Simpler Better Market Betas

Ivo Welch

August 17, 2020

Abstract

My paper proposes a robust and easy-to-implement one-pass beta estimator: Justified by the market-model itself, daily stock returns are first winsorized at –2 and +4 times the contemporaneous market return. The resulting "slope-winsorized" betas outpredict all other prominent estimators, including not only the Bloomberg-Merrill Lynch (Blume (1971)) beta (nearly ubiquitous on all financial websites) and the better Vasicek (1973) estimator, but also estimators that require intra-day data, super-computers, and financial statements. Adding simple age decay further improves the estimates.

Keywords: Market-Beta, Factor Exposures, Robust Estimation

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

Conflicts of Interest: None.

The most important moments in finance are expected rates of return, own volatilities, and factor exposures (especially market-beta).

Expected returns are the most interesting moment to forecast. They are also the most difficult. Lo and MacKinlay (1990), Harvey, Liu, and Zhu (2016), and others, have forcefully pointed out how historical data correlations can be oversearched and thus spurious. Mclean and Pontiff (2016) have pointed out how investors' research can itself erode past associations. This could even render average returns intrinsically unforecastable *despite* seemingly good empirical historical associations (Martin and Nagel (2020)).

Volatilities are important not only in the context of pricing options but also because they seem to have an empirical associations with future rates of return (Ang et al. (2009)). Volatilities are usually easier to estimate than expected returns, because second moments improve with higher-frequency sampling. Engle (1982), Glosten, Jagannathan, and Runkle (1993), and others, have even shown how to reliably predict time-varying volatility.

My paper posits that market betas are the third intrinsically interesting moment. Thus, it explores beta primarily for the sake of itself—as the market-related hedge ratio of assets. Beta's most useful application is in its original context, i.e., that of measuring how individual stocks contribute to the risk of a portfolio similar to the market. The market factor is also approximately the first principal component of all stock returns, and many investors (esp. index funds and smart-beta near-index funds) indeed hold highly-diversified portfolios not too dissimilar from it. Betas can inform such investors how tilting stocks in and out of the market portfolios changes the overall portfolio risk. The CFA exam covers market-beta and the Bloomberg-Merrill Lynch adjustment. The use of beta is also mandated in some regulatory contexts. All of this is uncontroversial.

My paper does not explore the controversial association of betas with average rates of return (although I discuss this issue in Appendix $\mathscr{D}$). My paper also does not estimate the betas of large well-diversified portfolios. This is because there is less to estimate—in the extreme, the beta of the market-portfolio is 1 and there is nothing to estimate. Beta is a useful measure only when starting from or considering disaggregated assets. And my paper does not discuss betas with respect to other factors. Unlike for the stock market, it is difficult not only to obtain agreement on what the important factors are, but also to hedge many of these factors (due to the lack of shortable factor-tracking portfolios).

The new estimator proposed in my paper is easy to use. It is a single-pass estimator based on a prior that the relation between stock returns and the market should not be too far from 1.0. This prior is *not* used in a Bayesian framework,[1] but in a robust estimation framework. Stock returns exceeding extreme slope bounds are deemed more likely to be uninformative outliers than informative observations and are thus winsorized.

The full estimation recipe is as follows:

0. Using between one and three year's worth of *daily* stock rates of return,
1. winsorize each daily rate of return at $-2 \cdot r_{m,d}$ and $+4 \cdot r_{m,d}$, and
2. run a plain OLS market-model regression on these winsorized daily rates of return.

I refer to this estimator as bsw (for "beta slope-winsorized").

Despite its ease of implementation, this estimator outperforms all other prominent estimators I am aware of.[2] This includes not only versions of the OLS (and Bloomberg-Merrill-Lynch variants, nowadays nearly ubiquitous on financial sites), Vasicek (1973), and Dimson (1979) estimators, but also estimators that require intra-day data (Ait-Sahalia, Kalnina, and Xiu (2014)), super-computer based intensive calculations (Martin and Simin (2003)), and financial statement information (Cosemans et al. (2016)). They do so also in samples of the biggest 1,000 or the biggest 3,000 stocks, which are the universes of most interest to fund managers.

In terms of magnitude, in an idealized scenario with data-matching outlier presence, the bsw estimator should offer an RMSE improvement of 40-45% over the Bloomberg beta and 5 to 15% over daily-data-based Vasicek and OLS estimators. In forecasting actual future daily OLS betas in the data (which are themselves betas measured with error), the

[1] I show that my beta outperforms the Vasicek (1973) estimator, which is based on a Bayesian framework.

[2] The primary criterion to evaluate the performance of slope-winsorized and other betas is how well they forecast *realized future OLS betas* (Rosenberg and Guy (1976b), Harrington (1983)). The secondary criterion 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 even if these estimates serve some other useful function, a researcher should still use the slope-winsorized betas when in predicting their *future* realizations.

improvement of bsw over the Bloomberg beta is about 30%, 8-10% over the OLS beta, and 4-5% over the Vasicek beta—about as suggested by the simulations.

Furthermore, an even (modestly) better version replaces the OLS market-model with a WLS market-model, in which the weight of each observation decays with age. A good decay parameter is 2/252 per day, suggesting a half-life of about 90 trading days. I refer to this estimator as bswa (for "beta slope-winsorized aged"). Its complete 10(!)-line implementation in R is in Section III.

The two parameters (1.0 ± 3.0 on the winsorization, and 90 trading days on the age decay) are low-dimensional, static, and not fragile. The slope-winsorized beta estimators do not use different parameters per stock, per time unit, or per stock-time unit. They use only these two static fixed numbers for the *entire* CRSP sample, and they were only moderately searched from a rather coarse grid (mostly before the estimation period). Winsorization levels of 1.0 ± 2.0 or 1.0 ± 4.0 (instead of 1.0 ± 3.0), and half-lives from 75 to 120 days give nearly identical performance. Further investigation suggests that any potential gains from further fine-tuning the two parameters are likely to be small. Even *ex-post* optimal parameters within many subsamples improve the predictive performance only modestly (Appendix $\mathscr{B}$).

I Data and OLS Betas

The data set for the analysis is the early-2020 CRSP data base. It contains 75,230,516 daily stock returns with share codes 10 and 11 from 1926 to 2019. For each stock-month, I run a 12-month OLS market-model regression on daily returns.[3] The market-model dependent variable is each stock's own rate of return net of the risk-free rate of return (from Ken French's website). The independent variable is the CRSP value-weighted market rate of return, again net of the same risk-free rate of return. When a stock does not have at least 20 trading-days worth of stock returns, or when a return is not present on the last trading day,

[3] 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 windows has a more natural relation to the calendar and performs nearly as well.

no beta is calculated. This daily-returns-based OLS estimator and its resulting estimates are abbreviated as bols.

A "naive" concurrent historical bols yields a beta forecast that is an obvious benchmark against which alternative historical beta forecasts should be measured. Moreover, bols also supplies the future estimation target that is to be forecast, because the true underlying stock betas are unobservable. Beyond being the direct sample equivalent of the population parameter—i.e., beyond being a reasonable proxy for the true beta—the *realized* OLS beta is itself of interest, too: It is the best *realized ex-post* hedge ratio.[4]

The forecasting betas are always calculated strictly before the OLS beta they usually need to forecast. There is never any overlap between predicting and predicted beta.

In addition, end-of-month marketcap ranks from CRSP (contemporaneous with the predicting beta) are used to limit the analysis to the prevailing biggest 1,000 or 3,000 stocks. (Not shown, similar results obtain in the sample of all CRSP stocks.) My analysis begins in 1973 (just after NASDAQ came online). About 0.5% of the biggest 1,000 firms' months and 5% of the biggest 3,000 firms' months are lost to survival attrition.

[4]For an analogy, researchers predict future *realized* returns with historical models that are often based on factors aggregating and smoothing these return predictions. Their tests neither predict factor-model based predicted returns nor use the historical own realized average returns as the (sole) predictor. It is the same with betas. Researchers predict future realized betas (which are the ex-post perfect realized market-hedges) not only with historical beta estimates but also with beta estimates obtained with other models.

II Existing Beta Estimators

Table 1 provides a glossary of beta estimators used in my paper for quick reference.

[Insert Table 1 here: **Glossary of Market-Beta Estimators**]

Vasicek (1973): Shrinkage / Random-Effects Betas: The Vasicek (1973) estimator, bvck_i, is calculated with daily-frequency stock returns in my paper. It can be viewed either as a Bayesian shrinkage estimator or as a random-effects panel estimator. It requires first computing bols_i and its standard error "σ_i^2" for each stock. 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 then

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

bvck can be an optimal estimator, but the necessary assumptions are unrealistic in our stock-return context. The underlying betas are themselves time-changing. Even the cross-sectionally presumed-constant moments ($\overline{\text{bols}}_t$ and $\overline{\sigma}_t$) are time-changing and contribute to the good performance of bvck over bols..[5] My paper shows that most of the good performance of the Vasicek comes from its good handling of outliers (aggressively deemphasizing slopes when a stock return series has an outlier), although it was *not* specifically designed for this.

Monthly Stock-Return Based Estimators: Betas based on monthly-frequency stock returns are far and away the most common estimates of betas, both in practice and in research—presumably because they are easiest to calculate. To improve reliability, the estimation window has to be expanded from 1 year to 5 years, empirically a good

[5] The composition of stocks changes over time, too. bvck is often empirically badly biased. Levi and Welch (2017) suggest a debiased estimator of $0.20 + 0.75 \cdot \text{bvck}_i$. For predictive purposes, the R^2 performance metric of the Levi-Welch estimator is identical to that of bvck, because it is a simple linear transform. Thus the Levi-Welch estimator is omitted here.

self-predictive window and in wide use.[6] The OLS beta calculated from monthly stock returns is named bmols. The equivalent monthly-return Vasicek estimator is named bmvck.

Blume (1971) first suggested debiasing the OLS estimator using an empirically estimated linear correction. His Table 4 shows a mean adjustment of 0.64. The *Merrill-Lynch* (now Bloomberg) beta simplifies the shrinkage to 2/3. Covered on the CFA exam and distributed by Capital IQ to financial websites, it can be found, e.g., on the Yahoo!finance and Google finance sites. This beta is thus perhaps the one in most common use. In the paper, this beta is named bmblm.

Frazzini and Pedersen (2014): Deconstructed Betas: The Frazzini and Pedersen (2014) beta estimator (bfp) uses five years of daily stock return data to calculate correlations and one year data to calculate standard deviations. The estimator itself is the correlation multiplied by the ratio of the two standard deviations, debiased with a linear transform of $0.6 \cdot \hat{b} + 0.4$. The origin of the estimator is not clear, because Frazzini and Pedersen (2014) do not validate or benchmark it. Nevertheless, perhaps due to the success of bfp in predicting future average rates of return, bfp has become popular.

With their bfp as beta estimate, Frazzini and Pedersen (2014) have claimed a negative and empirically meaningful association. Yet, there is now some disagreement about this. The negative association has been disputed by Novy-Marx and Velikov (2018) and Han (2019), who show that the success of "betting-against-beta" was not primarily due to a better beta measure but due to (1) a mixing of time-varying volatility into the Frazzini-Pedersen estimated betas and (2) time-varying investment amounts.[7]

Martin and Simin (2003): Robust Betas: In the beta estimation context, robust estimation was first used by Martin and Simin (2003). They developed a non-linear two-equation system to establish a bandwidth in one equation that is used with a band-like shrinker in another equation.

