Simply Better Market Betas*

Ivo Welch

June 13, 2021

Abstract

This paper introduces a robust and easy-to-implement one-pass market-beta estimator. It only requires first winsorizing daily stock rates of return at –2 and +4 times the contemporaneous market rate of return. In predicting future market-betas, this “slope-winsorized” beta estimator predicts future betas better not only than OLS betas, Bloomberg betas (ubiquitous on financial websites), and Vasicek (1973) betas, but also published estimators that require intra-day data, super-computers, or financial statements. Moreover, using WLS to exponentially decay the weight of aged return observations (with a half-life of about four months) further improves the estimates.

Keywords: Market Beta, Factor Exposures, Robust Estimation

JEL Codes — C58 (Financial Econometrics), G11 (Portfolio Choice)

*Campbell Harvey kindly agreed to serve as the acting editor for this paper at the CFR due to a conflict of interest. (Any further questions about the handling process should be addressed to him). I think him and an anonymous referee for comments.

This paper introduces a new estimator for the unknown true market beta (β_i) of an individual stock i,

$$\beta_i = \frac{\text{cov}(\tilde{r}_i, \tilde{r}_m)}{\text{var}(\tilde{r}_m)} , \tag{1}$$

where $\tilde{r}$ indicates the (excess) rate of return and m indicates the market.

This new beta estimator is based on an easy-to-implement single-pass formula. It only requires winsorizing historical returns, as motivated by the linear relationship of the market model itself ($\tilde{r}_i = \alpha_i + \beta_i \cdot \tilde{r}_m + \tilde{\epsilon}$). On a daily basis, α_i should be so small that it can be ignored, leaving only $\beta_i \cdot \tilde{r}_m$ for guidance. When an observed stock rate of return $r_{i,t}$ is too far from the observed simultaneous market rate of return $r_{m,t}$ on the same day—specifically beyond what is implied by a beta of 1 ± 3 (i.e., $-2 \cdot r_{m,t}$ to $+4 \cdot r_{m,t}$, where t is a day index)—then this rate of return is winsorized. A return that is further from these bounds is more plausibly (but not certainly) an outlier.

For example, if the market rate of return on a particular day is +5%, the firm's rate of return on this day is winsorized at –10% and +20%. If the market rate of return is –5%, the firm's rate of return is winsorized at +10% and –20% instead. The standard OLS market-model regression using these winsorized rates of returns then gives the new estimator of beta. Because the winsorization threshold depends on the market's rate of return itself, I call this estimator a slope-winsorized beta estimator, bsw.

Despite its ease of implementation, in predicting future OLS market-betas, bsw outperforms all other prominent estimators that I am aware of. (The details will be explained in a moment.) These alternative beta estimators include not only the OLS estimator itself, but also its Bloomberg (formerly Merrill-Lynch) variant, as well as the Bayesian Vasicek (1973) estimator and the Dimson (1979) estimator. bsw also outperforms estimators that require intra-day data (Ait-Sahalia, Kalnina, and Xiu (2014)), supercomputer-intensive calculations (Martin and Simin (2003)), or financial statement information (Cosemans et al. (2016)). The superior performance appears among the largest 1,000 stocks, 2,000 stocks, 3,000 stocks, or all stocks. It appears in samples beginning in 1926 or 1973, or for that matter in almost every year in the sample. It appears when the predicted future OLS betas are themselves calculated with daily or monthly stock returns. It appears when the predicted future OLS betas are measured over one month, one year, five years, or ten years. Indeed, I am unaware of reasonable samples where it does not appear.

Furthermore, bsw also predicts not only the future OLS market-beta better than alternative estimators, but it also predicts future realized estimates of most alternative beta estimators better than their historical estimates themselves.

Furthermore, an even better and still one-pass version uses WLS instead of OLS to decay the influence of stock returns (in the market model) with age. This *beta slope-winsorized aged* (bswa) uses a decay parameter of 2/252 per trading day, which suggests a half-life of relevance of about 90 trading days (four months).

Both the winsorization and age decay parameters (±3.0 on the winsorization [around the mean beta of 1.0] and 90 trading days on the age decay) are low-dimensional, static, and robust. The parameters are not altered per stock, per time unit, per stock-time unit, or per other firm attributes. Instead, they are static and fixed for the *entire* CRSP sample. Winsorization levels from 1.0 ± 1.5 to 1.0 ± 5.0 (instead of 1.0 ± 3.0), with half-lives from 75 to 120 trading days (instead of 90 trading days) show similar performance.

My paper now proceeds as follows. Section I explains why the (future) realized OLS market beta is so useful to an investor, and when and why it matters that beta estimates can be horizon-dependent. Section II describes the estimators. It also shows that bsw performs well in simulations that match CRSP moments. Finally, it describes the age-decayed version, bswa. Section III describes the CRSP data used and shows some summary statistics. Section IV discusses the empirical test design. Section V investigates the empirical performance of the beta estimators. And Section VI concludes, followed by a 10-line R program that implements both bsw and bswa.

I Ex-Post OLS Market Betas as Estimation Targets

When an investor buys a security i and shorts w_i times the market, she earns $\tilde{R}_P \equiv \tilde{R}_i - w_i \cdot \tilde{R}_m$. The portfolio return volatility is

$$\mathrm{Var}(\tilde{R}_P) = \mathrm{Var}(\tilde{R}_i - w_i \cdot \tilde{R}_m) = \mathrm{Var}(\tilde{R}_i) + w_i^2 \cdot \mathrm{Var}(\tilde{R}_M) - 2 \cdot w_i \cdot \mathrm{Cov}(\tilde{R}_i, \tilde{R}_m) \ . \tag{2}$$

Dividing by $\mathrm{Var}(\tilde{R}_M)$ yields

$$\mathrm{Var}(\tilde{R}_P)/\mathrm{Var}(\tilde{R}_m) = \mathrm{Var}(\tilde{R}_i)/\mathrm{Var}(\tilde{R}_M) + w_i^2 - 2 \cdot w_i \cdot \beta_i \ . \tag{3}$$

Note that the beta in the market-model regression appears naturally, while the alpha is irrelevant to the optimal hedge. Because $\mathrm{Var}(\tilde{R}_m)$ is just a constant, the lowest volatility hedge is

$$\min_{w_i} \mathrm{Var}(\tilde{R}_P) \qquad \Longrightarrow \qquad w_i^* = \beta_i \ . \tag{4}$$

The β_i from the OLS formula is the investment weight in the market that minimizes the variance of the overall portfolio in its same realized sample of rates of return.

Furthermore, the CAPM is a model that relates this ability to hedge to equilibrium expected rate of return. This relationship requires many more assumptions and does not have strong empirical support. However, the hedging aspect does not require the CAPM. Hedging is also useful to all investors who want to minimize their tracking error to the market or lower their risk using market index instruments.

The relationships above hold not only for the *true unobservable* beta β_i (which minimizes the *true unobservable* volatility), but also for the *ex-post realized* beta b_i (which minimizes the *ex-post realized* volatility). An investor interested in the lowest risk portfolio (combining an investment in a firm with one in the stock market) would rather minimize the realized *ex-post* sample volatility than the expected *true* portfolio volatility, in the same sense that a lottery player would rather know the realized than the expected return on different number combinations. (However, this is of course not possible.) Equation (3) is also horizon independent, in that over any horizon, the OLS beta minimizes the portfolio variance in the sample of returns from which it has been computed.

Yet, an OLS beta estimate obtained from one set of (prior) returns does not necessarily minimize the variance in a set of different (later) returns. This is of course the application that analysts require. They can only use historical rates of return to estimate beta but are interested in the best hedge for future rates of return (i.e., the future beta).

This application is also why non-OLS beta estimators (on historical returns) can be superior in forecasting the subsequent OLS beta. A direct analog are asset-pricing models, in which analysts do not use historical average rates of return to predict future rates of return. (They would not expect a stock like Tesla with 500% return in 2020 to have a 500% return in 2021.) Instead, they predict better if they use a different model on historical returns—such as a factor model or a model based on analysts' prevailing predictions. The same is the case for estimate betas. Analysts would not necessarily want to use the historical OLS beta to estimate the future OLS beta.

If the underlying true return process (beta) were constant, the longest ex-ante return series should be used to produce a (then horizon-independent) estimate for the ex-post beta. However, this is not the case in reality. The underlying true betas are themselves time-varying—and specifically mean-reverting. Thus, unlike equation (3), the association between ex-ante forecasting beta estimates and the ex-post realization beta target is not horizon independent. The estimation interval matters and the best estimate of the instant beta is not the same as the best estimate of a longer-term beta. For this reason, Levi and Welch (2017) suggest more aggressive shrinkage when estimating the latter.

My paper primarily investigates the performance of various estimators in predicting the *ex-post* OLS beta, holding (most) estimation parameters constant (e.g., using the same sets of prior and subsequent returns to compare). Although the horizon could have mattered in when what beta estimator performs best, the empirical analysis shows that this is not the case in CRSP data for prediction horizons from one month to ten years. (I did not explore longer horizons.) The slope-winsorized beta estimators always performs better than their peers.

