

Better Hedge Ratios for Spread Trading

Paul Teetor

November 2011

1 Summary

This note discusses the calculation of hedge ratios for spread trading. It suggests that using *total least squares* (TLS) is superior to the normally-used technique of ordinary least squares (OLS). The advantage of TLS is that it accounts for the variance of both legs, unlike OLS which accounts for the variance of only one. The two techniques often give similar answers, but when they differ, the difference can be significant. I present an example from pairs trading where the TLS hedge ratio is 30% different than the usual OLS hedge ratio.

The note includes R code for calculating the TLS-based hedge ratio, which is quite simple for the two-asset case.

2 Introduction

When we trade a spread, we are long one market and short another. Let's call those markets P and Q , respectively. If we are long one unit of P , how many units of Q should we sell short? That quantity is called the *hedge ratio*.

Mathematically, the hedge ratio defines the spread when viewed as a time series, given by

$$s_t = p_t - \beta q_t \tag{1}$$

where s_t is the spread at time t , β is the hedge ratio, and p_t and q_t are the prices for P and Q , respectively, at time t .

The choice of hedge ratio is critical because it alters the behaviour and, ultimately, the profitability of the spread trade.

In this note, I am considering hedge ratios for which

- We calculate the hedge ratio from historical data; and
- We use historical prices, not returns.

This class may seem restrictive, but many important spreads fall into this class. An example is pairs trading, where the hedge ratio must be determined by the historical relationship between two stocks. Another example is currency spreads, where the hedge ratio is determined by regressing one currency's history against the other.

Spreads outside this class often have a hedge ratio that is either fixed or determined by external considerations and not dependent upon historical data. For example, calendar spreads are usually a simple 1:1 ratio; Treasury curve spreads may require an equal DV01 for both legs, which requires knowing spot prices but no historical data. (An Appendix catalogs several kinds of hedge ratio calculations.)

You may be surprised that I use prices, not returns, for determining hedge ratios. I find price-based hedge ratios necessary for practical spread trading because we are trading the difference of two prices ($p_t - \beta q_t$), not the difference of two returns. If your application depends upon returns-based hedging, you can still use TLS, but the exact sizing of your position will require additional steps. An Appendix outlines those steps.

3 Computing the Hedge Ratio

Which algorithm is best for computing hedge ratios? A common solution is ordinary least squares (OLS), but there are problems with that algorithm.

3.1 Problems with Ordinary Least Squares

In a previous note, I recommended using ordinary least squares (OLS) for computing hedge ratios by fitting X and Y to this simple linear regression and taking β_1 as the hedge ratio.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{2}$$

I used this approach myself until I notice something very peculiar: when I reversed the role of x and y , like this

$$x_i = \gamma_0 + \gamma_1 y_i + \epsilon_i \tag{3}$$

I expected to obtain a consistent hedge ratio. Specifically, I expected that $\gamma_1 = 1/\beta_1$, but it does not!

The OLS fit is not symmetrical because of a critical mathematical assumption behind the OLS algorithm; namely, that Y is a random variable and the sole source of variance, and that the X values are fixed constants with zero variance. The result is that the hedge ratio is not symmetrical.

This asymmetry is illustrated in Figure 1. In the first regression, $\text{lm}(Y \sim X)$, the algorithm minimizes the (squared) residuals of Y , shown by the dotted horizontal lines. The second regression, $\text{lm}(X \sim Y)$, minimizes the (squared) residuals of X , shown by the dotted vertical lines. The two regressions produce inconsistent hedge ratios, in the sense that long X / short Y is not the opposite of long Y / short X .