[6] My findings reported below are robust to window lengths from 36 to 120 months. Shorter windows yield poorer and less reliable estimates. Longer windows deteriorate due to drift in the underlying true beta.

[7] Novy-Marx and Velikov (2018) shows how this estimator is biased in the time-series, picking up firm-specific time-changing volatility patterns. Han (2019) shows that most of the average return performance comes from the time-series component.

Unfortunately, the Martin-Simin maximum-likelihood betas (bmm) are difficult to compute. Fortunately, Timothy Simin generously shared a set of similar estimates with me, henceforth named bmm. These betas are similar, but use a Huber (1964) loss function instead of an optimal one.[8] According to Simin, the Huber estimates are nearly identical, but their computations are five times faster. Nevertheless, it took days running distributed Matlab on the Penn State super-computer to calculate them. The Simin estimates begin in June 1966 and end in June 2018. They contain 2,023,162 firm-months. To be included, a firm-year had to have a CRSP share code 10 and 11 firms and at least 240 daily observations.

Ait-Sahalia, Kalnina, and Xiu (2014): Intra-Day Betas: Ait-Sahalia, Kalnina, and Xiu (2014) develop a non-parametric time series regression estimator without the usual assumption of piecewise linearity, using intra-day 5 minute data.[9] Unfortunately, the betas are difficult to replicate. Fortunately, Dacheng Xiu also generously shared their beta estimates, henceforth named bakx. Their beta estimates are calculated based on one month each.

Cosemans et al. (2016): Fundamental and Intra-Day Betas : Cosemans et al. (2016) propose a number of different estimators. For the S&P 100 stocks, they shrink rolling betas estimated using a semi-annual window of daily returns to a fundamentals-based prior beta[10] obtained from a Bayesian panel regression on monthly data (i.e, monthly stock and market returns and the monthly firm characteristics and business-cycle variables used as conditioning variables). For obtaining out-of-sample beta forecasts, the panel regression is estimated using an expanding window.[11] Unfortunately, the Cosemans et al. (2016) betas are involved enough to be difficult to replicate.

[8] Rousseeuw (1984) explains how these betas are "maximum likelihood type estimators."

[9] The paper is a tour de force. Because the intraday TAQ data is not particularly clean, the data analysis (and the econometric methodology, the estimations, and the large data storage requirements) require near-Herculean efforts. Not shown, I also experimented with a 12-month average. The forecasting results improved but not enough to change the conclusions.

[10] Other accounting-based estimates (e.g., as proposed in Rosenberg and Guy (1976b)) perform poorly. Further investigation suggests that they are of use only with monthly stock returns, and obsoleted by daily stock returns. Harrington (1983) examined only monthly-return based betas. Accounting data adds no useful information once daily-return based betas.

[11] For validation, they did calculate future intra-day betas based on 15 minute intervals over different horizons.

Fortunately, Mathijs Cosemans also generously shared their beta estimates for the S&P 100 stocks with me, henceforth named bcfsb.

Appendix $\mathscr{D}$ discusses the Dimson (1979) and Schols and Williams (1977) estimators. In brief, these have surprisingly high estimation efficiency costs and are dominated in this use case where stocks with daily closing prices are readily available. They cannot predict their future selves better than other estimators can predict them.

III The Slope-Winsorized Beta Estimators

My estimators belong to the class of robust winsorizing methods. Such estimators require an "aggressiveness" parameter, which sets the tradeoff between type-I and type-II errors: correctly winsorizing unpredictive outliers vs. incorrectly winsorizing predictive outliers. Good choices trade off being too lax (thereby not having any effect) vs being too strict (thereby pushing all beta estimates too close towards the same value).

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

The most common robust estimator winsorizes 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 prior centers not on 0.0 but on 1.0.[12]

Level winsorization is so common that analysts often apply it to data before the models are even considered. 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. A user should not want to reduce such outliers when estimating volatility.

[12] 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 of return. Appendix $\mathscr{E}$ discusses another alternative, band winsorization. Band winsorization works well, but it has not been used in the literature afaik and it does not outperform slope winsorization.

However, when estimating betas, she may want to 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, one may want to reduce outliers less aggressively when estimating volatilities than when estimating betas.

My paper introduces a novel "slope-winsorized" beta estimator (bsw). "Slope" winsorization limits returns based on minimal and maximal coefficient slopes:

$$\mathrm{rsw}_{i,d} \in \left(1.0 + [\,-\Delta, +\Delta\,]\right) \cdot \mathrm{r}_{m,d}\ , \tag{2}$$

where Δ is the winsorization parameter—set to 3.0 in my paper, i.e., leaving a return range limited to $(1.0 \pm 3.0) \cdot \mathrm{r}_{m,d}$, and $\in$ denotes the winsorization.[13] This linear structure between stock return and market return is provided by the market-model itself. The slope-winsorized beta is then

$$\mathrm{bsw}_{i,y} \equiv \frac{\mathrm{cov}\left[\ \mathrm{rbsw}_{i,d}(\Delta)\ ,\ \mathrm{r}_{m,d}\ \right]}{\mathrm{var}(\ \mathrm{r}_{m,d}\ )}\ . \tag{3}$$

The bottom plot in Figure 1 illustrates slope winsorization.[14] This estimator is custom-designed for our application: It relies on the knowledge that the mean market-beta is 1, that the daily intercept is near 0, and that few stocks have extreme market betas (see also Appendix $\mathscr{F}$).

[13] Delta was chosen roughly where the monotonic relation between historical and future OLS betas breaks down. Betas this extreme are rare. In the CRSP sample, only 0.17% of all bols estimates exceed slopes of –1 and +4.

[14] Note that it makes no sense to report the number of affected observations. Appendix $\mathscr{E}$ discusses band winsorization, an alternative (though never having been used in the literature) performs almost as well. The choice between band and slope winsorization comes down to a user preference for specification of the prior: is it more natural to specify reasonable (beta) slopes or reasonable (return) residuals? There is also no harm done (but not much gained, either) by combining the two.

IV The Properties of The Estimator

We first want to assess the properties of the bsw estimator under return process characteristics deemed relevant. There are two aspects of special interest: (1) The influence of stock return outliers; (2) The use of the information that the average true market-model beta is 1. This information is used not only by the slope-winsorized estimator, but also by bmblm and bvck estimators.[15] The assessment is generous to the bmblm estimator, in that it is assumed that the underlying betas do not change. Thus, stock returns from 5 years ago remain as relevant as more recent stock returns (as used in bols, bvck, and bsw).

Unfortunately, there are no closed-form expressions for the power of *any* winsorizing (and perhaps even non-linear) slope estimators. Moreover, stock-return outliers do not follow any known distribution.[16] Fortunately, simulations in large samples yield nearly exact results to the question of interest: how well can one expect these estimators to perform?

The appendix provides a more detailed description of the analysis and the complete R code. In brief:

Data Generation: The analysis is based on the properties of the top-3,000 stocks (based on the previous calendar-year end) from 1974 to 2019 with at least 126 returns in the calendar year. There are two guiding cases:

- A parameterized "normal" case, in which true market-betas are a linear function of log-normally distributed market-model sigmas, and market-model residuals are normally distributed.
- A non-parameterized "empirical" case, in which OLS market-betas and market-model sigmas are sampled from the actual empirical distribution of calendar-year market-betas (see also Appendix $\mathscr{F}$), as are the (sigma-scaled) market-model residuals. These residuals retain their empirical skewness and kurtosis. The actual return sampling approach is analogous to that in Brown and Warner (1985).

[15] The estimators also use other information. The Vasicek estimator uses information about the heterogeneity of betas. The slope estimator uses information that the alpha is close to zero and its weighting parameter.

[16] Kurtosis is difficult to estimate. The average kurtosis also depends on the market-model window length. Being right-tailed, it is even remarkably difficult just to generate realistically in simulations!

The analysis is based on thus-simulated daily stock returns computed for 252*5 days. Daily estimators had access to 252 data points, the monthly bmblm estimator had access to 60 compounded data points.

Specifically, the steps for each of 1 million draws were as follows: (1) daily value-weighted market returns are sampled after 1974. (2) A (presumed true) market-beta and market-model sigma are drawn, either from the empirical (winsorized) distribution or from a presumed linear relationship between market-model beta and market-model sigma. (3) Unit-normalized daily market-model residuals are sampled and rescaled to the drawn sigma. (4) The stock returns are calculated from the market-beta, market returns, and market-model residuals. (5) All generating information except returns are hidden from the estimators. The daily analysis receives 252 data points, the monthly analysis receives 60 data points (of compounded rates of return).

[Insert Table 2 here: **Idealized Properties of Market-Beta Estimators' Errors**]

Estimation: For each draw, we then estimate betas with various techniques, and compare them to the (hidden but known) true beta that underlay the random draw. Table 2 tabulates the performance results. It shows that monthly bmblm estimates are nearly unbiased, but have high RMSEs of 0.32 to 0.33. Both the plain OLS and Vasicek estimator 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, respectively. This suggests a relative improvement by the bsw estimator over bmblm of about 40-45%, and about 5-10% over the bols and bvck estimators.

V Age-Decayed Estimators

It has long been known that the underlying beta drifts. The common way to handle this drift is to use a block-samplers (Ghysels and Jacquier (2007)), i.e., a moving estimation window. All daily estimators discussed so far have been implemented as 1-year block samplers; the monthly bmblm as a 5-year block sampler.

A smoother alternative is to progressively disregard older observations. A weighted-least-squares regression (WLS) can offer smooth in-time age decay.

Block samplers require the window length as a parameter, while age decay requires the speed of decay as a parameter. My age-decayed slope-winsorized beta estimator has a constant decay of $2/256 \approx 0.78\%$ per day and is named bswa.[17] To keep its implementation simple, bswa uses only one fixed parameter, regardless of firm and time. Thus, like bsw, users can calculate bswa without the need for a first-stage regression. Moreover, unlike block samplers, age-decayed bswa can easily be updated without needing estimation restarts.

Empirically, bswa offers similar beta forecasts when each day receives a weight anywhere between about 0.7% to 0.9% higher than the preceding one. This implies half-lives somewhere between 75 and 120 days, with 90 days a good middle, $(1/(1+2/256))^{90} \approx 0.5$.

The parameter choice 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.

Estimators other than bsw could be similarly modified to reflect age decay. However, I know of no papers in the literature that have done so. In the absence of a compelling reason to enhance other estimators, age decay is used only to understand the viable improvement of the best block-sampling estimator, which is the bsw estimator. By showing the performance of both bsw and bswa, it is possible for the reader to understand the improvement due to winsorization vs. due to aging.

[17] The particular decay was chosen for its ability to express it as an integer divided by 256.

VI Implementation of BSWA

The complete R code is

```
## _bswb is an internal function doing most of the work
_bswb <- function( ri, rm, Delta, rho ) {
   stopifnot( (length(ri) > 250) & (length(ri) < 255*3 ) )

   ## 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 )

   ## a utility function
   olsbeta <- function(...) coef(lm(...))[2]

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

## 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 )

# d <- read.csv("dailyreturns.csv"); print(bsw(d$ri, d$rm))
```

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

VII Distinctiveness of Estimates

[Insert Table 3 here: **Descriptive Statistics, 1973-2019**]

Table 3 shows the data ranges, means, and standard deviations of the independent variables (beta predictors) used in my paper. 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 0.1 for the others. For the 3,000 biggest stocks, the bias can be more pronounced. It is particularly stark for bmm, although this is due to their more aggressive provision of smaller stocks. Beta estimates from different methods also differ in their standard deviations. bfp and bcfsb are least heterogeneous, while OLS estimates (both daily and monthly) and bakx are most heterogeneous.