My base specification predicts a 252-day OLS market-model regression on daily returns on each stock-month end.[1] My paper (and unreported robustness checks) have

[1] A better estimation window than 12 months would be 16–20 months. See also Foster and Nelson (1996) and Ghysels and Jacquier (2007). However, the 12-month window has a more natural relation to the calendar and moreover performs nearly as well.

also considered many other variants. The results never changed qualitatively. That is, I could not identify subsets or methods in which the bsw and bswa estimators did not outperform their peers in estimating future OLS betas. bsw and bswa also predicted the future estimates from most other beta estimators better than these could predict themselves.[2]

II The Beta Estimators

An investor can only use *ex-ante* information to estimate the future OLS beta (the minimum-variance forward-looking hedge ratio). The novel slope-winsorizing estimator bsw introduced here is one formula among many.

A The Slope-Winsorized Beta Estimators

The slope-winsorized beta estimator belongs to the class of robust winsorizing methods. Such estimators require an “aggressiveness” parameter, which sets the trade-off between Type-I and Type-II errors: correctly winsorizing unpredictive outliers versus incorrectly winsorizing predictive outliers. Good choices trade off being too lax (thereby not having any effect) versus being too strict (thereby pushing all beta estimates too close to the same value of 1.0).

[Insert Figure 1 here: **Winsorization Techniques**]

The most common robust estimator would winsorize extreme (dependent) stock returns. The top plot in Figure 1 illustrates such a “level winsorization.” By compressing the range of the dependent variable, the beta estimates in the market-model regression become biased towards zero. This bias is especially undesirable in our case, where the beta prior centers not on 0.0 but on 1.0.[3] Moreover, the level parameter requires an

[2]My paper does not entertain a approach that assumes an equilibrium model in which stocks with higher betas have higher expected rates of return, and then tests what betas best predict the future average rates of return. Such tests would have to lean hard on a good equilibrium model of returns that the profession just does not yet have.

[3]Empirically, level winsorization performed poorly, because there were many days on which the overall stock market itself had exceptionally positive or exceptionally negative rates of return. On these days, level winsorization incorrectly cut off too many informative large positive or negative individual rates

analysis of what reasonable daily rates of return should be for different stocks—and they would likely be different not only across different types of stock but also when the stock market experiences a crash versus on a day on which not much is happening. A good winsorization level would not be easy to judge.

Level winsorization is so common that analysts often apply it to data before the models are even considered (and commonly fail to report it). However, model-specific winsorization schemes can often do better. Consider an example in which stocks follow a fat-tailed return distributions with occasional large outliers. An investor should not want to reduce such outliers when estimating volatility. However, when estimating betas, she should recognize that such outliers may randomly and unrelatedly occur on days when the rest of the stock market happens to have gone up or down. Ergo, this investor may want to reduce such outliers less aggressively when estimating volatility than when estimating beta.

"Slope" winsorization limits firms' rates of return based on minimal and maximal coefficient slopes on the market rate of return:

$$\mathrm{rsw}_{i,d} \in \Big(1.0 + \big[\,-\Delta, +\Delta\,\big]\Big) \cdot \mathrm{r}_{m,d}\ , \tag{5}$$

where $\in$ denotes the winsorization and Δ is the winsorization parameter. In this paper, delta is set to 3.0, thus leaving a return range limited to $(1.0 \pm 3.0) \cdot \mathrm{r}_{m,d}$.[4,5] The final slope-winsorized beta is then

$$\mathrm{bsw}_i \equiv \frac{\mathrm{cov}\big[\ \mathrm{rsw}_{i,d}(\Delta)\ ,\ \mathrm{r}_{m,d}\ \big]}{\mathrm{var}(\ \mathrm{r}_{m,d}\ )}\ . \tag{6}$$

of return. Band winsorization is an alternative to slope winsorization, but it has not been used in the literature and it does not seem to outperform slope winsorization. It is also more natural to specify reasonable (beta) slopes than reasonable (return) residuals. The latter may very well differ more across different types of stocks.

[4] Delta was chosen roughly where the monotonic relation between historical and future OLS betas breaks down. Betas more extreme are rare. In the CRSP sample, only 0.17% of all bols estimates exceed slopes of –1 and +4. (In fact, negative betas rarely occur.) The delta choice is extremely robust. Even *ex-post* optimal parameters within many subsamples improve the predictive performance only modestly.

[5] Note that if the market rate of return is 0%, the firm's rate of return is itself set to 0%. This turns out to be irrelevant, because observations on days on which the market rate of return is (near) its mean (of zero) are uninformative in the market-model regression anyway.

The bottom plot in Figure 1 illustrates slope winsorization. It is custom-designed for our application: It relies on the knowledge that the mean market beta is 1, that the daily intercept (alpha) is near 0, and that, from an ex-ante perspective, few stocks are likely to have extreme rates of return, not justified by overall market returns appropriately modulated by reasonable market betas.

Note that unlike a Bayesian prior, the winsorization changes all beta estimates. It can do so even for stocks with OLS betas of 1.0. The presumption is not that betas shouldn't be too large, but that stock returns shouldn't be too large.

B Benchmark Alternatives

The literature already offers other estimators against which the performance of the slope-winsorized beta can be judged:

1. **bols:** The *ex-ante* OLS beta is obtained from a regression using *ex-ante* daily returns.

2. **bvck:** Like bols, the Bayesian shrinkage estimator in Vasicek (1973)—which is the same as a frequentist random-effects panel estimator—can be calculated with daily-frequency stock returns.[6] For each stock, it requires first computing bols_i and the coefficient's standard error (σ_i^2). Then it requires calculating cross-sectional statistics over all stocks' betas to obtain a cross-sectional mean ($\overline{\text{bols}_t}$) and standard deviation ($\overline{\sigma}_t^2$). For each stock i at time t, the Vasicek estimate is

$$\text{bvck}_i \equiv w_i \cdot \text{bols}_i + (1 - w_i) \cdot \overline{\text{bols}}_t \ , \quad \text{where } w_i \equiv \frac{\overline{\sigma}_t^2}{\sigma_i^2 + \overline{\sigma}_t^2} \ . \tag{7}$$

   The Vasicek estimator was derived under the assumption of betas that are not time-varying and that market-model residuals are normally distributed. It performs well with time-varying underlying betas and outliers, although this was not how its use was originally justified.

3. **bfp:** Frazzini and Pedersen (2014) suggested a hybrid estimator, which uses both daily and monthly frequency stock returns. The estimation details are explained

[6] Vasicek/random-effects estimators have shown superior performance also in predicting future alphas (Harvey and Liu (2018)).

in their paper. Frazzini and Pedersen did not validate the performance of their estimator. See also Novy-Marx and Velikov (2018) and Han (2022) for their critiques that bfp is not really an estimator of the true beta.

4. **bmols:** An equivalent OLS beta can be calculated from *monthly* stock returns. Instead of using about one year's worth of data, monthly estimators use 60 months of historical return data. (The one-year monthly-return frequency beta estimates perform far worse.) The abbreviations contain the letter *m* to distinguish them from the daily estimates.

5. **bmvck:** An equivalent Vasicek estimator calculated from monthly stock returns.

6. **bmblm:** Blume (1971) debiased the OLS estimator using an empirically estimated linear correction. His Table 4 shows a mean adjustment of 0.64. The Bloomberg (formerly Merrill-Lynch) market beta simplifies the shrinkage to 2/3. This estimator is covered on the CFA exam and distributed widely by Capital IQ to financial websites, such as *Yahoo!finance* and *Google Finance*.

Section V.E briefly discusses some other beta estimators.

C Simulated (Theoretically Expected) Performance of Estimators

An OLS estimator is known to be the best unbiased linear estimator ("BLUE") under standard OLS assumptions. Why is it then possible that the bsw estimator can perform better in predicting the future bols realization than the bols estimator?

The primary differences from the standard OLS assumptions are that (1) there are stock return outliers in the data, and (2) the bsw estimators can use information that the true mean beta is about 1.0, which is not used by bols, (3) the estimator does not require unbiasedness (as do other common shrinkage estimators), and (4) the underlying beta is mean-reverting. The extra information in (2) is used not only by the slope-winsorized estimator, but also by the bmblm and bvck estimators.

The first question of interest now is how well one can expect these estimators to perform in theory. Unfortunately, there are no closed-form expressions for the theoretical properties of *any* robust slope estimators even under the normal distribution. Moreover, stock-return outliers do not follow any known distribution. Fortunately, simulations in large samples can yield nearly exact results. Thus, I now examine the performance of bsw and other beta estimators in simulations designed to match some empirical moments:

1. Each simulation begins with a hypothetical set of five years of daily market rates of return, which are randomly resampled realizations from the empirical realizations of value-weighted market returns after 1974.

2. I now draw pairs of market-model beta and sigmas (the standard deviation of market-model residuals). The intent here is to retain the association between the two—in the CRSP data, firm-years with higher betas also have higher sigmas. I entertain two variants:

   (a) In a parameterized version, a sigma is first drawn from a lognormal distribution, and beta is assumed to be linearly related to the log-sigma plus an error. The moments (and relation) are matched to their distributions in the empirical data. The linearity is not intended to be realistic, but just a sketch.