This is not good. From a trading standpoint, we want the algorithm to treat the two markets symmetrically. We do not want to ignore that variance (volatility) of one market in favor of the other. Furthermore, if the trader

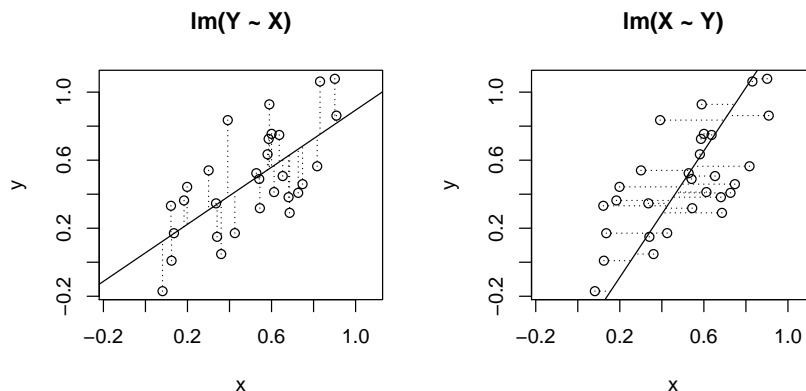


Figure 1: Residuals for Ordinary Least Squares

reverses his or her position, the reversed position should be the exact inverse of the original position.

Consequently, I no longer recommend using OLS for computing hedge ratios.

3.2 Total Least Squares

An alternative to ordinary least squares is *total least squares*, also called *orthogonal regression*¹. Total least squares (TLS) treats the X and Y variables symmetrically, so the resulting hedge ratio is consistent with respect to both markets. It accounts for the total variance of both variables, hence the “total” in “total least squares”.

Instead of minimizing either the residuals for X or the residuals for Y , TLS minimizes the distance from each (x_i, y_i) point to the regression line. This is illustrated in Figure 2. The distances to be minimized are vectors which are orthogonal to the regression line, hence the alternative name "orthogonal regression." That distance is unchanged if we exchange X and Y , so the resulting fit is unchanged, too. That gives the needed symmetry.

An Appendix shows the R code for the two-asset case. A principle components analysis (PCA) calculates both the slope and intercept of the regression, accounting for the variance in both legs.

¹Other names include *least circles regression* and *the errors-in-variables model*. The TLS technique has been independently “discovered” by many researchers over the several decades, leading to the profusion of names.

Total Least Squares

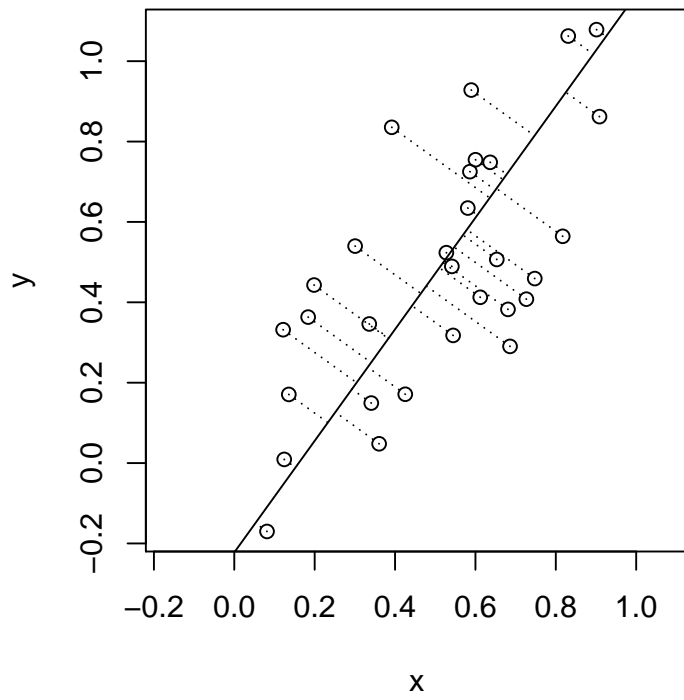


Figure 2: Residuals for Total Least Squares

4 Total Least Squares in R

Performing total least squares (TLS) regression in R is quite simple because principle components analysis (PCA) does the heavy lifting. Assume that p and q are the time series of prices for two markets, respectively, and you want to compute the hedge ratio, β , needed to form the spread $p_t - \beta q_t$.

```
r <- princomp( ~ p + q )
beta <- r$loadings[1,1] / r$loadings[2,1]
spread <- p - beta*q
```

You can easily compute the regression intercept, too, although that's not needed to form the spread.

```
inter <- r$center[2] - beta * r$center[1]
```

You can force a zero intercept by recentering the data around their means prior to the PCA.

```
p0 <- p - mean(p)
q0 <- q - mean(q)
r <- princomp( ~ p0 + q0 )
```

But note that recentering does *not* change the slope, so there may be no purpose in it.