[Insert Table 4 here: **Distinctness of Estimators, 1973-2019**]

This raises the question of whether the betas are distinct enough to make a difference. Table 4 shows the root-mean-squared differences across estimators in pairwise comparisons, thereby assuring that the set of firms is kept the same. Because all estimators are attempting to isolate the same underlying true beta, it is not surprising that they are somewhat similar. The two slope-winsorized estimators differ by less than 0.1 for the biggest 1,000 firms. The block-sampled bsw is also very similar to the block-sampled bvck. (As I will show below, much of the power of bvck derives from its effect on outliers.) The robust Martin and Simin (2003) bmm estimator is similar to bols, bvck, and bsw. The distances of the four basic estimators to other estimators (bmols, bmvck, bfp, bakx, and sometimes bcfsb) are larger, typically ranging from about 0.3 to 0, even for the biggest 1,000 firms. In particular, the estimators based on monthly stock-market rates of returns seem entirely different from their daily siblings, with RMSD's of 0.2 to 0.7.

The differences among estimators seem sensible. They are small but not trivial. For a CAPM use attempt 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 is about 60 to 200 basis points per annum—reasonably modest for estimators based on the same stock prices and seeking to estimate the same underlying true beta, but still economically meaningful.

VIII Forecasting Regression Test Design

A Performance Metrics

Like earlier papers, starting with Rosenberg and Guy (1976a), 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 *one-year-ahead* realized OLS beta, $bols_{i,y+1}$.

The first metric comes from predictive "gamma" regressions,

$$R^2\left(bols_{i,y+1} = \gamma_0 + \gamma_1 \cdot \hat{b}_{i,y} + \epsilon_{i,y+1} \right) . \tag{4}$$

The second metric is the direct RMSE

$$\sqrt{\sum_{i,y} (bols_{i,y+1} - \hat{b}_{i,y})/N} . \tag{5}$$

The regression R^2 metric is not affected by *bias* in the beta estimator, while the RMSE metric is. When beta is merely a control variable in a regression, the bias is harmless and methods with higher R^2 are better. When beta is used directly as a hedge ratio, methods with lower RMSE are better.

The tables below will 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 about a thousand observations in the cross-section and there are less 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 or vice-versa are used, and when forecasts compete in multivariate (rather than univariate) gamma regressions (Appendix $\mathscr{A}$).

B Errors In The Dependent Variable

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, tests suffer from noise (e.g., lower R-squareds in the "gamma" regressions, which predict future realized betas with current beta estimates).

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

C Errors In The Independent Variable

It is more problematic that the independent variable is also measured with noise. If the underlying model is stable, then the asymptotic bias in a slope coefficient of past on future values is $1/(1+\sigma_{\mathrm{e}}^2/\sigma_{\mathrm{b}}^2)$, 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 $1-1/(1+0.05/0.4^2) \approx 2\%$.[18] The 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.

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

D Hypothetically Achievable Prediction Accuracy

If both the dependent and the independent variable 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 Jegadeeshash et al. (2019)). For example, if the historical OLS beta, bols, can explain 50% of its future self, then it would suggest that bols could explain about $\sqrt{.50} \approx 71\%$ of the true unknown beta.[19]

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.

IX The Empirical Performance in a Horse Race

OOS tests can catch issues that were not preserved in the analysis of the estimators under the idealized circumstances in Section IV. Nevertheless, the tests that follow remain *specific*. They are relevant to the common need to forecast the performance of beta estimators for OLS (and non-OLS) market betas over the next 1 to 24 months. Other beta estimators could have advantages in contexts not considered in my paper. For example, an intra-day estimator could work better forecasting the one-day ahead beta.[20]

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

[20]However, it is also not a-priori clear why this would be the case. For example, the 1-year bols can predict a 1-month-ahead bols (based on about 20 trading days worth of stock returns) with an R^2 of about 17.5% for the largest 3,000 stocks. The 1-month bols predicts the same dependent variable (itself leading by 1 month) only with an R^2 of about 8%. If there is an advantage to using a beta predictor that is based on the same window length as the predicted beta, it is outweighed by the loss of precision in the estimator (not estimated) from the shorter window. Using the longer-window beta estimator is better even for predicting 1-month ahead betas.

A Predicting the OLS Beta With the Basic Daily Estimators

[Insert Table 5 here: **Predicting One-Year Ahead OLS Market-Beta (bols), 1973-2019**]

Table 5 shows the key result of the paper. It describes the univariate characteristics and the forecasting performances of the basic four estimators using daily stock-return data: the OLS estimator (bols), the Vasicek estimator (bvck), the slope-winsorized estimator (bsw), and the aged slope-winsorized estimator (bswa). All estimators use concurrent historical data, based on the same daily-frequency stock returns, and thus can be computed for the same stocks and months. Panel A focuses on the biggest 1,000 stocks using non-overlapping calendar year regressions, Panel B on the biggest 3,000 stock, using individual-month overlapping regressions.

The two left columns show the means and standard deviations of the independent variables. The columns to the right show the forecasting performance predicting the future bols—in particular, the forecasting regression coefficients γ_0 and γ_b, the forecasting R^2, and the direct RMSE.

The performance ordering is always the same. The historical bols performs worse than the bvck estimator, which performs worse than the two slope-winsorized estimators. The forecasting improvements are solid but modest and in-line with those suggested by Section IV and VIII. They roughly double the improvement provided by bvck over bols. With their ease of implementation, there is little reason not to use the better slope-winsorized estimators.

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

Figure 2 plots the relative performance of the estimators, with the bols estimator normalized to be 0.0. 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. Both plots show 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 principally in a small set of bunched years.

In sum, the figure shows that the superior performance of the two bsw estimators is not a fluke, but has occurred in (separate) samples for many years. It has also not diminished over time. Appendix $\mathscr{C}$ shows that it also appears in different subsets created based on market-beta or its standard error (itself a proxy for stock return volatility).

B Adding Beta Estimators from the Literature

[Insert Table 6 here: **Competitive Prediction, 1973-2019**]

Table 6 allows comparing the predictive performance of the four main estimators to other beta estimators. Each panel is careful to compare predictions for the same dependent variable (here, bols) in the same set of observations. Otherwise, comparisons would be meaningless.

Panel A shows that beta estimates based on *monthly* stock returns perform poorly predicting the future bols. Monthly betas should never be used to predict bols. Even the most common bmblm performs poorly, although it surprisingly performs better than the bmvck estimator on monthly stock returns. (Not shown, bmvck performs better among microcaps.) The RMSE of bswa is about 70% that of bmblm, again in line with what was suggested by Sections IV and VIII.

Panel B shows that Frazzini and Pedersen (2014) betas perform better than the monthly market betas, but remain inferior even to the bols estimator. bfp performs especially poorly on the R^2 metric, suggesting that its best quality is a lesser bias.

Panel C shows that the Martin and Simin (2003) betas perform well in both data sets. bmm performs as well as or better than bols, bvck, and bsw on the R^2 metric, and better than bols on the RMSE metric. Yet it cannot beat bswa on either metric or in either data set. However, as already mentioned, bmm's biggest drawback is not its performance but the fact that its computation requires access to a supercomputer and is thus unlikely to find wide use.

Panel D shows that the Ait-Sahalia, Kalnina, and Xiu (2014) betas perform poorly. They underperform on both metrics. Not shown, this was also the case for a 12-month moving average of bakx.

Panel E shows that the Cosemans et al. (2016) betas cannot outperform the basic four estimators on the R^2 metric in either data set. However, bcfsb performs better on the RMSE metric. The estimates are nearly unbiased and can therefore outperform both bols and bvck for the biggest 1,000 firms and bols for the biggest 3,000 firms. Yet, they cannot outperform bsw and bswa—and they are again very difficult to recreate.

C Predicting Self Rather than the OLS Beta

As already explained, practitioners usually care about *future* betas, not historical betas. Ergo, if another estimator $\hat{b}$ cannot predict its future self as well as bswa can predict $\hat{b}$, then even a user interested in this other estimator $\hat{b}$ today may want to use bswa instead of the $\hat{b}$ itself. Good self-prediction is a necessary but not a sufficient criterion. *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.[21] Estimators should instead be evaluated primarily based on their ability to predict future naive realized betas (i.e., bols).

In this subsection, we predict future beta estimates that are other than those obtained from bols. To the extent that all beta estimators attempt to uncover the same true beta signal in noisy stock return data, if the estimation error is iid, the best estimator (bswa) could predict not only the bols estimates better but also other (noisy) estimates of beta.

[21]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 Betas Other Than bols, 1973-2019**]

Segment (A) in Table 7 shows that bswa dominates bsw. Even a user interested in forecasting bsw should use bswa rather than bsw. Together with the evidence in Table 5 that bsw cannot predict bols better than bswa, it suggests that bsw should not be used.

Segment (B), together with Table 6, shows that bvck is similarly dominated—and not just by the return-aging bswa estimator, but also by the bsw block sampler.

Segment (C) shows that bmblm can predict itself better than bsw and bswa can predict it on on the RMSE metric, but not on the R^2 metric. A bias-matched bswa estimator would obsolete bmblm even in applications that are interested in the future bmblm per se.

Segments (D-E) show that the two other monthly-return beta estimators are also better forecast by bsw and bswa. Thus, even if bmols is the quantity of interest, the user is better off working with historical bswa data than historical bmols data.

Segment (E) shows that bfp cannot predict itself better than bswa can predict bfp on the R^2 metric. However, bfp is biased enough that it has a lower RMSE in predicting its future self than bswa. It is a different animal—a user interested in predicting its future realization must either first rebias bswa or stick to bfp.

Segment (F) shows that bmm can predict itself better than bsw can predict bmm on R^2 but not RMSE. As a direct measure (RMSE), both bsw and bswa are better predictors than bmm is for itself.

Segment (G) shows that bakx cannot predict itself better than the bswa estimates can predict bakx. For predicting future bakx estimates, bswa is preferable. Together with the evidence than bakx is also a poor predictor of bols, it should not be used in our context. This inference extends to a 12-month moving average of bakx.

Note that we do not predict bswa and bcfsb, because both are not block-sampled but updating estimators. Good self-predicting power is mechanically built in.

X Conclusion

Extensive appendices answer many questions not covered in this paper. They all point to the same conclusion that a slope-winsorized beta estimator estimated from daily returns predicts future stock market betas better than other known estimators. When age-decayed with a half-life of about 4-5 months, the resulting bswa performs even better. The most common beta estimator is also the worst one and by a wide margin. The Bloomberg-Merrill Lynch beta (also on many websites), based on a 5-year block-sampler using monthly rates of return, should never be used.

Both bsw and bswa are easy to implement and use. The complete R code is 10 lines and appears in Section VI. Moreover, WRDS now offers a complete set of slope-winsorized betas for download.

Using a computer software analogy, bvck based on daily stock returns (or bmm or bcfsb) seem good enough not to require an upgrade. However, the upgrade is free, it performs better, and it is easier to use. In contrast, the forecasting performances of many other beta estimators is so much worse that an upgrade would seem called for.[22]

[22] In their defense, the use of inferior betas can be harmless—mostly in situations in which one may as well use 1.0 as the universal beta estimate and not even bother with estimation—such as when one needs estimates of betas for large well-diversified portfolios. In the extreme, a portfolio similar enough to the value-weighted market portfolio has a beta of 1.0 by definition.