   (b) In a sampled version, sigma and beta are randomly drawn pairs from the empirical firm-year distribution of estimated market-models.

   (Note that the simulations remain unrealistic in one respect—the true beta is held constant.)

3. I now sample five years of daily market-model residuals from the empirical distribution. This is in order to retain the skewness and kurtosis of market-model residuals in the data.

4. Daily returns are the sum of the beta-scaled market rates of return plus the sigma-scaled market-model residuals.

All generating information, except stock and market returns, is hidden from the estimators. Monthly estimators (like bmblm) receive access to five years of data, albeit compounded into 60 monthly stock returns.[7] The daily estimator receives access only to one year, 252 daily stock returns.

[7] This is generous to the bmblm estimator, in that it assumes that the underlying betas do not change. Thus, stock returns from five years ago remain as relevant as more recent stock returns (as used in bols, bvck, and bsw).

Unlike in the empirical analysis in later sections, the experimenter knows the true (fixed) OLS betas in each simulation. Thus, there is no need for a sixth year of out-of-sample market betas to be predicted. The beta estimates can be directly compared to the true betas.

[Insert Table 1 here: **Idealized Properties of Market-Beta Estimators' Errors**]

Table 1 tabulates the results. For each draw, we calculate betas using various estimators, and compare them to the (hidden but known) true betas that underlay the random draws. The table shows that monthly bmblm estimates are nearly unbiased, but they have high RMSEs of 0.32 to 0.33 in both scenarios. Both the plain OLS and Vasicek estimators can provide better beta estimates using a shorter window with higher frequency data. Their RMSEs are 0.188 and 0.184 for bols, and 0.200 and 0.184 for bvck. The slope-winsorized beta estimator performs best, with RMSEs of 0.170 and 0.177. This suggests a relative improvement by the bsw estimator of about 40–45% over bmblm, and about 5–10% over the bols and bvck estimators.

D Age Decay Versus Block Sampling

It has long been known that the underlying beta drifts. The common way to handle this drift is to use a *block-sampler* (Ghysels and Jacquier (2007)), i.e., a moving estimation window. All daily estimators described so far are 252-day block samplers computed at the end of each month; all monthly estimators are 60-month block samplers.

An alternative is to progressively disregard older observations. A weighted-least-squares regression (WLS) can offer smooth in-time age decay. This is the age-decayed bsw estimator, mentioned in the introduction, named bswa.

Block samplers require a window length as a parameter, while age decay requires a speed of decay as a parameter. To keep its implementation simple, bswa uses only one fixed decay parameter, regardless of firm and time. Thus, like bsw, analysts can calculate bswa with no need for a first-stage regression. My age-decayed, slope-winsorized beta estimator has a constant decay of $2/256 \approx 0.78\%$ per day and is named bswa. Good half-lives range from about 75 to 120 days, with 90 days being a good middle, $(1/(1+2/256)^{90} \approx 0.5)$. Intuitively, this parameter implies that yesterday's stock returns should have about twice the weight of stock returns from four months ago, eight times the weight of those from one year ago, and sixty times the weight of those from two years ago. Three-year-old stock returns are effectively irrelevant.

The decay parameter itself was originally chosen based on an analysis of data prior to 1973. As just noted, the empirical performance of bswa remains very similar when each day's presumed information content decays not by 0.78% per day, but by 0.7% to 0.9% per day. That is, the precise choice is not sensitive. When optimized, the best decay also remains stably within this interval regardless of sample or years analyzed. In sum, the decay parameter could be chosen based on predictive performance in preceding years (or based on different stocks altogether). The choice would remain similar.

An advantage of age-decayed betas is that they can be easily updated in the same regression in time—unlike a block sampler which needs a new regression whenever more data becomes available and is to be used. A disadvantage of age-decayed betas is that they should include all data since the first appearance of a stock (though old-enough data barely matters)—unlike a block sampler which has a limited data requirement.

III Data

The data set for the analysis is based on the early-2020 CRSP database. It contains 75,230,516 daily stock returns from ordinary stocks (with CRSP share codes 10 and 11) from 1926 to 2019. The main data analysis itself begins in 1973 (just after NASDAQ came online).

In addition to the stock return data, end-of-month marketcap ranks from CRSP (contemporaneous with the *ex-ante* market-beta independent variable estimates) are sometimes used to limit the main analysis to the then biggest 1,000 or 3,000 stocks.

In the implementations and tests below, the dependent variable in the market model is each stock's own rate of return (from CRSP) net of the risk-free rate of return (from Ken French's website). The independent variable is the value-weighted market rate of return, again net of the same risk-free rate of return. The beta is not calculated when a stock does not have at least 20 trading-days of stock returns, or when a return is not present on the last trading day. In some later tables, other data requirements (such as 100 trading days within the year) are imposed.

[Insert Table 2 here: **Descriptive Statistics of *Ex-Ante* Beta Estimates, 1973–2019**]

Table 2 shows the data ranges, means, and standard deviations of the independent variables (predicting betas) used in the first set of tables. Each daily beta is based on 252 days of daily stock returns. Each monthly beta is based on 60 months of return.

Because the stock market index used in the market model is value weighted, the equal-weighted average beta across stocks in the sample is not 1.0. However, for the 1,000 biggest stocks, the distance from 1.0 is less than 0.03 for the first four estimators (bols, bvck, bsw, and bswa) and less than 0.07 for the others. For the 3,000 biggest stocks, the distance from 1.0 is modestly more pronounced, reaching 0.13 for bmols. Beta estimates based on different methods also differ in their standard deviations. The bfp estimates are least heterogeneous. The OLS estimates (both daily and monthly) are most heterogeneous.

[Insert Table 3 here: **Distinctness of *Ex-Ante* Beta Estimates, 1973–2019**]

Because all estimators attempt to isolate the same underlying true beta from the same returns, it is not surprising that they are somewhat similar. However, this raises the question of whether the choice of estimator matters. Table 3 shows the root-mean-squared differences across estimators in pairwise comparisons.

The largest estimated differences are observed across estimates using different return frequencies. Among daily estimates, the differences are about 0.1 for the 1,000 biggest firms and 0.15 for the 3,000 biggest firms. Among monthly estimates, the numbers are just a little higher. Across monthly or daily estimates, the distances are between 0.4 and 0.6. This holds true even for estimators using the same methods (such as bols and bmols).

The two slope-winsorized estimators (bsw and bswa) differ by 0.08 from each other for the biggest 1,000 firms and by 0.10 for the biggest 3,000 firms. The block-sampled 1-year bsw is also similar to the block-sampled 1-year bvck with a distance of 0.06. (The good performance of bvck was due to its [largely] coincidental handling of outliers.)

These differences among estimators seem sensible. They are small but not trivial. For a CAPM use with a risk premium of 6%, the typical inferred expected return differences due to beta estimators with RMSE differences between 0.1 and 0.4 would be about 60 to 200 basis points per annum. This is economically meaningful.

IV Test Design

A Forecasting Regressions and R^2 and RMSE as Performance Metrics

The primary criterion to evaluate the performance of slope-winsorized betas and other betas is how well they forecast *realized future OLS betas* (Rosenberg and Guy (1976), Harrington (1983)). The secondary criterion (to be explored in Table 7) is whether they can forecast themselves. If the slope-winsorized betas can also predict other beta estimators better than these beta estimators can predict themselves, it suggests that investors should still use the *ex-ante* slope-winsorized betas even if they were interested in these other estimators' *ex-post* realizations.

Like earlier papers, I consider two measures of forecasting success. Both are based on the ability of estimated historical betas (named $\hat{b}_{i,y}$) to predict the future realized OLS beta, $bols_{i,y+1}$, where y is a year index. (Each beta is estimated from daily returns within the year y, although Table 5 extends the forecast to multiple years.)

The first success metric is the R^2 from predictive "gamma" regressions, $bols_{i,y+1} = \gamma_0 + \gamma_1 \cdot \hat{b}_{i,y} + \epsilon_{i,y+1}$:

$$R^2 \equiv \frac{\mathrm{Cov}(\hat{b}_{i,y}, bols_{i,y+1})^2}{\mathrm{Var}(\hat{b}_{i,y}) \cdot \mathrm{Var}(bols_{i,y+1})} \;, \tag{8}$$

averaged over all observations (cross-section and time-series).

The second metric is for the use of a historical beta as a direct proxy,

$$\mathrm{RMSE} \equiv \sqrt{\sum_i \sum_y (bols_{i,y+1} - \hat{b}_{i,y}) \,/\, (N \cdot Y)} \;. \tag{9}$$

The R^2 metric is not affected by *bias* in the beta estimators, but the RMSE metric is. When beta is merely a control variable in a regression, such bias is often harmless and methods with higher R^2 are better. When beta is used directly to inform the hedge ratio, methods with lower RMSE are better.

Most of the remaining tables below show the results of pooled forecasting panel "gamma" regressions. The nature of the data means that these regressions are mostly cross-sectional. Each year has thousands of observations in the cross-section on fewer than 50 years. Not shown here, the results are also the same with Fama–MacBeth-like test specifications, when overlapping months rather than calendar-year forecasts are used (or vice-versa), and when betas compete in multivariate (rather than univariate) gamma regressions.

