from a normal distribution, $\mathcal{N}(\mu, \sigma)$, say $x_1, x_2, \ldots, x_n$, let $\hat{\mu}$ and $\hat{\sigma}$ be the sample estimates from Equation 4. The posterior is then

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_1/2, m_1\sigma_1^2/2\right), \\ \mu\left|\sigma^2\right. &\propto \mathcal{N}\left(\mu_1, \sigma^2/n_1\right), \end{aligned} \tag{17}$$

where

$$n_1 = n_0 + n, \qquad \mu_1 = \frac{n_0\mu_0 + n\hat{\mu}}{n_1}, \tag{18}$$

$$m_1 = m_0 + n, \qquad \sigma_1^2 = \frac{m_0\sigma_0^2 + (n-1)\hat{\sigma}^2 + \frac{n_0 n}{n_1}\left(\mu_0 - \hat{\mu}\right)^2}{m_1}. \tag{19}$$

This commonly used model can be trivially modified to one on the variance and the signal-noise ratio, where the former is a nuisance parameter. Transforming Equation 16, we arrive at

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_0/2, m_0\sigma_0^2/2\right), \\ \zeta\left|\sigma^2\right. &\propto \mathcal{N}\left(\frac{\mu_0}{\sigma}, 1/n_0\right), \end{aligned} \tag{20}$$

Marginalizing out σ^2, we arrive at a upsilon prior

$$\sqrt{n_0}\zeta \propto \Upsilon\left(\sqrt{n_0}\zeta_0, m_0\right), \tag{21}$$

where $\zeta_0 = \mu_0/\sigma_0$. In this case the upsilon coincides with the lambda prime distribution. The marginal posterior can be written as

$$\sqrt{n_1}\zeta \propto \Upsilon\left(\sqrt{n_1}\zeta_1, m_1\right), \tag{22}$$

where

$$n_1 = n_0 + n, \qquad \zeta_1 = \frac{n_0\zeta_0\sigma_0 + n\hat{\zeta}\hat{\sigma}}{n_1\sigma_1}, \tag{23}$$

$$m_1 = m_0 + n, \qquad \sigma_1^2 = \frac{m_0\sigma_0^2 + (n-1)\,\hat{\sigma}^2 + \frac{n_0 n}{n_1}\left(\zeta_0\sigma_0 - \hat{\zeta}\hat{\sigma}\right)^2}{m_1}, \tag{24}$$

where $\hat{\zeta} = \hat{\mu}/\hat{\sigma}$.

The unattributed model can be generalized to the factor model by following the standard Bayesian regression analysis. Again, we are assuming that the factors $\boldsymbol{f}_t$ are deterministic. The Bayesian regression prior is typically stated as

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_0/2, m_0\sigma_0^2/2\right), \\ \boldsymbol{\beta}\left|\sigma^2\right. &\propto \mathcal{N}\left(\boldsymbol{\beta}_0, \sigma^2{\Lambda_0}^{-1}\right), \end{aligned} \tag{25}$$

where σ_0^2, m_0 are the Bayesian hyperparameters for the coefficient and degrees of freedom of the error term, while $\boldsymbol{\beta}_0$ is that for the regression coefficient, and Λ_0 parametrizes uncertainty in the regression coefficient.

Again assume one has n observations of $\boldsymbol{f}_t$, stacked row-wise in the $n \times l$ matrix, F, and the corresponding returns stacked in vector $\boldsymbol{x}$. Again define $\hat{\boldsymbol{\beta}} =_{\text{df}} \left(\mathsf{F}^\top\mathsf{F}\right)^{-1}\mathsf{F}^\top\boldsymbol{x}$ and $\hat{\sigma} =_{\text{df}} \sqrt{\left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)^\top\left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)(n-l)^{-1}}$. The posterior distribution is

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_1/2, m_1\sigma_1^2/2\right), \\ \boldsymbol{\beta}\left|\sigma^2\right. &\propto \mathcal{N}\left(\boldsymbol{\beta}_1, \sigma^2{\Lambda_1}^{-1}\right), \end{aligned} \tag{26}$$

where

$$\Lambda_1 = \Lambda_0 + \mathsf{F}^\top\mathsf{F}, \quad \boldsymbol{\beta}_1 = {\Lambda_1}^{-1}\left(\Lambda_0\boldsymbol{\beta}_0 + \mathsf{F}^\top\mathsf{F}\hat{\boldsymbol{\beta}}\right), \tag{27}$$

$$m_1 = m_0 + n, \qquad \sigma_1^2 = \frac{m_0\sigma_0^2 + (n-l)\,\hat{\sigma}^2 + \hat{\boldsymbol{\beta}}^\top\mathsf{F}^\top\mathsf{F}\hat{\boldsymbol{\beta}} + {\boldsymbol{\beta}_0}^\top\Lambda_0\boldsymbol{\beta}_0 - {\boldsymbol{\beta}_1}^\top\Lambda_1\boldsymbol{\beta}_1}{m_1}. \tag{28}$$

A non-informative prior corresponds to $\Lambda_0 = \mathsf{0}, \boldsymbol{\beta}_0 = \mathbf{0}, \sigma_0^2 = 0, m_0 = 0$.