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.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang (2009). “High idiosyncratic volatility and low returns: International and further US evidence”. In: *Journal of Financial Economics* 91.1, pp. 1–23.

Back, Kerry, Nishad Kapadia, and Barbara Ostdiek (2015). *Testing Factor Models on Characteristic and Covariance Pure Plays*. Tech. rep.

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

Brown, Stephen J and Jerold B Warner (1985). “Using daily stock returns: The case of event studies”. In: *Journal of financial economics* 14.1, pp. 3–31.

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.

Engle, Robert F. (1982). “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation”. In: *Econometrica* 50, pp. 987–1007.

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

Fowler, David J and C Harvey Rorke (1983). “Risk measurement when shares are subject to infrequent trading: Comment”. In: *Journal of Financial Economics* 12.2, pp. 279–283.

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.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle (Dec. 1993). "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks". In: *The Journal of Finance* 48.5, pp. 1779–1801.

Han, Xing (2019). *Understanding the Performance of Components in Betting Against Beta*. Tech. rep. University of Otago, NZ.

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

Harvey, Campbell R., Yan Liu, and Heqing Zhu (Jan. 2016). "...and the Cross-Section of Expected Returns". In: *Review of Financial Studies* 29 (1), pp. 5–68.

Hoberg, Gerard and Ivo Welch (Dec. 2009). *Optimized vs. Sort-Based Portfolios*. Tech. rep. USC and UCLA.

Huber, Peter J. (1964). "Robust Estimation of a Location Parameter". In: *Annals of Statistics* 53.1, pp. 73–101.

Jegadeeshash, 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.

Karolyi, G. Andrew (1992). "Predicting Risk: Some New Generalizations". In: *Management Science* 38, pp. 57–74.

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.

Lo, Andrew W and A Craig MacKinlay (1990). "Data-snooping biases in tests of financial asset pricing models". In: *The Review of Financial Studies* 3.3, pp. 431–467.

Martin, Ian and Stefan Nagel (Dec. 2020). *Market Efficiency in the Age of Big Data*. Tech. rep. LSE and Booth.

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

Mclean, R. David and Jeffrey Pontiff (Feb. 2016). "Does Academic Research Destroy Stock Return Predictability?" In: *The Journal of Finance* 71.1, pp. 5–32.

Novy-Marx, Robert and Mikhail Velikov (Nov. 2018). *Betting Against Betting Against Beta*. Tech. rep. University of Rochester and FRB of Richmond.

Rosenberg, Barr and James Guy (May 1976a). “Prediction of Beta from Investment Fundamentals (Part One)”. In: *Financial Analysts Journal*, pp. 60–72.

— (July 1976b). “Prediction of Beta from Investment Fundamentals (Part Two)”. In: *Financial Analysts Journal*, pp. 62–70.

Rousseeuw, Peter J. (1984). “Least Median of Squares Regression”. In: *Journal of the American Statistical Association* 79.388, pp. 871–881.

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

Tukey, John W. (1960). “A survey of sampling from contaminated distributions”. In: *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Ed. by I. Olkin et al. Stanford University Press, pp. 448–485.

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.

Table 1: Glossary of Market-Beta Estimators

12 Months of Daily Stock Return Data

Abbrev	Long Name	Remarks
bols	OLS	Base OLS Estimator, Daily Stock Returns.
bvck	Vasicek	Random-Effects Panel Estimator. Vasicek (1973).
bsw	Slope-Winsorized	Firm returns first winsorized at $-2 \cdot r_{m,d}$ and $+4 \cdot r_{m,d}$.
bmm	Robust Maximum-Likelihood	Martin and Simin (2003), Rousseeuw (1984).

Partly or Fully Intra-Day Returns

Abbrev	Long Name	Remarks
bcfsb	Prior Firm Information	Intra-day and Fundamentals, Cosemans et al. (2016).
bakx	Intra-Day (1 Month!)	Intra-day, Ait-Sahalia, Kalnina, and Xiu (2014).

Extended (60 Months) of Stock Return Data

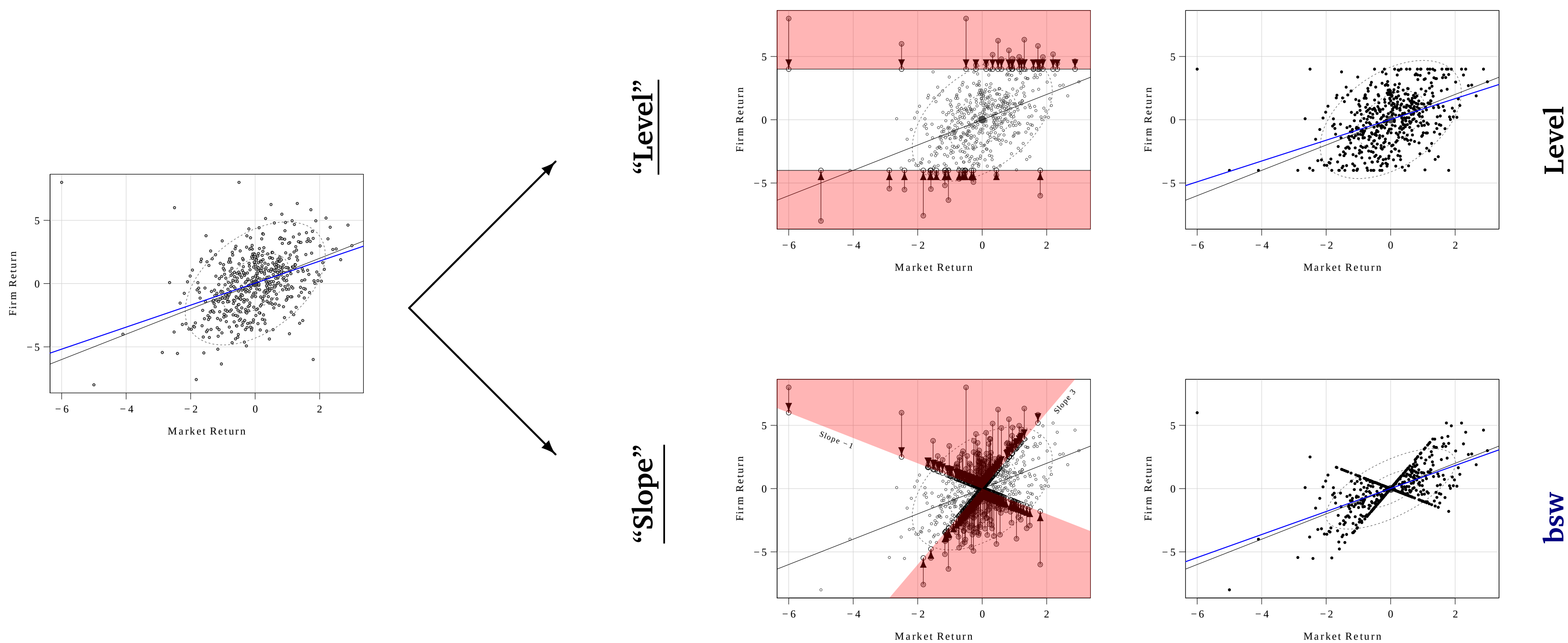
Abbrev	Long Name	Remarks
bmblm	Bloomberg/ML/CIQ/Web	calculated from 60 monthly stock returns.
bmols	Monthly OLS	same
bmvck	Monthly VCK	same
bfp	Frazzini-Pedersen	Daily-frequency returns, 12 months for variances, 60 months for correlation, Frazzini and Pedersen (2014).

Daily Stock Return Data As Long As Available

Abbrev	Long Name	Remarks
bswa	Slope-Winsorized and Aged	Like bsw but $(1 + 2/252) - 1 \approx 0.8\%$ decay per day.

Explanations: bols, bvck, bsw, bswa, bmblm, bmols, bmvck, and bfp were computed by myself. All stock returns and the value-weighted market return were net of the prevailing risk-free rate from Ken French's website. bmm, bakx, and bcfsb were provided by their respective authors. The bfp betas were checked against those estimated in Novy-Marx and Velikov (2018).

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 based on a beta prior. The right column shows the winsorized returns and the new OLS estimated line. (Appendix $\mathscr{E}$ shows band winsorization.)

Interpretation: Level winsorization flattens the estimated slope. Band winsorization is harshest on positive return outliers when the market was bearish and on negative return outliers when the market was bullish. 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 2: Idealized Properties of Market-Beta Estimators' Errors

Estimator	Return Inputs	"Normal Scenario" Mean	SD	RMSE	"Empirical Scenario" Mean	SD	RMSE
bmblm	1,260 Days	0.002	0.331	0.331	0.002	0.320	0.320
bols	252 Days	–0.000	0.188	0.188	–0.000	0.184	0.184
bvck	same	–0.016	0.199	0.200	–0.016	0.183	0.184
bsw	same	–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 have access only to 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 ($2/3 \cdot$ bols $+ 1/3 \cdot 1$) using compounded monthly returns, bols is the standard OLS estimator, bvck is the Vasicek 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 outperforms the alternatives.

Table 3: Descriptive Statistics, 1973-2019

	From To	1,000 Stocks N	Mean	SD	3,000 Stocks N	Mean	SD
bswa	197301 201812	529,951	0.98	0.42	1,562,546	0.91	0.46
bsw			0.98	0.43		0.92	0.47
bols			1.02	0.51		0.96	0.60
bvck			0.99	0.46		0.93	0.52
bmblm	197308 201812	451,845	1.05	0.37	1,856,824	1.05	0.49
bmols			1.08	0.55		1.08	0.73
bmvck			1.06	0.46		1.04	0.50
bfp	197301 201812	486,701	1.07	0.30	2,403,123	1.00	0.39
bmm	197301 201806	480,840	0.96	0.48	2,255,641	0.70	0.60
bakx	199601 201712	248,986	1.01	0.59	1,078,196	0.90	1.06
bcfsb	199601 201112	15,660	0.98	0.34	16,228	0.99	0.34

Explanations: Simple univariate statistics for the variables. Variables are explained in Table 1. Not shown but inferrable, the attrition rate is about 0.5% for the biggest 1,000 stocks, and about 5% for the largest 3,000 stocks.

Interpretation: Estimators vary in terms of bias and standard deviation.

Source: benchmarks/1univariate.Rout

Table 4: Distinctness of Estimators, 1973-2019

Panel A: 1,000 Biggest Firms

	bswa	bsw	bols	bvck	bmblm	bmols	bmvck	bfp	bmm	bakx	bcfsb
bswa	0	0.08	0.17	0.12	0.34	0.44	0.38	0.26	0.14	0.39	0.19
bsw	0.08	0	0.14	0.08	0.35	0.45	0.39	0.27	0.10	0.41	0.21
bols	0.17	0.14	0	0.09	0.38	0.46	0.42	0.31	0.10	0.42	0.24
bvck	0.12	0.08	0.09	0	0.37	0.46	0.40	0.29	0.09	0.41	0.23
bmblm	0.34	0.35	0.38	0.37	0	0.19	0.12	0.27	0.39	0.48	0.26

Panel B: 3,000 Biggest Firms

	bswa	bsw	bols	bvck	bmblm	bmols	bmvck	bfp	bmm	bakx	bcfsb
bswa	0	0.10	0.25	0.16	0.45	0.59	0.46	0.33	0.18	0.62	0.19
bsw	0.10	0	0.22	0.11	0.46	0.60	0.47	0.34	0.15	0.64	0.21
bols	0.25	0.22	0	0.17	0.50	0.61	0.51	0.40	0.15	0.65	0.26
bvck	0.16	0.11	0.17	0	0.48	0.61	0.49	0.37	0.14	0.64	0.24
bmblm	0.45	0.46	0.50	0.48	0	0.22	0.14	0.35	0.52	0.71	0.26

Explanations: These are the root mean square distances between beta estimators in their pairwise overlapping samples.