B Errors In Beta Variables

Realized OLS betas are a measure of realized diversification benefits for an investor holding the market portfolio. They are themselves of interest.

However, they are not the true underlying betas. To the extent that realized betas differ from true betas, the prediction tests suffer from noise (e.g., lower R-squareds in the "gamma" regressions, which predict future realized betas with current beta estimates).

For the dependent variable, the future bols, measurement noise is a minor concern. One expects ex-ante estimators that perform better in predicting the future realized bols also to be able to predict better the unknown true (expected) betas. Put differently, even if the interest is the ability to predict the true future market-beta, errors in the dependent variable are a benign complication from an econometric perspective. It is what OLS was designed for.

For the independent variable, the noise is more of a concern if the gamma coefficient is to be interpreted as a measure of how the true beta would predict the future beta. If the underlying model is stable, then the asymptotic bias in a slope coefficient of past on future values is $1/(1+\sigma_e^2/\sigma_b^2)$, where σ_e^2 is the squared standard error of the noise and σ_b^2 is the cross-sectional dispersion in the (beta) predictor. The average estimated beta standard error in the market-model regression is a rough estimate for σ_e^2. It is about 0.05 (per day). The average dispersion of estimated betas in the cross-section is a rough estimate for σ_e^2. It is about 0.40 (per firm). Thus, an over-the-envelope estimate for the γ_1 bias is about $1-1/(1+0.05/0.4^2) \approx 2\%$.[8] The actual empirical bias is larger because the underlying beta is also *not* stable. In addition to time-varying betas, our sample also has time-varying heterogeneity in firm-size, and with it time-variation in cross-sectional beta means and standard deviations.

[8] These approximations have ignored the panel nature of the data. However, unreported simulations suggest that they are reasonably applicable in our panel sample, too.

However, the purpose of my forecasting regression is not to test whether the gamma estimates are biased relative to the case in which the independent variables had been the true betas. They surely are[9] if one wanted to use the gammas to assess how the *true unknown beta* would predict the future (OLS) beta. Instead, the purpose is to determine which of the *beta estimates* predicts the future OLS beta better. It is the estimated betas that are the object of interest, not the true betas. In this sense, there is no errors-in-variables concern.

However, we can assess how the beta estimates relate to the true unknowable betas. If both the dependent and the independent variables are proxies drawn with error from an unknown true but stable normal variable, the R^2 of a cross-sectional regression of one proxy on the other yields an R^2 that is the square of the R^2 in an (infeasible) regression of the true (unknown) beta parameter on the realized OLS beta (see also Jegadeesh et al. (2019)). For example, if the *ex-ante* OLS beta, bols, can explain 50% of its *ex-post* self, it would suggest that bols could explain about $\sqrt{.50} \approx 71\%$ of the true unknown beta.[10]

In sum, potential improvement in predictions for *any* beta estimator above and beyond that provided by the OLS beta itself are effectively limited by two aspects. First, the predicted beta is not the true beta, but a noisy measure of the true beta. Second, if two betas are different by 0.2, the maximum improvement that one estimator can provide over the other is 0.2.

V Empirical Evidence

A Predicting One-Year Daily OLS Betas

Predicting out-of-sample betas is the problem investors face when they want to determine their minimum-variance hedge ratios. Table 4 shows the key result of the paper: the performance of estimators in predicting the future OLS market beta. Here, betas are calculated from 252 daily stock returns.

[9] The above calculation suggest that the biased would be modest, however.

[10] If the underlying betas are changing, the estimated square root of the R^2 is a lower bound. The association of an instant proxy with the true instant beta would be higher.

[Insert Table 4 here: **Predicting Future *Ex-Post* (One-Year-Ahead) OLS Market Beta, 1973–2019**]

The first four rows focus on estimators based on daily stock returns. The slope-winsorized betas outperform both the OLS and Vasicek betas. The performance improvement of bsw over bvck is about the same as the performance improvement of bvck over bols. The age-decayed bswa performs better than the block-sampled bsw. The standard errors show that, with their 351,155 (overlapping) and 68,267 (nonoverlapping) firm-months observations, respectively, the forecast improvements are always statistically significant.

The next rows show that the other four beta estimators perform much worse. The bfp estimator, a hybrid (still based partly on daily stock returns), performs worse than any of the pure daily-return-based estimators. The OLS and Vasicek estimators from monthly stock returns perform even worse. And the nearly ubiquitous Bloomberg estimator pales in comparison to *any* of the betas obtained from estimators based on daily stock returns. Beta estimators based on monthly stock returns should not be used for forecasting the daily-frequency OLS beta.

This forecasting improvements also align roughly with those suggested in Section II.C. The improvement of bsw over the Bloomberg and other monthly betas was about 30%, 8–10% over the OLS beta, and 4–5% over the Vasicek beta. The superior performance of bsw is remarkably consistent and all the more impressive, because the ex-post betas used in the performance tests are not the true betas but are themselves measured with error.

[Insert Figure 2 here: **Year-by-Year Relative Predictive Performance of Estimators, Big 3,000 Stocks, Dec 1973–2019**]

Figure 2 plots the relative performance of the estimators by year. Each point is the difference between the forecasting RMSE obtained by bols minus the forecasting RMSE of the named estimator. A positive number means that the estimator outperforms the bols estimator. The plot shows that the RMSE performance of the three estimators is better in the second half than in the first half. The slope-winsorized estimators, both plain and age-decayed, visibly outperform both the bols and bvck estimators—and not just in a some year clusters. The superior performance of the two slope-winsorized estimators is not a fluke. It occurs in (completely separate) samples over many years.

The performance has also not diminished over time. Not shown, other subsamples (e.g., based on marketcap, assets, stock volatility, *ex-ante* beta, etc.) also suggest that slope-winsorized estimators are better than their peers.

Not shown, when the static delta and rho parameters are replaced by the (unobtainable) *ex-post* year-specific delta and rho parameters that would optimize performance, the performance of the slope-winsorized betas of course improves—but not greatly so. This suggests that any potential gains from further fine-tuning these two parameters are likely to be small.

B Predicting Long-Term OLS Betas

[Insert Table 5 here: **Multiyear Performance, 1973–2019, 2000 Stocks, Annual**]

It is a good question whether a better short-term beta hedge is also the better long-term beta hedge. Thus, Table 5 predicts longer-horizon market-betas. For variety, this table now uses the largest 2,000 stocks and non-overlapping annual calendar year regressions. Both panels show that the bsw and bswa estimators always perform better than the other estimators. The rightmost column further shows that the daily OLS beta estimators also predict the future monthly OLS beta estimates better than the monthly OLS beta estimates predict themselves. Even a (CAPM) user should use daily-return frequency betas, not monthly-return frequency betas. Furthermore, monthly betas are especially problematic if they are used as direct proxies and not first debiased.[11]

[11] The monthly betas perform "only" badly, but not "as" badly, in predicting future monthly betas. This may be the case because the same estimation bias/noise appears in both ex-ante and ex-post monthly betas and differently across different kinds of stocks. Note however that the samples in the 5y daily and 5y-MO monthly regressions are different, too, and thus the R^2 are not strictly comparable. In any case, in the same set of observations, monthly betas remain dominated by daily betas.

C Predicting Hedged Portfolio Returns

[Insert Table 6 here: **Hedged Return Performance, 1928–2019, 3,000 Stocks**]

Table 6 uses a different sample (starting in 1928), annual non-overlapping gamma regressions, and 3,000 stocks. Its purposes is to examine a different success metric—the variance of the market-hedged stock portfolio over the following year. In effect, in this objective function, the portfolio cannot be rebalanced (rehedged) during the calendar year, because the beta-hedge is applied to year-compounded firm and market rates of return. The table shows that the portfolio is best hedged with the ex-ante bsw or bswa rather than with the ex-ante bvck or bols estimators. However, the bvck estimator shows good performance, too.

D Predicting Self Rather than the OLS Beta

Investors who want to establish a minimum-variance hedge presumably care only about the *future* beta. Historical data is useful only insofar as it informs the estimate. To the extent that all beta estimators attempt to uncover the same true beta (β) signal in noisy stock return data, the slope-winsorized estimators could predict better the *ex-post* realizations of not only bols estimates but also other (noisy) estimators.

To elaborate, consider a thought experiment. Assume that an investor is interested in some other beta estimate for its own sake (e.g., to place a bet with a friend). For example, this investor may want to bet on what the future Bloomberg beta estimate will be. Would this investor prefer to use the historical Bloomberg beta as her best predictor or would she prefer to use bswa? Because bswa predicts the true beta better, it could also predict the future Bloomberg beta (which partly contains the true beta) better. And if this were so, the *prevailing* historical Bloomberg beta number would be of even lesser interest. Good self-prediction is a necessary but not a sufficient criterion for an estimator to seem useful.[12]

[12] *Self-prediction is not enough.* For example, a "beta estimator" claiming that beta is the firm's first CUSIP number could predict itself perfectly well, but it would not be considered a good estimator of beta. This qualification is also important, because it has sometimes been argued that industry, size, or other aggregated betas should be used in lieu of firm betas, because they are more stable. Although it is indeed true that they are more stable, even cursory examination reveals that industry betas are *very poor* predictors of any firm's own firm betas. They should *never* be used as proxies for individual stock betas. See also Levi and Welch (2017).