One problem is that the PCA computes the point estimate of β , but neither the confidence interval nor the standard error, unlike `lm` which can provide regression statistics. The confidence interval could be useful for evaluating the effectiveness and stability of the hedge ratio. Forming the spread two ways and backtesting — once with the interval's lower limit, and once with the interval's upper limit — would give a sense of the spread's range of behaviour. Wildly different results, for example, could indicate an unstable ratio.

Unfortunately, I am unaware of how those regression statistics can be easily calculated from the `princomp` results. One possibility is using a statistical time-series bootstrap to form the confidence interval. Another possibility is to apply the TLS algorithms in Van Huffel and Vandewalle (1991) for the N -asset case with $N = 2$. Under the assumption of Normal distribution of residuals, those algorithms will yield regression statistics.

5 Example: The Russell 2000 / VIX spread

Suppose we notice an unexpected correlation between two ETFs: IWM, which tracks the Russell 2000 index, and VXZ, which tracks the VIX index. We think VXZ would be a useful hedge for an IWM position, and we want to calculate the hedge ratio.

This example calculates the hedge ratio using the two algorithms, OLS versus TLS; and it shows two different applications of each algorithm, one for IWM versus VXZ, and the other for the opposite spread, VXZ versus IWM. The purpose is to highlight the inconsistency of OLS and the consistency of TLS.²

We can use the popular `quantmod` package to create a two-column history of the ETF prices.

```
library(quantmod)

getCloses <- function(sym) {
  ohlc <- getSymbols(sym, from="2009-01-01", to="2011-01-01",
    auto.assign=FALSE, return.class="zoo")
  Cl(ohlc)
}

closes <- merge(IWM=getCloses("IWM"),
  VXZ=getCloses("VXZ"), all=FALSE)
```

We define two functions for computing hedge ratios, one using OLS and the other using TLS.

```
olsHedgeRatio <- function(p, q) {
  m <- lm(p ~ q)
  coef(m)[2]
}

tlsHedgeRatio <- function(p, q) {
  r <- princomp(~ p + q)
  r$loadings[1,1] / r$loadings[2,1]
}
```

(For the OLS calculation, we do not force a zero intercept because it's not logical that VXZ would necessarily be zero if IWM was zero.) Then we apply the functions to our data.

```
with(closes, {
  cat("OLS for IWM vs. VXZ =", olsHedgeRatio(IWM, VXZ), "\n")
  cat("OLS for VXZ vs. IWM =", olsHedgeRatio(VXZ, IWM), "\n")
  cat("TLS for IWM vs. VXZ =", tlsHedgeRatio(IWM, VXZ), "\n")
  cat("TLS for VXZ vs. IWM =", tlsHedgeRatio(VXZ, IWM), "\n")
})
```

²The code is available on github at

<https://github.com/pteetor/public/tree/master/BetterHedgeRatios>.

This is the output.

```
OLS for IWM vs. VXZ = -0.6583409
OLS for VXZ vs. IWM = -1.03481
TLS for IWM vs. VXZ = -0.7611657
TLS for VXZ vs. IWM = -1.313774
```

The OLS hedge ratios are inconsistent because $1 / -0.658 = -1.520$, not -1.035 , which is a substantial difference. The TLS ratios, however, are consistent because $1 / -0.761 = -1.314$.

Notice that the hedge ratios are *negative*. This may be unfamiliar to some readers. It occurs because these ETFs are negatively correlated. The spread is formed either by being long both legs or by being short both legs, unlike a positively correlated pair.

6 Acknowledgements

I am grateful to the R/Finance Conference 2011 committee for the opportunity to present this material during a lightning talk. I am also grateful to Jordan Erenrich for his useful comments on the first draft of this paper, and to Garrett See for identifying a serious typo.

7 References

The article by Petras and Podlubny is a very readable tutorial on the subject of total least squares, which they call “least circles” from the geometric interpretation of the technique. The book by Van Huffel and Vandewalle details both the mathematical logic, the statistical analysis, and the algorithmic implementation of TLS.