Interpretation: The four basic daily stock-return based beta estimators and bmm tend to have typical mutual distances between 0.1 and 0.2 for the largest 1,000 stocks, and about 50% higher for the largest 3,000 stocks. Daily estimators have distances between 0.3 and 0.4 with estimators that are based on other frequencies, either monthly stock returns or intra-day stock returns.

Source: benchmarks: 1compare.Rout, May 16, 2020

Table 5: Predicting One-Year Ahead OLS Market-Beta (bols), 1973-2019

Panel A: 1,000 Biggest Stocks, Calendar-Year Forecasting Regressions

N	Mean	SD	Predictor	γ_0	γ_b	$\bar{R}^2$	RMSE
44,190	1.02	0.52	bols	0.29	0.71	0.5286	0.3815
44,190	1.00	0.47	bvck	0.22	0.79	0.5305	0.3639
44,190	0.99	0.43	bsw	0.15	0.87	0.5385	0.3529
44,190	0.99	0.42	bswa	0.11	0.92	**0.5642**	**0.3405**

Panel B: 3,000 Biggest Stocks, Overlapping Forecasting Regressions

N	Mean	SD	Predictor	γ_0	γ_b	$\bar{R}^2$	RMSE
1,562,546	0.96	0.60	bols	0.36	0.63	0.4265	0.4917
1,562,546	0.93	0.52	bvck	0.26	0.76	0.4568	0.4453
1,562,546	0.92	0.47	bsw	0.19	0.84	0.4636	0.4322
1,562,546	0.91	0.46	bswa	0.16	0.88	**0.4872**	**0.4201**

Explanations: The dependent variable is always the 1-year ahead OLS market-beta (bols), calculated from daily stock returns. The estimators are defined in Section II (and listed in Table 1). bols, bvck, and bsw are 1-year block samplers. The means and standard deviations in the left two columns refer to the independent variable. The sample is defined in Section I. To be included, a future one-year beta (bols) had to be available. (The sample thus begins in 1973/1974.) The 1,000 or 3,000 stocks were ranked by marketcap contemporaneous with the independent beta estimate.

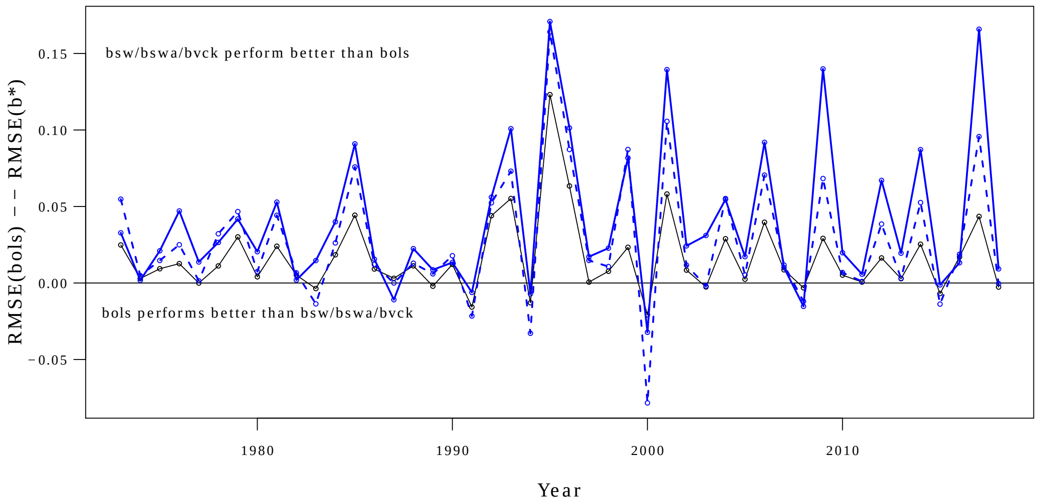
The reported gamma coefficients and $\bar{R}^2$s are from pooled predictive regressions. (Not shown, the typical standard errors for γ_b were below 0.002, i.e., below the displayed two-digit coefficient precision.) The best performer in each panel is bold-faced. The root-mean-squared-errors (RMSEs) are the square-roots of the mean distances between the future bols realizations and the predicting beta estimates. Panel B uses all months (incl. overlapping ones), Panel A uses only calendar-year estimators.

Interpretation: The slope-winsorized market-betas (bsw, bswa) outperform both the OLS and Vasicek estimates.

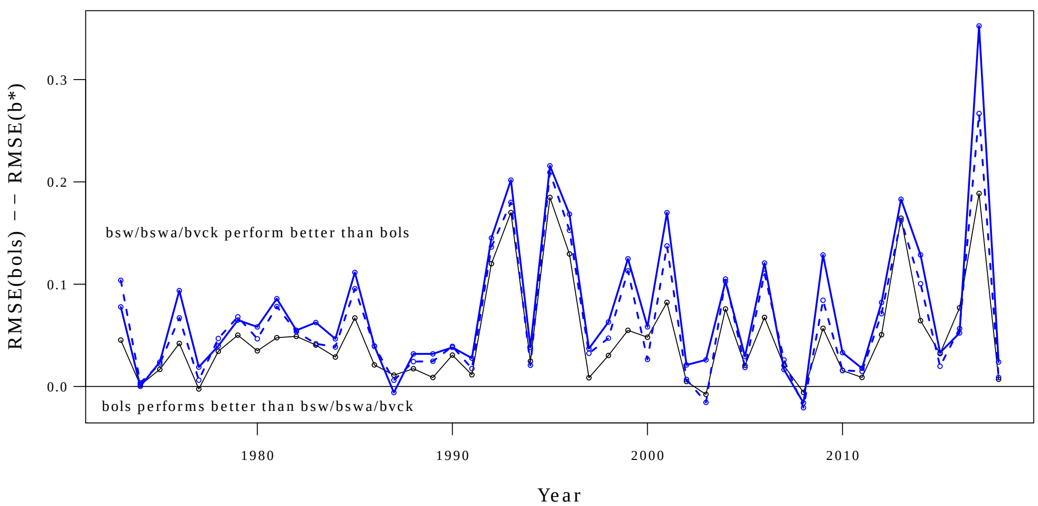
Source: 2analysis.Rout, May 16, 2020.

Figure 2: Year-by-Year Relative Predictive Performance of Estimators, Dec 1973-2019

Panel A: Biggest 1,000 Stocks



Panel B: Biggest 3,000 Stocks



Explanations: This plot is equivalent to Table 5, but shows the differences in the RMSE's of the predictive regressions when run year by year (Decembers only). The zero line is the OLS predictive performance. The solid blue line is for the bswa estimator, the dashed blue line is for the bsw estimator, the black line is for the bvck estimator. Higher values imply better performance.

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

Table 6: Competitive Prediction, 1973-2019

Panel A: Monthly Stock-Return Based Betas: (1973 to 2018, N is 451,845 and 1,121,845)

	Biggest 1,000 Firms		Biggest 3,000 Firms	
Predictor	$\bar{R}^2$	rmse	$\bar{R}^2$	rmse
bols	0.5546	0.3405	0.4802	0.4312
bvck	0.5547	0.3294	0.4908	0.4087
bsw	0.5585	0.3202	0.4984	0.3947
bswa	**0.5849**	**0.3081**	**0.5245**	**0.3815**
bmblm	0.2906	0.4190	0.2059	0.5387
bmols	same	0.5059	same	0.6587
bmvck	0.2851	0.4565	0.2132	0.5566

Panel B: Frazzini and Pedersen (2014): bfp (1973 to 2018, N is 486,701 and 1,288,271))

	Biggest 1,000 Firms		Biggest 3,000 Firms	
Predictor	$\bar{R}^2$	rmse	$\bar{R}^2$	rmse
bols	0.5573	0.3477	0.4743	0.4440
bvck	0.5564	0.3365	0.4856	0.4198
bsw	0.5596	0.3276	0.4937	0.4056
bswa	**0.5859**	**0.3155**	**0.5198**	**0.3924**
bfp	0.4252	0.3746	0.3278	0.4685

Panel C: Martin and Simin (2003): bmm (1973 to 2018, N is 480,840 and 1,265,066)

	Biggest 1,000 Firms		Biggest 3,000 Firms	
Predictor	$\bar{R}^2$	rmse	$\bar{R}^2$	rmse
bols	0.5576	0.3481	0.4781	0.4420
bvck	0.5566	0.3368	0.4878	0.4189
bsw	0.5595	0.3279	0.4958	0.4045
bswa	**0.5855**	**0.3159**	**0.5216**	**0.3914**
bmm	0.5698	0.3397	0.4959	0.4328

(pto)

Panel D: Ait-Sahalia, Kalnina, and Xiu (2014): bakx (1996 to 2017, N is 248,986 and 717,101)

Predictor	Biggest 1,000 Firms $\bar{R}^2$	Biggest 1,000 Firms rmse	Biggest 3,000 Firms $\bar{R}^2$	Biggest 3,000 Firms rmse
bols	0.5197	0.3802	0.3656	0.5061
bvck	0.5140	0.3675	0.3875	0.4634
bsw	0.5220	0.3569	0.3946	0.4493
bswa	**0.5531**	**0.3432**	**0.4239**	**0.4346**
bakx	0.4182	0.4708	0.2270	0.7141

Panel E: Cosemans et al. (2016): bcfsb (1996 to 2011, N is 15,660 and 16,008)

Predictor	Biggest 1,000 Firms $\bar{R}^2$	Biggest 1,000 Firms rmse	Biggest 3,000 Firms $\bar{R}^2$	Biggest 3,000 Firms rmse
bols	0.5509	0.3258	0.5542	0.3351
bvck	0.5571	0.3156	0.5602	0.3249
bsw	0.5607	0.3074	0.5615	0.3179
bswa	**0.5862**	**0.2956**	**0.5847**	**0.3068**
bcfsb	0.5285	0.3105	0.5180	0.3259

Explanations: This table is equivalent to Table 5, but includes other beta estimators. To be a valid horse race, the included samples and dependent variables had to be identical, necessitating multiple panels. The best predictor in each panel is boldfaced.

Interpretation: bols, bmvck, bfp, and bakx underperform even bols estimates when predicting future bols in both samples and on both metrics. bmm and bcfsb can outperform bswa, but can at least in beat bols in one sample on one metric.