[Insert Table 7 here: **Predicting Future *Ex-Post* Betas Other Than the OLS Beta, 1973–2019**]

Table 7 predicts *ex-post* estimates that differ from those obtained from bols. The table shows that, based on the $\bar{R}^2$ metric, the *ex-ante* age-decayed bswa is the best predictor of the *ex-post* beta metric obtained from any other formula (estimator). The one-year block-sampled bsw is not as good but never far behind. For example, consider an investor who wants to estimate an ex-post 5-year beta based on monthly returns (bmols). On average, the investor will miss the mark by 0.557 when using the historical bmols' ex-ante equivalent. But, the investor's estimate will deviate by only 0.484, on average, if she uses the historical bswa ex-ante equivalent.

Two estimators have persistent stock-specific estimation biases. Thus, based on the RMSE metric, the bswa estimates cannot predict the Frazzini–Pedersen bfp or the Bloomberg bmblm estimates as well as these two estimators can explain themselves. In this sense, the Frazzini-Pedersen bfp and the Bloomberg bmblm estimators can predict their future selves better on the RMSE metric than the bswa estimator can explain them. In the case of the bmblm metric, the intuition is easy. The superior self-performance is caused by the obvious fixed shrinkage towards 1.0—a metric that would shrink 99% towards 1.0 would have an even more persistent bias. Absent their persistent biases (i.e., measuring performance on the R^2 metric), their self-prediction advantage disappears. For the bfp estimator, the reason is the volatility mix into the estimator (Novy-Marx and Velikov (2018), Han (2022)). bfp is not primarily an estimator of the OLS market-beta.

In sum, with their easier implementation, there is little reason not to use the better performing slope-winsorized estimators from daily stock returns in all cases.

E Other Estimators

E.1 The Dimson Non-Synchronicity Estimator

Dimson (1979) and Scholes and Williams (1977) betas serve a different purpose. They intend not to estimate the *ex-post* OLS beta, but rather a "synchronicity-adjusted" equivalent. Thus, the appropriate criterion to assess their empirical performance is how they predict themselves, not how they predict the future OLS beta, at least not in cases in which one would suspect non-synchronicity. The Dimson beta is more common, perhaps because it is also very easy to compute. It is the sum of three regression coefficients, with one lead and one lag on the market rate of return included (in addition to the contemporaneous market rate of return).

[Insert Table 8 here: **Summary Statistics by Dollar Trading Volume Deciles, 1974–2019**]

My first step is to identify subsets of stocks in which there likely is a lot of non-synchronicity vs. subsets where there is not. Table 8 shows that there are four orders of difference in magnitude across trading volume deciles, when trading volume is defined as the average daily dollar trading volume over the last 22 trading days of the year.[13] The next row shows that the price on CRSP is a bid-ask quote rather than a trade (marked as a negative number) in one in three firm-years in the least-traded deciles. It could be argued that such a quote suffers from less non-synchronicity than a price from a trade that would have occurred earlier in the day. More importantly, it confirms that the lowest decile truly suffers from infrequent trades and thus more likely from non-synchronicity. In contrast, in decile 10, stocks almost always trade on their last day.

The table further lists the means and standard deviations of the bols, bswa, and bdim (Dimson) estimators by dollar trading volume. Firms that trade less have lower betas on all metrics. This could be due to inadequate non-synchronicity control or due to the fact that these firms are just different. bswa has the lowest heterogeneity, bdim has the highest.

[Insert Table 9 here: **Performance of Dimson Estimator, 1974–2019**]

Table 9 shows predictive-regression performance metrics analogous to those in earlier tables (e.g., Table 4).

In Panel A, the predicted variable is the ex-post bols (based on 1 year of daily stock returns, calendar years only). It is often overlooked that Dimson estimates do not provide superior performance for free. When there is no non-synchronicity concern, bdim should perform worse, because it fails to take advantage of the restriction that stock returns should be uncorrelated over time. Table 9 shows this effect. In the most-traded decile, where there is little non-synchronicity and thus the ex-post bols is a good target, bdim

[13]The sort is by trading volume within each calendar year, but the dollar volumes themselves are not inflation-adjusted. Thus, the magnitude of the reported numbers in Table 8 should be viewed only as meaningful in relative but not absolutely terms.

as a predictor is greatly inferior even to bols as a predictor. bdim also performs poorly predicting the OLS beta among the least-traded stocks, but this could be because the ex-post bols is meaningless for stocks with such low trading volume.

When non-synchronicity is a serious concern, it may make more sense to predict the ex-post Dimson estimate, presumably a better measure of the true beta. Panel B shows that bdim performs better in explaining its future self than bols in the least-traded decile, at least on the R^2 metric. The Dimson estimator does perform worse on the RMSE metric than the OLS estimator in predicting the future Dimson estimate. Even in the least-traded decile, the Dimson estimator should not be used as a direct proxy. Instead, it should be appropriately mean-adjusted first.

Moreover, in general, it is never advisable to use the Dimson estimator non-discriminately on all stocks. This is because bdim performs *much* worse even than the plain bols in the other nine deciles.

However, bswa ultimately obsoletes bdim on both the R^2 and the RMSE metrics and does so in all deciles. Ergo, even though non-synchronicity is concerning for stocks that almost never trade, the bswa estimator remains preferred to bdim. It may be possible to combine the outlier handling of bsw and the non-synchronicity aspect of bdim to offer an even better estimator. This is left to future research.

A final word of caution: Although the Dimson estimator cannot predict its own future realizations as well as other daily estimators can predict them, the evidence here pertains only to publicly-traded stocks. The poor performance of bdim is perhaps not surprising, because stocks that do not trade are still marked-to-market at the end of the day. The Dimson estimator may remain more useful for assets in which daily asset returns are not readily available and even monthly asset returns are of dubious quality.

E.2 More Beta Estimators

The bsw and bswa estimators also outperform other estimators:

Martin and Simin (2003) propose a more complex nonlinear two-equation robust estimator. Unfortunately, the computation of their betas requires access to a supercomputer and is thus unlikely to find widespread use. Fortunately, the authors generously shared their estimates. Their estimators perform about as well as the bvck and bsw estimates. They perform worse than the bswa estimates, both in predicting *ex-post* OLS betas and in predicting themselves.

Ait-Sahalia, Kalnina, and Xiu (2014) develop a nonparametric time-series regression estimator without the usual assumption of piecewise linearity, using *intra-day* data. Unfortunately, these betas are difficult to replicate, partly because the TAQ input data require extensive cleaning, partly because the data are voluminous, and partly because the econometrics is complex. Fortunately, the authors also generously shared their estimates. These estimates perform poorly in the task set to them here. They underperform even the historical OLS betas in predicting the future OLS betas.

Cosemans et al. (2016) suggest a number of different estimators, based on underlying stock fundamentals, business-cycle variables, and Bayesian techniques. The data frequencies range from semi-annual to intra-day data. These estimators are quite complex. Fortunately, the authors also generously shared their estimates for the S&P 100 stocks. The Cosemans et al. (2016) betas underperforms both bsw and bswa (and on both performance metrics) when predicting the future bols.[14]

Fundamentals-Smoothed Betas, i.e., estimates in which CRSP or Compustat data are used, yield an improvement if the inferior monthly-frequency, stock return-based betas are used. They yield no economically meaningful improvements for daily slope-winsorized betas.

In addition, I explored other estimators and tests. The results always showed superior slope-winsorized beta performance. For example, I also investigated whether selection or survival effects mattered. For example, I explored predicting future betas calculated over 1 month of daily stock returns instead of 1 year. For example, I experimented with removing instead of winsorizing observations, with firm-year specific winsorization (of all kinds) and/or with industry-related, peer-related, marketcap-related, volatility-related, and trading volume based winsorization. Some tuned versions of bsw and bswa indeed improved modestly on the bsw and bswa estimators reported here—not surprising given the extra free parameters. However, none improved the bsw and bswa in a meaningful way.

[14]As with bswa, self-prediction of the Cosemans et al. beta makes little sense, because the performance is affected by overlap due to infinite reach-back (use of history).

VI Conclusion

The most common Bloomberg (Merrill Lynch) beta estimator, which is prominently used on many financial websites, performs poorly in forecasting the future market-beta, as do other betas calculated from monthly stock returns. Betas computed from daily stock returns perform better. This is the case even if the object of interest are future betas calculated from monthly stock returns.

Among daily-frequency based beta estimators, the simple, one-pass, slope-winsorized bsw and bswa (respectively block-sampled and exponentially age-adjusted) seem to perform best. Most importantly, these estimators are also trivially easy to implement. The complete R code is 10 lines long and in the Appendix. There is no reason not to use them.

References

Ait-Sahalia, Yacine, Ilze Kalnina, and Dacheng Xiu (2014). "High-frequency factor models and regressions". In: *Journal of Econometrics* 216, pp. 86–105. DOI: 10.1016/j.jeconom.2020.01.007.