We can collapse the prior or posterior 'along' the direction $\boldsymbol{v}$ via

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_i/2, m_i\sigma_i^2/2\right), \\ \boldsymbol{v}^\top\boldsymbol{\beta}\,\middle|\sigma^2 &\propto \mathcal{N}\left(\boldsymbol{v}^\top\boldsymbol{\beta}_i, \sigma^2\boldsymbol{v}^\top{\Lambda_i}^{-1}\boldsymbol{v}\right), \end{aligned} \tag{29}$$

where $i = 0$ for the prior and $i = 1$ for the posterior. As in the unattributed model, marginalizing out σ^2, we have a upsilon prior and posterior:

$$\left(\boldsymbol{v}^\top{\Lambda_i}^{-1}\boldsymbol{v}\right)^{-1/2}\frac{\boldsymbol{\beta}^\top\boldsymbol{v}}{\sigma} = \left(\boldsymbol{v}^\top{\Lambda_i}^{-1}\boldsymbol{v}\right)^{-1/2}\zeta_g \propto \Upsilon\left(\left(\boldsymbol{v}^\top{\Lambda_i}^{-1}\boldsymbol{v}\right)^{-1/2}\zeta_{g,i}, m_i\right), \tag{30}$$

where $\zeta_{g,i} =_{\text{df}} {\boldsymbol{\beta}_i}^\top\boldsymbol{v}/\sigma_i$.

4.1 Credible and posterior prediction intervals

One can construct credible intervals on the signal-noise ratio based on the posterior, so-called *posterior intervals*, via quantiles of the upsilon distribution. [10] For example, a $(1-\alpha)$ credible interval on ζ is given by

$$\frac{1}{\sqrt{n_1}}\left[\Upsilon_{\alpha/2}\left(\sqrt{n_1}\zeta_1, m_1\right), \Upsilon_{1-\alpha/2}\left(\sqrt{n_1}\zeta_1, m_1\right)\right], \tag{31}$$

where $\Upsilon_q\left(\boldsymbol{t}, \boldsymbol{\nu}\right)$ is the q quantile of the upsilon distribution with coefficient $\boldsymbol{t}$ and $\boldsymbol{\nu}$ degrees of freedom. For the case of noninformative priors (corresponding to $n_0 = m_0 = 0$), this becomes

$$\frac{1}{\sqrt{n}}\left[\Upsilon_{\alpha/2}\left(\sqrt{n}\sqrt{\frac{n}{n-1}}\hat{\zeta}, n\right), \Upsilon_{1-\alpha/2}\left(\sqrt{n}\sqrt{\frac{n}{n-1}}\hat{\zeta}, n\right)\right] \tag{32}$$

This is equivalent to the Frequentist confidence intervals given in 3.2, but replacing n for $n-1$ in the degrees of freedom for σ. Asymptotically, the Bayesian credible interval for an noninformed prior is the same as the Frequentist confidence interval. Alternatively, a Bayesian quant could argue that her Frequentist cousin is a confused Bayesian with prior $n_0 = 0, m_0 = -1, \sigma_0^2 = 0$.

For the factor model signal-noise ratio, the $(1-\alpha)$ credible interval on ζ_g is

$$\frac{1}{\sqrt{\boldsymbol{v}^\top\Lambda_1\boldsymbol{v}}}\left[\Upsilon_{\alpha/2}\left(\sqrt{\boldsymbol{v}^\top\Lambda_1\boldsymbol{v}}\zeta_{g,1}, m_1\right), \Upsilon_{1-\alpha/2}\left(\sqrt{\boldsymbol{v}^\top\Lambda_1\boldsymbol{v}}\zeta_{g,1}, m_1\right)\right]. \tag{33}$$

A Bayesian prediction interval is an interval which contains some fixed proportion of our posterior *belief* about the Sharpe ratio of some future observations. This is very similar to the Frequentist prediction interval, but dances around the issues of frequency and belief that separate the Frequentist and Bayesian.