Source: benchmarks/1analysis.Rout, May 16, 2020

Table 7: Predicting Future Betas Other Than bols, 1973-2019

	Prediction		Biggest 1,000 Firms		Biggest 3,000 Firms		
	Y	X	$\bar{R}^2$	rmse	$\bar{R}^2$	rmse	X Range
(A)	bsw	self (12mo)	0.5766	0.2953	0.5187	0.3490	1973-2017
	(12mo)	bswa	**0.6025**	**0.2820**	**0.5428**	**0.3346**	
(B)	bvck	self (12mo)	0.5526	0.3328	0.4875	0.4010	1973-2017
	(12mo)	bsw	0.5597	0.3184	0.4957	0.3815	
		bswa	**0.5855**	**0.3057**	**0.5192**	**0.3682**	
(C)	bmblm	self (60mo)	0.2375	**0.3712**	0.1920	**0.4460**	1973-2014
	(60mo)	bsw	0.2728	0.3917	0.2263	0.4701	
		bswa	**0.2844**	0.3826	**0.2363**	0.4600	
(D)	bmols	self (60mo)	0.2375	0.5568	0.1920	0.6691	1973-2014
	(60mo)	bsw	0.2728	0.4916	0.2263	0.5929	
		bswa	**0.2844**	**0.4843**	**0.2363**	**0.5853**	
(E)	bmvck	self (60mo)	0.2510	0.4636	0.2245	0.5003	1973-2014
	(60mo)	bsw	0.2872	0.4310	0.2549	0.4872	
		bswa	**0.2991**	**0.4229**	**0.2659**	**0.4778**	
(F)	bfp	self (60mo)	0.1967	**0.3266**	0.1771	**0.3709**	1973-2014
	(60mo)	bsw	0.2736	0.3859	0.2603	0.4401	
		bswa	**0.2851**	0.3766	**0.2725**	0.4285	
(G)	bmm	self (12mo)	0.6051	0.3166	0.5495	0.3944	1973-2017
	(12mo)	bsw	0.5898	0.3087	0.5390	0.3767	
		bswa	**0.6161**	**0.2962**	**0.5644**	**0.3633**	
(H)	bakx	self (12mo)	0.2845	0.5646	0.0950	0.8879	1996-2016
	(12mo)	bsw	0.3156	0.4907	0.1514	0.7146	
		bswa	**0.3302**	**0.4831**	**0.1594**	**0.7075**	

(pto)

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

Interpretation: For an estimator to be useful to a researcher interested in future rather than contemporaneous market betas, it is a necessary but not a sufficient that this estimator can predict its future self better than alternatives. By this criterion, bmols, bmvck, or bakx are not useful. (With a caveat, neither is bmblm.) The contemporary bswa is a superior replacement. For bmm, this is also the case for the biggest 3,000 stocks but not the biggest 1,000 stocks. In contrast, bfp is sufficiently idiosyncratic (perhaps not underlying beta- but variance-related) that bswa cannot fully replace it. bcfsb can always predict itself better than bswa.

Source: benchmarks/1analysis.R, May 16, 2020

APPENDIX

These are just some appendixes describing further data analyses. (Not shown, for example, is that the slope-winsorized beta estimators also outperform in the full CRSP universe and not just the biggest 1,000 or 3,000 stocks; and that their performances are quite robust with respect to the age and decay parameters.) Others are available upon request.

$\mathscr{A}$ Multivariate Regressions

[Insert Table $\mathscr{A}$.1 here: **Multivariate Predictive Regressions, 1973-2019**]

Table $\mathscr{A}$.1 enters multiple beta estimators into a competitive multivariate regression. For the biggest 1,000 stocks, the Vasicek estimator is completely subsumed by either slope-winsorized estimator. For the biggest 3,000 stocks, the Vasicek estimator can still add modestly to the prediction above and beyond bsw and bswa.

These multivariate results should not be overread. All estimators seek to measure the same true beta, are based on the same stock returns, and are highly correlated. The coefficient estimates are highly multicollinear and not robust to small changes. For the same reason, the additional $\bar{R}^2$ obtained by the two extra variables is modest, as indicated by the table.

Table $\mathscr{A}$.1: Multivariate Predictive Regressions, 1973-2019

tbl:mvregs in beta-appendix.tex, August 17, 2020

Biggest	bX	γ_0	γ_{bols}	γ_{bvck}	γ_{bx}	$\bar{R}^2$	N
1,000 stocks	bsw	0.20	0.42	–0.16	0.55	54.81%	529,951
	bswa	0.15	0.40	–0.35	0.83	57.24%	529,951
	bswa	0.11			0.92	56.42%	529,951
3,000 stocks	bsw	0.20	–0.05	0.31	0.57	46.61%	1,562,546
	bswa	0.16	–0.07	0.19	0.76	48.82%	1,562,546

Explanations: This table is equivalent to Table 5, but performs multivariate instead of bivariate regressions.

Interpretation: The slope-winsorized betas subsume all explanatory power from the Vasicek betas within the biggest 1,000 firms. They are more important than but do not subsume all power of bvck within the biggest 3,000 firms.

Source: work/1analysis.Rmd, May 16, 2020

B Heterogeneous Parameters?

Both the winsorizing and time-decaying parameters are fixed in advance. They are not fragile and were (mostly) obtained from estimations prior to 1974. Similarly, they could also be obtained from subset of stocks not included in a specific performance measurement, as shown below. Even if they had been estimated together with betas *in time*, they would have consumed only two degrees of freedom in analyses based on 75 million return observations.

As in Martin and Simin (2003), the parameters could also have been determined from conditional data analysis and/or (in-sample) first-stage estimations. Or they could be different for the upper or lower bounds. However, the spirit of a specification-based[23] estimator is to keep the parameters as simple as possible. No first-stage regressions are used.

The empirical evidence suggests that such changes are unlikely to improve the estimation greatly. I explored whether the best unknowable *ex-post* winsorization and decay parameters within various subgroups (e.g., within years, marketcap, etc.) could improve the forecasts. Although this was by necessity so, the forecasting gains were quite modest.

These ex-post optimal parameters are not even easily forecastable. Tukey (1960)'s point that it is less important which robust method is used, just that some robust method is used, seems to hold here, too. In sum, fine-tuning of the two parameters is not likely to be a fruitful avenue of pursuit.

[23]The "specification-based" moniker is intended to suggest that bsw is not based on a multi-step model-parameter-based estimation, but that it is a much simpler estimator based just on a pre-specified prior.

𝒞 Subset Performance

The superior performance of bsw and bswa holds by and large in subsets based on obvious variable classifications. It does not seem specific to a particular set of firms.[24]

Year by Year: Figure 2 plotted the year-by-year RMSE performance of the four estimators.

By Market Cap: Table 𝒞.1 divides stocks into three tertiales based on marketcap. The performance ordering (bswa, bsw, bvck, and bols) remains as before, with the exception of the (approximately) 150 biggest stocks, where bvck and bsw mildly underperform bols on the R^2 metric. Simply put, the largest 150 firms have not enough outliers to allow other estimators to gain an advantage over bols on R^2. However, outliers still bias bols, which is why the forecasting advantage of the two estimators returns on the RMSE metric. In contrast, the age-adjusted bswa estimator always improves the prediction and on both metrics, and also when only the biggest tertiale of firms are included.[25]

[Insert Table 𝒞.1 here: **Predictive Regressions By (Log) Marketcap, Dec 1973-2019**]

By OLS Beta: Table 𝒞.2 divides stocks into three tertiales based on the OLS beta. Again, bswa dominates on all metrics and data sets. Again, for the biggest 3,000 stocks, the ordering (bswa, bsw, bvck, and bols) always holds on both metrics. Again, on RMSE, for the biggest 1,000 stocks, the ordering always holds. However, bvck underperforms on R^2 within the biggest 1,000 stocks in the middle and upper bols tertiales, and bsw underperforms in the upper bols tertiales.

[Insert Table 𝒞.2 here: **Predictive Regressions By OLS Market-Beta, Dec 1973-2019**]

By OLS Beta Standard Error: Table 𝒞.3 divides the 1,000 stocks into three tertiales based on the standard error of the OLS beta. (The same results obtain when raw volatility

[24] Thus, each subset could have been used to fix the winsorization and decay parameters for the other subsets, yielding similar inference.

[25] Of course, here this is not due to outlier handling, but better decay handling.

or market-model residual volatility are used instead to split the sample.) Both slope-winsorized alternatives outperform bols and bvck in all three tertiales on both R^2 and RMSE, although the differences are small when betas are very reliable. The RMSE ordering always ranks bswa highest, bsw next, then bvck, and finally bols.

[Insert Table $\mathscr{C}$.3 here: **Predictive Regressions By OLS Market-Beta S.E., Dec 1973-2019**]

Similar results obtain when the classifications are based on dollar trading volume, spreads, and many Compustat-based financial-statement variables. The ordering is almost always the same. The bswa estimator performs best, nearly always followed by bsw and then by bvck, followed by bols.

Further analysis shows that it is possible to use conditioning information to improve the estimators. However, these improvements are very small. The best variable to improve performance would be based on dollar trading volume (or, absent volume, marketcap).[26]

Financial statement information was of surprisingly little practical use. I suspect that earlier recommendations of their use were due to the use of much noisier monthly stock returns in the estimation, rather than daily stock returns.

[26] The beta should be modestly increased for stocks with high volume and modestly decreased for stocks with lower volume.

Table 𝒞.1: Predictive Regressions By (Log) Marketcap, Dec 1973-2019

Panel A: Small Tertiale: Mean Selection Variable is 13.60 for Top-1000.

Univariate			Esti-	1,000 Stocks				3,000 Stocks	
N	Mean	SD	mator	γ_0	γ_b	R^2	RMSE	RMSE	Δ RMSE
14,330	1.01	0.57	bols	0.29	0.70	0.5200	0.4224	0.3145	0.5998
14,330	0.98	0.51	bvck	0.22	0.80	0.5259	0.3980	0.3725	0.5169
14,330	0.96	0.46	bsw	0.15	0.88	0.5378	0.3842	0.3798	0.5008
14,330	0.97	0.45	bswa	0.10	0.92	0.5626	0.3715	0.3989	0.4901

Panel B: Medium Tertiale: Mean Selection Variable is 14.36 for Top-1000.

Univariate			Esti-	1,000 Stocks				3,000 Stocks	
N	Mean	SD	mator	γ_0	γ_b	R^2	RMSE	RMSE	Δ RMSE
14,605	1.02	0.53	bols	0.29	0.71	0.5238	0.3912	0.4409	0.5050
14,605	0.99	0.47	bvck	0.22	0.80	0.5266	0.3725	0.4676	0.4585
14,605	0.98	0.43	bsw	0.15	0.88	0.5345	0.3617	0.4774	0.4445
14,605	0.98	0.42	bswa	0.11	0.92	0.5607	0.3490	0.5015	0.4311

Panel C: Big Tertiale: Mean Selection Variable is 15.79 for Top-1000.

Univariate			Esti-	1,000 Stocks				3,000 Stocks	
N	Mean	SD	mator	γ_0	γ_b	R^2	RMSE	RMSE	Δ RMSE
15,255	1.04	0.46	bols	0.27	0.73	0.5463	0.3279	0.5216	0.3863
15,255	1.03	0.42	bvck	0.23	0.79	0.5405	0.3193	0.5242	0.3675
15,255	1.02	0.39	bsw	0.17	0.85	0.5437	0.3113	0.5332	0.3558
15,255	1.02	0.38	bswa	0.12	0.90	0.5701	0.2993	0.5600	0.3428

Explanations: This table is equivalent to Table 5, but divides firms into tertiales based on log market cap contemporaneous with the predicting market-beta.

Interpretation: The performance ordering is usually the same, regardless of firm set and metric, with one exception: bvck and bsw underperforms bols for the big-firm tertiale among 1,000 stocks on the R^2 metric.