Blume, Marshall E. (1971). "On the assessment of risk". In: *The Journal of Finance* 26.3, pp. 1–10.

Cosemans, Mathijs, Rik Frehen, Peter C. Schotman, and Rob Bauer (2016). "Estimating security betas using prior information based on firm fundamentals". In: *The Review of Financial Studies* 29.4, pp. 1072–1112.

Dimson, Elroy (June 1979). "Risk measurement when shares are subject to infrequent trading". In: *Journal of Financial Economics* 7.2, pp. 197–226.

Foster, Dean P. and Dan B. Nelson (Jan. 1996). "Continuous Record Asymptotics for Rolling Sample Variance Estimators". In: *Econometrica* 64.1, pp. 139–174.

Frazzini, Andrea and Lasse Heje Pedersen (2014). "Betting against beta". In: *Journal of Financial Economics* 111.1, pp. 1–25.

Ghysels, Eric and Eric Jacquier (2007). *Market Beta Dynamics and Portfolio Efficiency*. Tech. rep. University of North Carolina and HEC Montreal.

Han, Xing (2022). "Understanding the Performance of Components in Betting Against Beta". In: *Critical Finance Review*.

Harrington, Diana R (1983). "Whose beta is best?" In: *Financial Analysts Journal* 39.4, pp. 67–73.

Harvey, Campbell R and Yan Liu (2018). "Detecting repeatable performance". In: *The Review of Financial Studies* 31.7, pp. 2499–2552.

Jegadeesh, Narasimhan, Joonki Noh, Kuntara Pukthuanthong, Richard Roll, and Junbo Wange (2019). "Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation". In: *Journal of Financial Economics* 113, pp. 273–298.

Levi, Yaron and Ivo Welch (Apr. 2017). "Best Practice for Cost-of-Capital Estimates". In: *Journal of Financial and Quantitative Analysis* 52 (2), pp. 427–463.

Martin, R. Douglas and Timothy T. Simin (Sept. 2003). "Outlier-Resistant Estimates of Beta". In: *Financial Analysts Journal* 59.5, pp. 56–69.

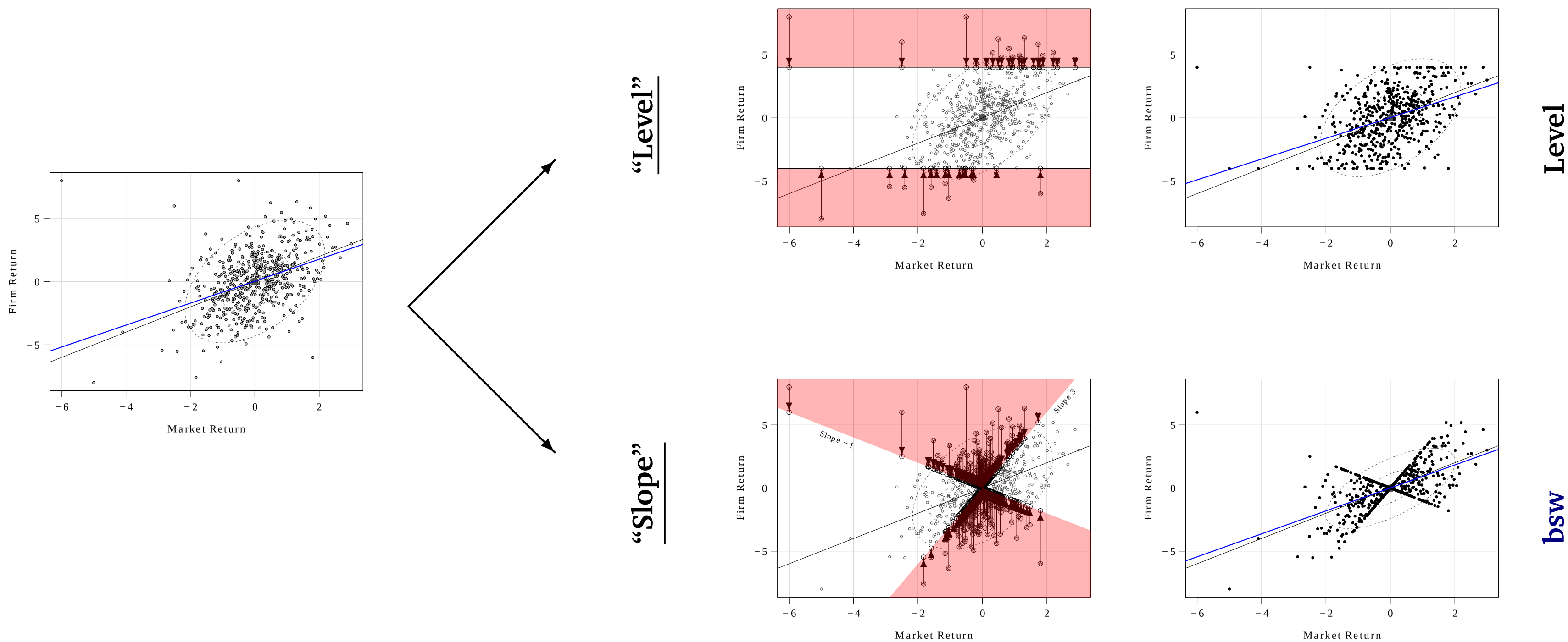
Novy-Marx, Robert and Mikhail Velikov (Nov. 2018). "Betting Against Betting Against Beta". In.

Rosenberg, Barr and James Guy (July 1976). "Prediction of Beta from Investment Fundamentals (Part Two)". In: *Financial Analysts Journal*, pp. 62–70.

Scholes, Myron and Joseph Williams (Dec. 1977). "Estimating betas from nonsynchronous data". In: *Journal of Financial Economics* 5.3, pp. 309–327.

Vasicek, Oldrich A. (1973). "A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas". In: *The Journal of Finance* 28.5, pp. 1233–1239.

Figure 1: Winsorization Techniques



Explanations: The left column shows data drawn from a true $r_i = 0 + 1 \cdot r_m + \epsilon$ (black line). The estimated OLS line (in blue) happens to be a little flatter in this draw. The middle column illustrates the method of winsorization. Level winsorization truncates the dependent variable (the firm's own stock return). Slope winsorization truncates returns based on a beta prior. The right column shows the winsorized returns as well as the new OLS estimated line.

Interpretation: Level winsorization flattens the estimated slope. Slope winsorization is harshest where it matters least for the regression coefficient estimates, i.e., around r_m=0.

Source: fig-conceptual/conceptualplots.Rmd

Table 1: Idealized Properties of Market-Beta Estimators' Errors

		Distance from True Beta					
	Return	"Normal Scenario"			"Empirical Scenario"		
Estimator	Inputs	Mean	SD	RMSE	Mean	SD	RMSE
bols	252 Days	–0.000	0.188	0.188	–0.000	0.184	0.184
bvck	252 Days	–0.016	0.199	0.200	–0.016	0.183	0.184
bmblm	1,260 Days	0.002	0.331	0.331	0.002	0.320	0.320
bsw	252 Days	–0.015	0.170	0.170	–0.017	0.176	0.177

Explanations: The table is based on 1 million resampled draws of each (252×5) market rates of return, (252×5) resampled/simulated market-model residuals, and one resampled/simulated market-model sigma with associated beta. The "normal" columns parameterize the firm-specific inputs (log-normal market-model sigma, betas linear in log sigma (with normal error), and normal market-model residuals). The "empirical" columns sample directly from the empirical joint distribution of market-model sigmas and betas (winsorized at 0.1% and 99.9%), and from the empirical distribution of unit-normalized market model residuals. The estimators use only the stock and market rates of returns. This table summarizes the beta prediction errors of four estimators. bmblm is the Bloomberg-Merrill-Lynch (Blume) estimator (2/3 × bols + 1/3 × 1) using compounded monthly returns, bols is the standard OLS estimator, bvck is the Vasicek (1973) estimator (given both the true mean and the true standard deviation of market betas in this sample), and bsw is the slope-winsorized estimator.

Interpretation: The slope-winsorized market beta, with its lower RMSE, outperforms the alternatives.

Source: simul/1simul-emp.Rmd

Table 2: Descriptive Statistics of *Ex-Ante* Beta Estimates, 1973–2019

		Return	1,000 Big Stocks		3,000 Big Stocks	
Beta Estimates	Abbrev	Freq	mean	sd	mean	sd
OLS Beta	bols	Daily	0.99	0.49	0.95	0.55
Vasicek (1973) Beta	bvck	Daily	0.97	0.45	0.93	0.51
Slope Winsorized	bsw	Daily	0.96	0.42	0.92	0.46
Slope Winsorized/Age-Decay	bswa	Daily	0.96	0.41	0.92	0.45
Frazzini and Pedersen (2014)	bfp	Hybrid	1.05	0.30	1.03	0.33
OLS Beta	bmols	Monthly	1.07	0.55	1.12	0.64
Vasicek (1973) Beta	bmvck	Monthly	1.04	0.46	1.08	0.50
Bloomberg	bmblm	Monthly	1.04	0.37	1.08	0.43
			N=351,155 (Overlapping Month Ends)		N=68,267 (Dec Only)	

Explanations: The 1,000-stock sample overlaps 252-day stock returns; the 3,000-stock sample uses only January-to-December non-overlapping data. Daily-based estimators use 252 days of stock returns. Monthly-based estimators use 60 months of stock return data. The sample period is 1973 to 2019. The 1,000 or 3,000 stock subsamples were ranked by marketcap contemporaneous with the beta estimates. These two subsamples are used in all subsequent tables.