1. Petras and Podlubny (2010), "Least Squares or Least Circles? A comparison of classical regression and orthogonal regression", *Chance*, Vol. 23, No. 2 (Spring 2010)
2. Van Huffel and Vandewalle (1991), *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM Press (ISBN 0-89871-275-0)

A Appendix: R Code for Ordinary Least Squares Hedge Ratio

I do not recommend using ordinary least squares (OLS) for computing hedge ratios, but here is the R code, in case you are interested. Assume that p and q are the time series of prices for two markets, respectively, and you want to compute the hedge ratio, β , needed to form the spread $p_t - \beta q_t$.

For many spreads, there is a natural, economic constraint: if one leg has zero value, the other leg should have zero value, too. For example, if the price of gasoline is zero, the price of heating oil should be zero or, at least, very close to zero.

We can use this constraint to eliminate a source of variability in the calculation: we force a zero intercept for the linear regression. In R, that is done by adding "+ 0" to the model formula.

```
m <- lm(p ~ q + 0)      # Force zero intercept
beta <- coeffs(m)[1]    # First coeff is slope
```

For some markets, however, a price of zero is either nonsensical or even impossible. Examples are Treasury futures and volatility (VIX) futures. In those cases, a zero intercept would distort the result, so we do not force one.

```
m <- lm(p ~ q)          # Allow non-zero intercept
beta <- coeffs(m)[2]    # Second coeff is slope
```

Regression statistics are easily obtained by either `summary(m)` or `confint(m)`, depending upon your needs.

B Appendix: Some Varieties of Hedge Ratios

There are many types of spreads, and the hedge ratio calculation can be quite different, depending on the spread. This Appendix distinguishes a few varieties of hedge ratios, among many.

Fixed vs. Variable vs. Time-Varying

Some hedge ratios are fixed by *a priori* considerations; some are variable and computed (once) from market data; and some vary of over time.

A example of a fixed spread is the spread between Fed Funds futures and Eurodollar futures. The Fed Funds valBP is \$41.66, and the Eurodollar valBP is \$25.00. A market-neutral spread is 3 Fed Funds for every 5 Eurodollar futures, giving each leg a valBP of \$125.00. We could express that hedge ratio as 3:5 or 1:1.666 or $\beta = 1.666$, and that ratio does not vary over time.

An example of a variable spread is IBM vs Microsoft (MSFT) in the stock market. What is the correct number of shares to trade? The answer depends

upon the historical relationship between those two stocks, and the hedge ratio must be computed from the historical data.

Some hedge ratios are highly dynamic and can fluctuate significantly over time. If your hedge ratio is time-varying, then β should be subscripted to become β_t , in which case the spread becomes:

$$s_t = p_t - \beta_t q_t \tag{4}$$

State-space models such as the Kalman filter are one way to calculate time-varying hedge ratios.

Time-varying hedge ratios are an important, interesting topic, but beyond the scope of this note, so I assume here that your hedge ratio, β , is fixed and does not vary significantly over time.

Historical vs. Spot (Instantaneous) Data

If a hedge ratio is calculated from market data, it might use *historical data* or it might use *spot data*, meaning the market quotes available right now. The regressions described in this note use historical data, such as the two year's data used in the example of IWM *versus* VXZ, above.

An example of using spot data would be a market-neutral spread between two Treasury bonds. The DV01 (or duration) can be calculated using spot prices and does not require historical data. Another example is a spread between a stock and its options based on a delta-neutral hedge. The option deltas are calculated from their market price and an assumed (spot) volatility.

Choice of metric: price, returns, volatility, or sensitivity

The regression techniques, OLS and TLS, fit their models to historical data. The question is, which data? In my experience, academics automatically assume regression on *returns*, and practitioners (traders) automatically assume regression on *price*.

For my trading, I use regression on price. I need to model and control the P&L carefully, and that is easier with a price-based regression. Regression on returns is possible but results in a time-varying hedge ratio. If your trading requires a returns-based model, an Appendix, below, shows the practical calculations for converting a returns-based regression into concrete position sizes.