Source: 2analysis.Rout, May 16, 2020.

tbl:key-rsq-mbeta in beta-appendix.tex, August 17, 2020

Table $\mathscr{C}$.2: Predictive Regressions By OLS Market-Beta, Dec 1973-2019

Panel A: Low bols Tertiale: Mean Selection Variable is 0.53 for Top-1000.

Univariate			Esti-	1,000 Stocks				3,000 Stocks	
N	Mean	SD	mator	γ_0	γ_b	R^2	RMSE	RMSE	Δ RMSE
14,780	0.53	0.23	bols	0.30	0.65	0.2069	0.3220	0.1692	0.4623
14,780	0.54	0.21	bvck	0.21	0.80	0.2697	0.3007	0.2854	0.4024
14,780	0.56	0.21	bsw	0.17	0.84	0.2844	0.2910	0.2976	0.3873
14,780	0.58	0.21	bswa	0.14	0.86	0.3092	0.2811	0.3198	0.3760

Panel B: Medium bols Tertiale: Mean Selection Variable is 0.95 for Top-1000.

Univariate			Esti-	1,000 Stocks				3,000 Stocks	
N	Mean	SD	mator	γ_0	γ_b	R^2	RMSE	RMSE	Δ RMSE
14,347	0.95	0.15	bols	0.24	0.77	0.1230	0.3096	0.1855	0.4002
14,347	0.94	0.15	bvck	0.26	0.76	0.1167	0.3115	0.1815	0.4019
14,347	0.94	0.15	bsw	0.19	0.82	0.1350	0.3072	0.1951	0.3970
14,347	0.95	0.16	bswa	0.16	0.85	0.1735	0.2999	0.2298	0.3875

Panel C: High bols Tertiale: Mean Selection Variable is 1.57 for Top-1000.

Univariate			Esti-	1,000 Stocks				3,000 Stocks	
N	Mean	SD	mator	γ_0	γ_b	R^2	RMSE	RMSE	Δ RMSE
15,063	1.57	0.42	bols	0.46	0.61	0.2623	0.4836	0.1520	0.6184
15,063	1.50	0.34	bvck	0.35	0.71	0.2369	0.4553	0.1791	0.5334
15,063	1.45	0.29	bsw	0.19	0.86	0.2443	0.4386	0.1866	0.5130
15,063	1.43	0.29	bswa	0.07	0.95	0.2975	0.4207	0.2224	0.4981

Explanations: This table is equivalent to Table 5, but divides firms into tertiales based on OLS market beta.

Interpretation: The performance ordering is usually the same, regardless of firm set and metric, with one exception: bvck and bsw underperforms bols for the highest bols tertiale among 1,000 stocks on the R^2 metric.

Source: 2analysis.Rout, May 16, 2020.

Table 𝒞.3: Predictive Regressions By OLS Market-Beta S.E., Dec 1973-2019

Panel A: Low OLS SE Tertiale: Mean Selection Variable is 0.09 for Top-1000.

Univariate			Esti-	1,000 Stocks				3,000 Stocks	
N	Mean	SD	mator	γ_0	γ_b	R^2	RMSE	RMSE	Δ RMSE
15,629	0.76	0.34	bols	0.17	0.82	0.5736	0.2517	0.6102	0.2846
15,629	0.76	0.34	bvck	0.16	0.83	0.5750	0.2500	0.6125	0.2818
15,629	0.76	0.33	bsw	0.14	0.85	0.5772	0.2473	0.6174	0.2768
15,629	0.77	0.33	bswa	0.12	0.87	0.5943	0.2407	0.6311	0.2704

Panel B: Medium OLS SE Tertiale: Mean Selection Variable is 0.09 for Top-1000.

Univariate			Esti-	1,000 Stocks				3,000 Stocks	
N	Mean	SD	mator	γ_0	γ_b	R^2	RMSE	RMSE	Δ RMSE
13,614	0.97	0.39	bols	0.26	0.73	0.4651	0.3193	0.4478	0.4235
13,614	0.96	0.37	bvck	0.24	0.76	0.4680	0.3144	0.4509	0.4123
13,614	0.96	0.35	bsw	0.20	0.80	0.4745	0.3070	0.4569	0.3993
13,614	0.97	0.35	bswa	0.16	0.84	0.4969	0.2977	0.4835	0.3865

Panel C: High OLS SE Tertiale: Mean Selection Variable is 0.21 for Top-1000.

Univariate			Esti-	1,000 Stocks				3,000 Stocks	
N	Mean	SD	mator	γ_0	γ_b	R^2	RMSE	RMSE	Δ RMSE
14,947	1.34	0.60	bols	0.48	0.60	0.3866	0.5207	0.2484	0.7094
14,947	1.28	0.51	bvck	0.39	0.70	0.3811	0.4859	0.2866	0.6034
14,947	1.24	0.45	bsw	0.28	0.82	0.3909	0.4674	0.2921	0.5827
14,947	1.24	0.43	bswa	0.19	0.89	0.4282	0.4489	0.3179	0.5671

Explanations: This table is equivalent to Table 5, but divides firms into tertiales based on the noisiness, here measured by the standard error of the OLS market beta.

Interpretation: The performance ordering is usually the same, regardless of firm set and metric, with one exception: bvck underperforms bols for the highest bols tertiale among 1,000 stocks on the R^2 metric.

Source: 2analysis.Rout, May 16, 2020.

$\mathscr{D}$ Other Observations

This paper is the distillation of many months of exploration and hundreds of tables, classifications, and robustness checks. This section discusses a number of omitted but relevant observations and findings.

$\mathscr{D}$.1 Estimation-Related Observations

- The online appendix discusses the commonalities and differences in the use of the prior in my robust context vs. its use in a Bayesian context. Suffice it to state that my use of the prior is not Bayesian—the paper does cover one Bayesian estimator bvck, which it outperforms. Slope winsorization is, to the best of my knowledge, a novel robust method (Tukey (1960)). It performs as well as band winsorization, and carries the advantage that the prior can be specified over the regression coefficients rather than over the residuals. It is specialized to this application, where the intercept is zero and the mean slope is known.

- When predicting the 1-month ahead beta (based on daily stock returns) instead of the 1-year ahead beta, the dependent variable is much noisier. When predicting 2-year ahead betas, the predictions become again less reliable. This is because the underlying betas are themselves not stable but mean-reverting.[27]

  However, the ordering of estimator performance always seems to remains the same—bswa outperforms bsw outperforms bvck outperforms bols. See Table $\mathscr{D}$.1.

- Survivorship bias seems unimportant. The inference is similar when predicting betas for stocks that have betas based on a full set of returns in the forecasting period vs. stocks that disappear in the second half of the forecasting period.

- Not shown, results with Fama-Macbeth-like specifications are always similar to those from the pooled regressions. However, F-M is ill-suited to this application. Unlike stock returns, betas are not uncorrelated over time.

[27] The deterioration could have been due to outliers, but investigation shows that this is not the case. It appears as strongly with predictors that have winsorized the outliers as those that have not.

tbl:1mopred in beta-appendix.tex, August 17, 2020

Table $\mathscr{D}$.1: Predicting 1-Month Ahead 1-Month Beta

Data Set	N	Predicted	Predictor	R^2
1,000	529,951	bols 1mo	bols 12 mo	0.2776
			bvck	0.2765
			bsw	0.2753
			bswa	0.2908
3,000	1,562,546	bols 1mo	bols 12 mo	0.1750
			bvck	0.2765
			bsw	0.2753
			bswa	0.2908

$\mathscr{D}$.2 Predicting Future Average Returns With Betas

- Not reported, in my sample, the relationship between estimates from *all* beta estimators and future average rates of return was generally negative but insignificant and unreliable.[28] It seems highly unlikely that a better ex-ante beta estimator will ever show a reliable positive relationship between betas and future returns *in my 1973-2019 CRSP sample*.

  Of course, this does not invalidate the use of beta to tilt an index-like portfolio to obtain a lower overall portfolio risk and thus a better mean-variance tradeoff.

- Sort-and-split portfolios are unsuited to investigating the performance of estimators that are as highly collinear as the different betas considered here. When the number of splits is modest, most stocks remain in the same portfolio regardless of choice of beta estimator. If the test portfolio is constructed to disentangle and optimize beta exposures (as in Hoberg and Welch (2009) or in Back, Kapadia, and Ostdiek (2015)), the inference is equivalent to that presented in my paper (forecasting beta).

[28] The exception is bfp, which has a significant reliable negative relationship. However, this is more likely due to its non-beta related aspects.

- There is also another category of beta users: managers and regulators are often required by law to use a CAPM—regardless of whether the model holds or not. The differences among estimators are meaningful. For a CAPM user with a risk-premium of 6%, the typical inferred expected return differences due to beta estimators is about 60 to 200 basis points per annum. However, it is not clear whether better betas result in better cost-of-capital estimates.[29]

$\mathscr{D}$.3 Reflections on Other Estimators

All popular beta estimators should be viewed as empirical methods that emphasize certain features of the financial data. None of them are based on decision theory that reflects empirical reality.[30] The theoretical properties of all estimators in realistic samples are therefore largely unknown. However, the most common use of beta rarely requires in-sample properties. Instead, it requires good out-of-sample forecasting power.

$\mathscr{D}$.3.1 Other Estimators

- The online appendix also considers some other estimators, specifically a multi-month bakx-like estimator, a level-winsorized estimator and a band-winsorized estimator. None of them had superior performance. The band-winsorized beta estimator (with an optimized bandwidth) can show performance almost but not quite as good as the slope-winsorized beta estimator.

- I also experimented with industry-average betas, peer-average betas, size-average betas, and combinations thereof. All performed terribly for forecasting the betas of their individual constituents. Firms are too idiosyncratic.

[29] Levi and Welch (2017) discuss the capital-budgeting perspective.

[30] Most importantly, none can account for the key facts [1] that the underlying betas are mean-reverting and [2] that there are outliers. Moreover, there is neither an a-priori known process for how (underlying true) betas mean-revert, nor do we know the distribution of rare outliers—itself very difficult to estimate because of their rare incidences.

- I also experimented with estimators that use accounting information, as suggested in Rosenberg and Guy (1976b). Estimates thus obtained did not meaningfully outperform the estimates presented here. However, estimation and comparisons become more difficult, because not all stocks have Compustat data and measures are often seasonal. It is perhaps not a surprise that accounting information did not help greatly—after all, the data is only updated once every three months.

$\mathscr{D}$.3.2 The Vasicek Estimator

- Other than the slope-winsorized betas, the Vasicek beta seems to predict the future OLS beta better than other common estimators. Nevertheless, Vasicek estimators have had demonstrably lower rates of adoption than inferior alternatives.

  The neglect may be for several reasons. First, Vasicek estimators unintuitively entangle beta estimates with those of unrelated stocks. Including or excluding unrelated stocks in the sample changes every other beta estimate. Second, there is extra effort involved. Even the simplest versions require running a first-step OLS regression, calculating cross-sectional statistics, and then going back to readjusting the OLS estimates based on (cross-sectional and time-series) standard errors. The Vasicek-derived estimators in Karolyi (1992) and Levi and Welch (2017) require even more steps. Third, Vasicek betas have never become *the* standard. They are more often not used than used. Fourth, in the absence of recent public performance benchmarks, their superior performance may not have been fully appreciated. Fifth, researchers often want to investigate another issue and not add effort by first investigating how to estimate betas. Their interest in beta may be perfunctory. Many papers investigate different issues and are content merely to add some (any) control for beta. Worse control may not necessarily be harmful. Sixth, diversified portfolio betas (rather than individual stock betas) are more forgiving: in the extreme, any reasonable estimator for the the market of the market portfolio should yield the number 1.