Table 3: Distinctness of *Ex-Ante* Beta Estimates, 1973–2019

Panel A: 1,000 Biggest Stocks, Overlapping, on Month Ends

	Daily Returns			Hybrid	Monthly Returns		
	bvck	bsw	bswa	bfp	bmols	bmvck	bmblm
bols	0.05	0.11	0.14	0.31	0.46	0.42	0.39
bvck	0	0.06	0.11	0.29	0.45	0.40	0.37
bsw		0	0.08	0.27	0.45	0.39	0.35
bswa			0	0.26	0.44	0.38	0.34
bfp				0	0.41	0.33	0.27
bmols					0	0.12	0.19
bmvck						0	0.12

Panel B: 3,000 Biggest Stocks, December Only

	Daily Returns			Hybrid	Monthly Returns		
	bvck	bsw	bswa	bfp	bmols	bmvck	bmblm
bols	0.08	0.15	0.19	0.38	0.59	0.51	0.49
bvck	0	0.09	0.14	0.36	0.59	0.49	0.47
bsw		0	0.10	0.33	0.58	0.47	0.45
bswa			0	0.32	0.58	0.46	0.44
bfp				0	0.51	0.37	0.34
bmols					0	0.21	0.22
bmvck						0	0.14

Explanations: These are the pooled root mean square distances between beta estimators, $\sqrt{1/T \cdot \sum_t (b_t^{(A)} - b_t^{(B)})^2}$, where (A) and (B) denote different beta estimates for the same stock in month t, in the two subsamples from Table 3.

Interpretation: Beta estimates based on daily stock returns are quite different from those based on monthly stock returns.

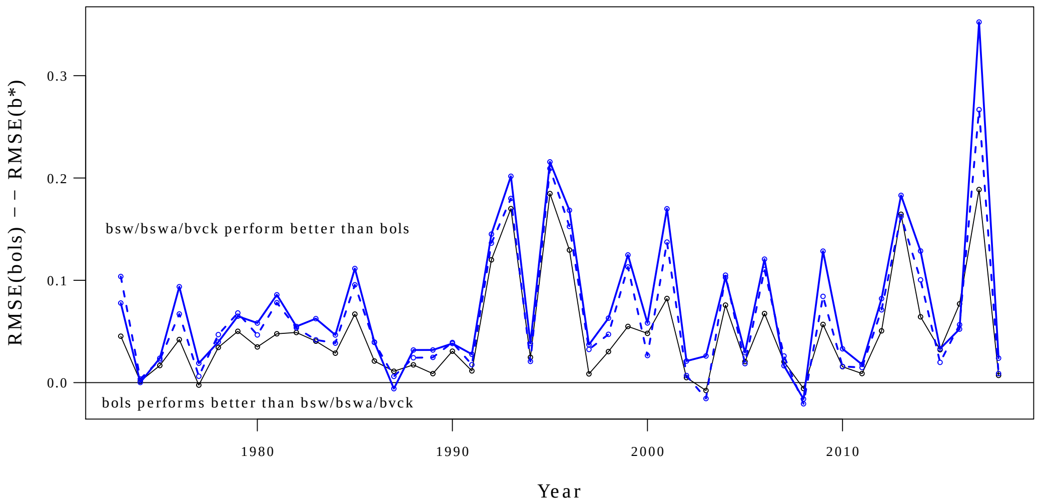
Table 4: Predicting Future *Ex-Post* (One-Year-Ahead) OLS Market Beta, 1973–2019

(Ex-Post) Predicted	(Ex-Ante) Predictor	1,000 Big Stocks γ_b	$se(\gamma_b)$	$\bar{R}^2$	RMSE	3,000 Big Stocks γ_b	$se(\gamma_b)$	$\bar{R}^2$	RMSE
bols	bols	0.744	0.004	0.580	0.332	0.703	0.003	0.517	0.411
	bvck	0.798	0.004	0.578	0.322	0.776	0.003	0.526	0.391
	bsw	0.862	0.004	0.580	0.314	0.859	0.004	0.534	0.377
	bswa	0.897	0.004	**0.604**	**0.302**	0.895	0.004	**0.554**	**0.366**
	bfp	1.063	0.007	0.447	0.360	0.986	0.008	0.362	0.441
	bmols	0.469	0.005	0.297	0.501	0.391	0.005	0.215	0.643
	bmvck	0.550	0.005	0.289	0.456	0.513	0.006	0.220	0.551
	bmblm	0.703	0.007	0.297	0.417	0.587	0.008	0.215	0.529

Explanations: The two samples are the same as in the previous tables. The reported statistics pool cross-section and time-series. γ_b, $se(\gamma_b)$, $\bar{R}^2$, and RMSE are from the same regression. (The intercept is included, but not reported. The set of 1,000 stocks predicts overlapping stock-months (thus forecasting one *ex-post* beta from, say, January-to-December followed by another from February-to-January). The set of 3,000 stocks predicts only calendar year-end months (from January-to-December). The overlapping months regressions on the left use 12 Newey–West lags for their standard errors.) RMSE is the root-mean-squared prediction error between x and y (which is not impervious to differential mean bias). The dependent variable is always the 1-year ahead ("*ex-post*") OLS market beta, calculated from 252 days of daily stock returns. The independent variables are the estimates from methods defined in Section II. The first four estimators are based on *ex-ante* daily stock return data, the last three estimators are based on *ex-ante* monthly stock return data, and bfp uses both daily and monthly returns. bols and bmols are OLS betas. bvck and bmvck are Vasicek betas. bmblm is the Bloomberg beta also popular on most finance websites. bfp is the Frazzini and Pedersen (2014) beta. bsw and bswa are the novel slope-winsorized betas. bswa weights older returns progressively less.

Interpretation: The slope-winsorized market betas outperform all alternatives.

Figure 2: Year-by-Year Relative Predictive Performance of Estimators, Big 3,000 Stocks, Dec 1973–2019



Explanations: This plot is equivalent to Table 4, but shows the differences in the RMSEs of the predictive regressions when run year by year (Decembers only, top-3,000 stocks) versus an *ex-ante* OLS predictor. The zero line is the OLS predictive performance. Higher values indicate superior performance over the OLS beta. The solid blue line shows the bswa estimator, the dashed blue line shows the bsw estimator, the black line shows the bvck estimator.

Interpretation: The performance of bvck, bsw, and bswa is better in the second half of the sample. The estimators do not seem to change their performance ordering over the sample.

Table 5: Multiyear Performance, 1973–2019, 2000 Stocks, Annual

Panel A: Adjusted R^2, in percent

	Daily-Frequency Returns, Future Beta over					Monthly Freq
	1y	2y	3y	5y	10y	5y-MO
bmols	24.42	17.49	12.77	7.17	4.68	21.35
bmvck	24.47	17.61	12.83	7.07	4.54	21.36
bols	51.99	39.03	30.97	20.21	5.92	25.25
bvck	52.62	39.97	31.94	20.86	6.10	25.17
bsw	53.45	40.74	32.62	21.32	6.59	25.35
bswa	56.23	42.38	34.08	22.38	7.12	26.59
N	66,851	62,400	56,626	45,872	25,179	44,462

Panel B: RMSE

	Daily-Frequency Returns, Future Beta over					Monthly Freq
	1y	2y	3y	5y	10y	5y-MO
bmols	0.5913	0.6217	0.6451	0.6715	0.6206	0.6190
bmvck	0.5163	0.5361	0.5555	0.5799	0.5501	0.5627
bols	0.3939	0.4410	0.4740	0.5218	0.5878	0.5742
bvck	0.3768	0.4154	0.4450	0.4895	0.5539	0.5624
bsw	0.3642	0.3959	0.4223	0.4635	0.5244	0.5497
bswa	0.3502	0.3852	0.4117	0.4531	0.5176	0.5401
N	66,851	62,400	56,626	45,872	25,179	44,462

Explanations: This table shows predictions of longer-horizon market-betas. Except for 5y-MO, all ex-post betas are calculated with daily-frequency stock returns (of which at least 100 had to be available to estimate a beta). The predictive regressions are run on a calendar-year basis (December only). The regression R^2 and plain RMSE in predicting the future OLS beta are akin to those in Table 4.

Interpretation: The daily stock return-based betas always predict future market-betas better regardless of target beta horizon and frequency used to calculated the target beta. Among daily-frequency-based betas, the same performance ordering always holds: bols is worse than bvck, which is worse than bsw, which is worse than bswa.

Source: Output comes from 1longtermols.R Jun 1, 2021.