As in the Frequentist case, our belief is that ζ_2, based on n_2 future observations, will be drawn from a compound non-central Sharpe ratio distribution with non-centrality parameter drawn from the posterior distribution. Effectively this is a 't of lambda prime' distribution, as was the case in the Frequentist setting. We can summarize this as

$$\begin{aligned} \sqrt{n_1}\zeta &\propto \Upsilon\left(\sqrt{n_1}\zeta_1, m_1\right), \\ \sqrt{n_2}\hat{\zeta}_2\,|\zeta &\propto t\left(\sqrt{n_2}\zeta, n_2 - 1\right), \end{aligned} \tag{34}$$

although this jumbles up the usual notation, since $\hat{\zeta}_2$ is a quantity one can eventually observe, not a population parameter. Nevertheless, the intent of these equations should be clear. The consequence is that we can find posterior prediction intervals, as in the Frequentist case, by using the upsilon distribution. A $1 - \alpha$ prediction interval on $\hat{\zeta}_2$ is given by $[\zeta_{lo}, \zeta_{hi}]$ where ζ_{lo} is chosen such that the upsilon distribution with coefficient

$$\boldsymbol{t} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} [\zeta_1, -\zeta_{lo}]^\top,$$

and degrees of freedom $\boldsymbol{\nu} = [m_1, n_2 - 1]^\top$, is less than zero with probability $\alpha/2$, and ζ_{hi} is defined *mutatis mutandis*.

5 An example: Momentum and the January effect

The methods above are illustrated using real asset returns, ignoring the assumption of Gaussian returns and deterministic factors. The monthly returns of 'the Market' portfolio, the small cap portfolio (known as SMB, for 'small minus big'), the value portfolio (known as HML, for 'high minus low'), and the momentum portfolio (known as UMD, for 'up minus down'), as tabulated by Kenneth French, were downloaded from Quandl. [9, 24] These are the celebrated three factor portfolios of Fama and French plus Carhart's momentum factor. [8, 4]

The data are distributed as monthly relative returns, quoted in percents. The Market return is quoted as an excess return, with the risk free rate subtracted out. The risk free rate is also tabulated. For our purposes, the raw market returns are needed, so the risk free rate is added back to the market returns. The set consists of 1056mo. of data, from Jan 1927 through Dec 2014.

Now we will consider the hypothesis that the momentum factor has smaller Sharpe ratio in January than the remainder of the year, under a factor model where we attribute returns to the remaining three factors, a so-called 'January effect'. We are effectively testing the *information coefficient* of UMD. Since the returns are independent[6], we can apply the independent two sample test.

Under the factor model, the Sharpe ratio for the Januaries was computed to be $-0.137\text{mo.}^{-1/2}$, while for non-Januaries, it was computed to be $0.299\text{mo.}^{-1/2}$. To test the null hypothesis that the factor model signal-noise ratios are equal, we consider quantiles of the upsilon distribution with coefficient $[-0.993, -2.175]^\top$ and degrees of freedom $[84, 964]^\top$. The 6 term Cornish Fisher approximate 0.005 and 0.995 quantiles were computed to be $[-5.751, -0.578]$, which does not contain zero. The 8 term Edgeworth approximation to the CDF was computed at zero to be 0.999, larger than $1 - \alpha/2$ with $\alpha = 0.01$. Thus we reject the null hypothesis of equality at the 0.01 level in favor of the alternative hypothesis that the UMD portfolio has smaller information coefficient in January than the remainder of the year.

[6] Independence ignores the small 'bounce' effect caused by any mispricing of assets in the portfolio at the time boundaries. This is likely to be small compared to the error in the other assumptions.

References

[1] Masafumi Akahira. A higher order approximation to a percentage point of the noncentral t-distribution. *Communications in Statistics - Simulation and Computation*, 24(3): 595–605, 1995. doi: 10.1080/03610919508813261. URL http://www.tandfonline.com/doi/abs/10.1080/03610919508813261.

[2] Masafumi Akahira, Michikazu Sato, and Norio Torigoe. On the new approximation to non-central t-distributions. *Journal of the Japan Statistical Society*, 25(1):1–18, 1995. URL https://www.jstage.jst.go.jp/article/jjss1995/25/1/25_1_1/_pdf.

[3] S. Blinnikov and R. Moessner. Expansions for nearly Gaussian distributions. *Astronomy and Astrophysics Supplement*, 130:193–205, May 1998. doi: 10.1051/aas:1998221. URL http://arxiv.org/abs/astro-ph/9711239.