Other applications require volatility-matched legs. Typically, those cases involve negatively correlated legs. We match the volatility of the legs so that the net volatility of the spread is kept to a minimum. In those cases, we would calculate the hedge ratio based on asset volatility.

Another choice of metric is sensitivity, meaning partial derivatives or the "Greeks" (delta, gamma, etc.). If we are trading a market-neutral Treasury spread, we might require equal DV01s for each leg. The DV01 is the sensitivity

of the bond price to a change in yield. If we are trading stocks against options, we might require that the option's dollar delta equal the stock's dollar value. The delta is the sensitivity of the option's price to a change in stock price.

Examples

These examples illustrate several kinds of hedge ratios.

Crack spread The spread is determined simply by the technology of petroleum distillation: three barrels of crude oil produces (approximately) two barrels of gasoline and one barrel of heavy oil (e.g., heating oil). So the ratio is 3:2:1 when trading energy futures for crude oil, gasoline, and heating oil. This hedge ratio is:

- Fixed at 3:2:1

Treasury curve steepeners and flatteners - This is a DV01-based spread: the DV01 measures the sensitivity of each leg to yield change. The DV01 of both legs must be equal, giving a fixed hedge ratio of 1:1. The DV01 changes as the prices changes, however, so this hedge ratio is:

- Fixed at 1:1
- A spot value (because the DV01 calculation requires market prices but no historical data)
- Based on a sensitivity (DV01)

Pairs trading Here, I mean trading pairs of stocks. This usually requires the hedge ratio calculation described in the body of this note. So the hedge ratio is:

- Variable
- Based on historical data
- Calculated from prices

Calendar spread Here, I mean a calendar spread for futures or options: buy one expiration, sell another. For a simple calendar spread, the ratio is usually fixed a 1:1, so we buy one contract for every contract sold. But calendar spreads can be based on volatility, too. One might buy a certain number of options contracts and calculate their dollar-delta; then sell another expiration with the same dollar-delta. That sizing is based on deltas, not prices or returns. In that case, the hedge ratio is:

- Time-varying
- Based on spot value of the delta, not the historical delta
- Calculated from a sensitivity: delta

C Appendix: Sizing a Returns-Based Spread

When trading a price-based spread, the position size is easily determined by the coefficient of the regression equation:

$$P_{A,t} = \beta P_{B,t} + \epsilon_t$$

The regression suggests we trade β units of B for each unit of A .³

Sizing the position of a returns-based spread is more complicated. The hedge ratio is intrinsically time-varying. This Appendix presents the mathematics for calculating $N_{B,t}$, the number of units (contracts or shares) of B that we trade at time t for a returns-based spread.

By construction, the returns-based regression gives us this:

$$r_{A,t} = \beta r_{B,t} + \epsilon_t$$

where

- $r_{A,t}$ is the return on A at time t .
- $r_{B,t}$ is the return on B at time t .
- ϵ_t is the error term.

We want the dollar variation in both legs to be equal, which is expressed by this equation:

$$r_{A,t}V_{A,t} = r_{B,t}V_{B,t}$$

where

- $V_{A,t}$ is the dollar value of A at time t , and
- $V_{B,t}$ is the dollar value of B at time t .

But $V_{A,t} = N_{A,t}P_{A,t}$ where $N_{A,t}$ is the number of units of A at time t ; and $V_{B,t} = N_{B,t}P_{B,t}$ where $N_{B,t}$ is the number of units of B at time t ; giving:

$$r_{A,t}N_{A,t}P_{A,t} = r_{B,t}N_{B,t}P_{B,t}$$

We can replace $r_{A,t}$ based on the regression:

$$\beta r_{B,t}N_{A,t}P_{A,t} = r_{B,t}N_{B,t}P_{B,t}$$

We assume that $N_{A,t} = 1$, then solve for $N_{B,t}$:

$$N_{B,t} = \beta \frac{P_{A,t}}{P_{B,t}}$$

This, then, is the number of B units we trade for each unit of A . Notice that it changes when $P_{A,t}$ or $P_{B,t}$ change over time, which they certainly do.

³For this discussion, I am assuming the regression intercept is zero.