Table 6: Hedged Return Performance, 1928–2019, 3,000 Stocks

	SD	SDpct	Min	1%	5%	25%	Median	Mean
$r_i - bols_i \cdot r_m$	0.5491	0.3316	–1.88	–0.943	–0.6375	–0.2371	–0.0121	0.0347
$r_i - bvck_i \cdot r_m$	0.5463	0.3301	–1.67	–0.909	–0.6250	–0.2348	–0.0118	0.0377
$r_i - bsw_i \cdot r_m$	0.5458	0.3290	–1.58	–0.899	–0.6177	–0.2336	–0.0116	0.0392
$r_i - bswa_i \cdot r_m$	0.5457	0.3285	–1.60	–0.899	–0.6180	–0.2333	–0.0120	0.0390

Explanations: This table varies both the time-period and the objective function. For each firm-year, the quantity $r_{i,y} - \hat{b}_{i,y-1} \cdot r_{m,y}$ (the hedged portfolio rate of return) is calculated, where r are the annual compounded rates of return and betas were estimated using daily stock returns in the prior year. (At least 100 returns had to be available to estimate a beta.) The reported statistics are over all firm years. SDpct is the percentile equivalent of the standard deviation under the normal distribution (i.e., the difference between the returns at the (approx) 20th to the 70th percentile). There are 182,379 non-overlapping firm-calendar-years in the four predictive regressions.

Interpretation: The hedged portfolio is less risky when the hedge is based on the prior bswa or bswa.

Source: Output comes from 0mkbetas-retvar.Rmd May 30, 2021.

Table 7: Predicting Future *Ex-Post* Betas Other Than the OLS Beta, 1973–2019

(Ex-Post)	(Ex-Ante)	1,000 Big Stocks				3,000 Big Stocks			
Predicted	Predictor	γ_b	se(γ_b)	$\bar{R}^2$	RMSE	γ_b	se(γ_b)	$\bar{R}^2$	RMSE
bswa	bsw	0.759	0.003	0.600	0.279	0.742	0.003	0.571	0.319
	bswa	0.791	0.003	**0.624**	**0.266**	0.772	0.003	**0.591**	**0.307**
bvck	bvck	0.748	0.003	0.583	0.308	0.722	0.003	0.542	0.364
	bsw	0.810	0.004	0.588	0.296	0.801	0.003	0.553	0.345
	bswa	0.844	0.003	**0.612**	**0.284**	0.834	0.003	**0.573**	**0.333**
bfp	bfp	0.451	0.006	0.192	**0.322**	0.450	0.006	0.186	**0.361**
	bsw	0.381	0.004	0.270	0.381	0.384	0.004	0.263	0.431
	bswa	0.397	0.004	**0.282**	0.372	0.402	0.004	**0.275**	0.420
bmols	bmols	0.482	0.006	0.237	0.557	0.424	0.007	0.194	0.668
	bsw	0.680	0.007	0.273	0.492	0.642	0.009	0.229	0.592
	bswa	0.709	0.007	**0.284**	**0.484**	0.669	0.009	**0.238**	**0.584**
bmvck	bmvck	0.499	0.006	0.251	0.464	0.462	0.006	0.226	0.500
	bsw	0.591	0.006	0.287	0.431	0.530	0.006	0.258	0.485
	bswa	0.616	0.006	**0.299**	**0.423**	0.552	0.007	**0.268**	**0.476**
bmblm	bmblm	0.482	0.006	0.237	**0.371**	0.424	0.007	0.194	**0.445**
	bsw	0.454	0.005	0.273	0.392	0.428	0.006	0.229	0.469
	bswa	0.473	0.005	**0.284**	0.383	0.446	0.006	**0.238**	0.458

Explanations: In this table, the predicted beta is no longer the future bols, but other future beta estimates (as indicated in the left-most column). To avoid overlap, monthly independent variables were lagged by 60 months.

Interpretation: The *ex-ante* bswa also predicts other *ex-post* estimators better than these estimators can predict their future *ex-post* selves.

Table 8: Summary Statistics by Dollar Trading Volume Deciles, 1974–2019

		Least Traded								Most Traded	
		Lo 1	2	3	4	5	6	7	8	9	Hi 10
Mean	$V	0.01	0.04	0.1	0.3	0.6	1.4	2.8	6.0	14.9	95.4
	No Trade	35%	21%	14%	10%	7%	5%	2%	1%	0%	0%
Mean	bols	0.23	0.37	0.52	0.67	0.79	0.87	0.94	1.00	1.08	1.19
	bswa	0.36	0.46	0.58	0.71	0.80	0.86	0.91	0.96	1.02	1.12
	bdim	0.32	0.49	0.63	0.78	0.89	0.98	1.06	1.13	1.20	1.23
SD	bols	0.56	0.55	0.58	0.62	0.62	0.59	0.57	0.56	0.56	0.57
	bswa	0.29	0.33	0.38	0.42	0.44	0.42	0.42	0.41	0.41	0.42
	bdim	0.76	0.73	0.76	0.75	0.74	0.72	0.69	0.68	0.68	0.67

Explanations: Firm-years are sorted into deciles, based on the mean dollar trading volume of the last 22 days of the (contemporaneous) calendar year. The sorts are first by calendar year, then trading volume. Decile 1 (10) are the least (most) traded stocks. To be included, a firm had to have a valid bols, bswa, and bdim in the following calendar year. There is no firm-size restriction in this sample, but at least 100 returns had to be available to estimate beta. Stocks that had no beta the following year were omitted. The observations are the same within each column, but different across rows. “No Trade” is a dummy taking one if CRSP reports a negative price on the last day with a valid rate of return in its calendar year.

Interpretation: There are four orders of magnitude difference in trading activity across deciles. More frequently-traded stocks have higher betas. bswa has the lowest heterogeneity across and within trading deciles, bdim the highest.

Table 9: Performance of Dimson Estimator, 1974–2019

Panel A: Predicting Future Calendar Year OLS Beta (bols)

	with bols			with bswa			with bdim		
\$V	γ_b	$\bar{R}^2$	RMSE	γ_b	$\bar{R}^2$	RMSE	γ_b	$\bar{R}^2$	RMSE
Lo 1	0.09	0.007	0.773	0.44	0.047	0.601	0.11	0.020	0.890
2	0.22	0.042	0.709	0.53	0.092	0.580	0.16	0.040	0.842
...									
5	0.53	0.274	0.608	0.84	0.336	0.516	0.34	0.159	0.759
6	0.55	0.304	0.561	0.83	0.362	0.477	0.36	0.192	0.711
...									
9	0.60	0.387	0.478	0.86	0.438	0.411	0.45	0.318	0.598
Hi 10	0.68	0.493	0.432	0.94	0.523	0.384	0.54	0.439	0.517

Panel B: Predicting Future Calendar Year Dimson Beta (bdim)

	with bols			with bswa			with bdim		
\$V	γ_b	$\bar{R}^2$	RMSE	γ_b	$\bar{R}^2$	RMSE	γ_b	$\bar{R}^2$	RMSE
Lo 1	0.14	0.009	0.921	0.49	0.036	0.779	0.14	0.017	1.012
2	0.26	0.033	0.877	0.57	0.062	0.772	0.16	0.023	0.983
3	0.38	0.084	0.806	0.64	0.115	0.715	0.24	0.057	0.926
...									

Explanations: Decile 1 (10) are the least (most) traded stocks, as in Table 8. All betas are calculated over all daily stock returns within their calendar years. Panel A predicts the future calendar-year OLS beta. Panel B predicts the future Dimson beta. At least 100 days of complete data had to be available to calculate beta. The dependent beta variables were for the succeeding calendar year. γ_b and $\bar{R}^2$ are from the forecasting regression. RMSE is the root-mean-squared error (difference). Predictions in each row are for the same firm-years and thus directly comparable. This is obviously not the case for columns.

Interpretation: Panel A: The high-trading decile (where non-synchronicity is implausible) shows the bdim efficiency loss. Panel B: bdim cannot predict its future self even in the low-trading decile better than bswa can predict it.

Source: 1dimson.R, June 1, 2021

𝒜 Appendix: Implementation of BSWA

The complete R code is

```
## _bswb is an internal function doing most of the work
.bswb <- function( ri, rm, Delta, rho ) {

   ## function to winsorize r based on lower and upper bounds
   wins.rel <- function( r, rmin, rmax ) {
      rlo <- pmin(rmin,rmax); rhi <- pmax(rmin,rmax)
      ifelse( r<rlo, rlo, ifelse( r>rhi, rhi, r ) )
   }

   wri <- wins.rel( ri, (1-Delta)*rm, (1+Delta)*rm )

   ## note: ri and rm must be ordered in time (increasing)
   bsw <- coef(lm( wri ~ rm, w=exp(-rho*(length(ri):1)) ))[2]
}

## the externally visible wrapper functions
bsw <- function( ... ) .bswb( ... , Delta=3.0, rho=0.0 )
bswa <- function( ... ) .bswb( ..., Delta=3.0, rho=2.0/256.0 )

# Test:
# d <- read.csv("dailyreturns.csv")
# with(d, cat("BSW Beta=", bsw(ri, rm), "; BSWA Beta=", bswa(ri,rm),"\n"))
```

Stock returns can also be added progressively to update the estimator in time, so an alternative C program (not shown) takes less than one minute on a good desktop computer to calculate all bswa betas for the entire CRSP universe.

The implementation code is also available from the author's and journal's website.