[4] Mark M Carhart. On persistence in mutual fund performance. *Journal of Finance*, 52(1):57, 1997. URL http://www.jstor.org/stable/2329556.

[5] Ching-Hui Chang and Nabendu Pal. A revisit to the Behrens-Fisher problem: Comparison of five test methods. *Communications in Statistics - Simulation and Computation*, 37(6): 1064–1085, 2008. doi: 10.1080/03610910802049599. URL http://dx.doi.org/10.1080/03610910802049599.

[6] John Howland Cochrane. *Asset pricing*. Princeton Univ. Press, Princeton [u.a.], 2001. ISBN 0691074984. URL http://press.princeton.edu/titles/7836.html.

[7] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001. doi: 10.1080/713665670. URL http://personal.fmipa.itb.ac.id/khreshna/files/2011/02/cont2001.pdf.

[8] Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *Journal of Finance*, 47(2):427, 1992. URL http://www.jstor.org/stable/2329112.

[9] Kenneth French. Fama/French factors (monthly). Privately Published, 2014. URL http://www.quandl.com/KFRENCH/FACTORS_M.

[10] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. ISBN 9781439840955. URL http://books.google.com/books?id=ZXL6AQAAQBAJ.

[11] Stefan R. Jaschke. The Cornish-Fisher-expansion in the context of Delta - Gamma - Normal approximations. Technical Report 2001,54, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, 2001. URL http://www.jaschke-net.de/papers/CoFi.pdf.

[12] J. D. Jobson and Bob M. Korkie. Performance hypothesis testing with the Sharpe and Treynor measures. *The Journal of Finance*, 36(4):pp. 889–908, 1981. ISSN 00221082. URL http://www.jstor.org/stable/2327554.

[13] Andreas Krause. An overview of asset pricing models. 2001. URL http://people.bath.ac.uk/mnsak/Research/Asset_pricing.pdf.

[14] David Irvine Laurensen. *Indoor radio channel propagation modelling by ray tracing techniques*. PhD thesis, University of Edinburgh, 1994. URL http://www.see.ed.ac.uk/~dil/thesis.pdf.

[15] Bruno Lecoutre. Two useful distributions for Bayesian predictive procedures under normal models. *Journal of Statistical Planning and Inference*, 79:93–105, 1999. URL http://www.researchgate.net/publication/242997315_Two_useful_distributions_for_Bayesia

[16] Bruno Lecoutre. Another look at confidence intervals for the noncentral T distribution. *Journal of Modern Applied Statistical Methods*, 6(1):107–116, 2007. URL http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/telechargements/Lecoutre_Another_look

[17] Yoong-Sin Lee and Ting-Kwong Lin. Algorithm AS 269: High order Cornish-Fisher expansion. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(1):pp. 233–240, 1992. ISSN 00359254. URL http://www.jstor.org/stable/2347649.

[18] Pui-Lam Leung and Wing-Keung Wong. On testing the equality of multiple Sharpe ratios, with application on the evaluation of iShares. *Journal of Risk*, 10(3):15–30, 2008. URL http://www.risk.net/digital_assets/4760/v10n3a2.pdf.

[19] Andrew W. Lo. The Statistics of Sharpe Ratios. *Financial Analysts Journal*, 58(4), July/August 2002. URL http://ssrn.com/paper=377260.

[20] Kevin P. Murphy. Conjugate Bayesian analysis of the Gaussian distribution. Privately Published, 2007. URL http://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf.

[21] Steven E. Pav. *sadists: Some Additional Distributions*, 2015. URL https://github.com/shabbychef/sadists. R package version 0.2.1.

[22] Jacques Poitevineau and Bruno Lecoutre. Implementing Bayesian predictive procedures: The K-prime and K-square distributions. *Computational Statistics and Data Analysis*, 54(3):724–731, 2010. URL http://arxiv.org/abs/1003.4890v1.

[23] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL http://www.R-project.org. ISBN 3-900051-07-0.

[24] Raymond McTaggart and Gergely Daroczi. *Quandl: API Wrapper for Quandl.com*, 2015. URL http://CRAN.R-project.org/package=Quandl. R package version 2.6.0.

[25] Stephen A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976. URL http://linkinghub.elsevier.com/retrieve/pii/0022053176900466.

[26] William F. Sharpe. Mutual fund performance. *Journal of Business*, 39:119, 1965. URL http://ideas.repec.org/a/ucp/jnlbus/v39y1965p119.html.

[27] Min-ge Xie and Kesar Singh. Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review*, 81(1):3–39, 2013. ISSN 1751-5823. doi: 10.1111/insr.12000. URL http://dx.doi.org/10.1111/insr.12000.