  The slope-winsorized beta estimates sidestep these problems. I hope their ease of use will lead to wide adoption.

- The Vasicek beta is a weighted average of the firm's own beta and the cross-sectional mean beta. The latter is *not* 1.0 in the full CRSP cross-section, but varies year by year with the number of firms in the cross-section. Surprisingly, for the bottom marketcap tertiale of CRSP stocks (which this paper did not include), the average beta is also a better predictor than the own beta, i.e., when the two are not averaged but considered individually. For these firms, the own bols is a terrible benchmark and estimator of future bols's.

- The Vasicek estimator has predicted well for a reason hitherto not widely understood. Contrary to popular belief, most of the Vasicek superior performance does not come from its presumed optimal Bayesian or random-effects properties, as manifested in *proportional* shrinkage. Instead, most of its good performance comes from its handling of extreme outliers.

  It would be interesting to explore other settings in which random-effect panel techniques have been proposed as solutions to one problem, only to be inadvertently rescued by their coincidental pull on outliers. One could also estimate a parameter that modulates the shrinkage of near vs far residuals.

$\mathscr{D}$.3.3 The Dimson Estimator

The online appendix includes the analysis of the Dimson (1979) estimator.[31] Unlike the other estimators, it cannot be expected to forecast the future empirical bols, because it seeks to estimate a different liquidity-adjusted beta. If liquidity-caused deviations from bols are time-persistent, it should predict itself well, however.

- The Dimson (1979) beta estimator is perhaps the second-most-widely-used estimator in the literature behind the monthly-return OLS estimator. It corrects for non-synchronous trading by including leads and lags of the market rate of return in

[31] Schols and Williams (1977) offered an earlier non-synchronicity-adjusted market-beta estimator. It does not perform better, was criticized by Dimson for its lack of efficient use of data, and is nowadays used less often.

the market-model regression. The Dimson beta is then the sum-total of the coefficients on differently timed market rates of return.[32] The Dimson estimator is easier to code than the Vasicek estimator. No cross-sectional first-pass statistics are needed.

The original Dimson paper used its namesake estimator only on low-frequency (monthly) returns of (typically decile) portfolios. It also suggested Vasicek (1973) shrinkage and Blume (1971) regression inspired enhancements. However, the subsequent literature has mostly ignored these enhancements and latched only onto the summed coefficients. Moreover, the literature has also generously applied the Dimson beta in other contexts, such as in the context of individual stocks and/or daily stock returns.

Although the Dimson beta is intuitive and appealing, it suffers from a large efficiency loss. Empirical evidence suggests that there is a large performance gap between the Dimson beta and the OLS beta even for the ten most liquid stocks, and the historical Dimson beta does not predict the future Dimson better than the OLS beta.

More important for our perspective, even for the most illiquid stocks in our sample, bswa predicts the future Dimson beta better than the Dimson beta predicts itself. This suggests that it is preferable to use bswa rather than the Dimson beta for publicly-traded stocks (with daily stock returns), even for illiquid stocks where non-synchronous trading is a concern.

Importantly, its lack of usefulness in our application here does not negate its broader usefulness. Many assets (such as private equity and many non-stock-based funds) are not only relatively less liquid than privately traded stock, but also do not provide daily returns. In these cases, a Dimson-like correction would seem highly desirable. And this is what the estimator was originally developed for!

The referee can request including more detailed information.

[32]Fowler and Rorke (1983) point out that the Dimson (1979) estimator is incorrect. It does not yield consistent estimates. However, in practice, this is unlikely to be a big problem, and it is the original Dimson and not the Fowler-Rorke corrected estimator that remains in wide use.

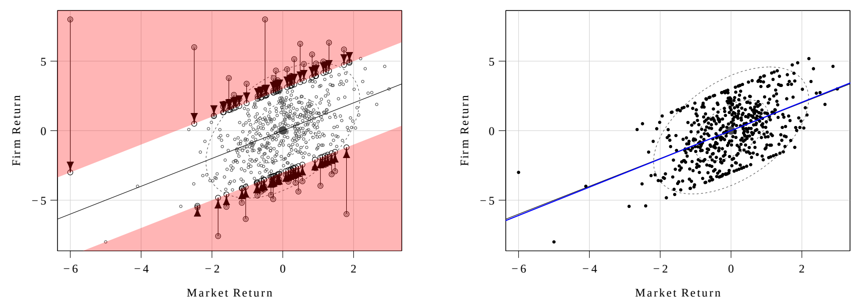
$\mathscr{E}$ Band Winsorizing Estimators

An alternative robust method that would take into account that the winsorization threshold should shift with the stock market performance on a given day would be a "band-winsorizer" around the slope coefficient 1.0. Repeating Figure 1, the following plot illustrates the band winsorizer.

Figure $\mathscr{E}$.1: Band Winsorization Concept

fig:conceptband in beta-appendix.tex, August 17, 2020

"Band"



Band

Band winsorization is more aggressive than slope winsorization with positive outliers when the market return is negative and with negative outliers when the market return is positive. Slope winsorization is more aggressive where it matters least for regression coefficient estimates—around market returns of about zero.

In cases in which the null hypothesis is zero, the band estimator collapses to the level winsorizer. In our case, the band estimator would winsorize stocks' rates of returns differently based on market movements with a 1-to-1 slope. By compressing the range of the dependent variable diagonally, each beta estimate becomes biased towards 1. The Martin and Simin (2003) estimator is a variant of a band winsorizer, in which the bandwidth is estimated together with model. To the best of my knowledge, a simple band winsorizer has not been used in the market-beta context in the literature. In the beta context, both estimators perform well, but the slope estimators performed just a little better. As Tukey (1960) notes, it is more important to use *any* robust estimator than it is to use a specific one. (Having said this, level winsorization should not be used.)

Use of either band or slope winsorization would seem a matter of preference. In the market-beta context, slope winsorization makes sense. Its winsorization parameter space seems more natural. For band winsorization, the winsorization parameter is in return space. The user has to provide a prior on what a reasonable stock return residual should be. For slope winsorization, it is in beta space. The user has to provide a prior on what a reasonable beta slope should be. In the context of beta, where reasonable ranges of betas are known and unlikely to differ with stock characteristics, slope winsorization seems easier than band winsorization. In the context of other applications that require multivariate predictions with intercepts and/or many variables, the reverse seems the case.

$\mathscr{F}$ Regression Tendencies of OLS Market-Beta Estimates

In the Monte-Carlo simulations, the true OLS beta distribution is assumed to come from a sampled OLS beta distribution, albeit with winsorization at the 0.001 and 0.999 levels. This weak a winsorization disadvantages the bsw estimator relative to other estimators. Thus, the suggested performance advantage of bsw in the simulation is conservative.

Figure $\mathscr{F}$.1: Future vs Current OLS Market Beta

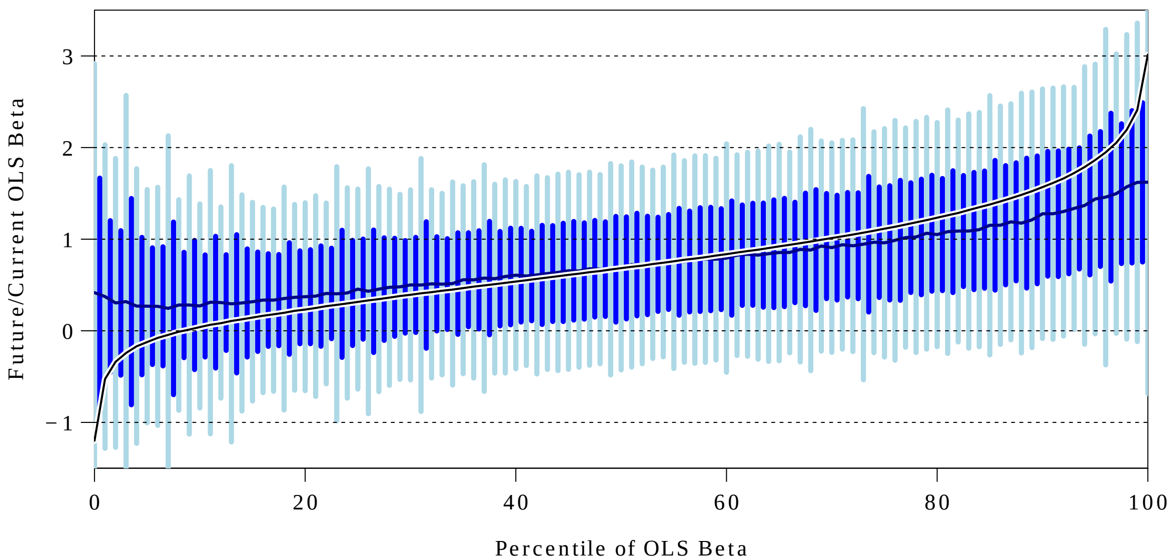


Figure $\mathscr{F}$.1 sorts historical 1-year OLS betas into 100 bins. It then plots the mean beta and its one- and two-standard deviations intervals within each bin for the 1-year beta one year ahead. (In addition, the black line plots the historical beta itself.)

The figure shows that the lowest 5 percentiles and top 2 percentiles are no longer monotonically predictive of future OLS betas. Winsorizing OLS betas smaller than –0.25 and greater than about 2.2 (assuming that there are no lower or no higher betas) would therefore improve the prediction.

$\mathscr{G}$ Some More Details on the Monte-Carlo Simulations

[Insert Table $\mathscr{G}$.1 here: **Drawn (Presumed-True) Monte-Carlo Inputs**]

Table $\mathscr{G}$.1 gives some statistics on the market-model residuals, sigmas, and betas drawn in the analysis from the assumed true distributions. The sharpest difference is that the empirical distribution retains more outliers and more skewness.

[Insert Figure $\mathscr{G}$.1 here: **Performance Under (Unknown) True Betas**]

Figure $\mathscr{G}$.1 shows the performance of the slope-winsorized market-beta bsw and the two other commonly-used alternatives (the Blume and the Vasicek betas bvck) as a function of the true market-beta.[33]

The estimators trade off performance depending on the (unknown) true market-beta. All three estimators outperform the OLS estimator for betas between 0.7 and 1.4. This should not be surprising, because all three estimators use information that the mean beta estimator is 1.0—information that the plain bols estimator ignores. (bmblm uses it in an obvious manner; bvck uses it in a Bayesian manner [*and* uses information about the true standard deviation of market-beta in the sample]; bsw uses it in a robust manner.) The range (of 0.7 units around 1.0) where all estimators easily outperform bols is about one standard deviation in the distribution of market-betas. As the true market-beta deviates further from 1.0, the slope-winsorized market-beta keeps its performance advantage longer relative to bvck, but eventually underperforms relative to bols, too.

The simulation (generations and estimation) code in R is available upon request. It is not included, because it is too long. (It is a few pages worth.)

[33] The figure is not symmetric even for the normal scenario, because stocks with higher sigma are given higher betas, on average.

Table $\mathscr{G}$.1: Drawn (Presumed-True) Monte-Carlo Inputs

tbl:mm in beta-appendix.tex, August 17, 2020

Panel A: Annual Market Model Residuals

	Gaussian		Empirical	
	Mean	SD	Mean	SD
Ann Mean	0.000	0.002	0.000	0.002
Ann Variance	0.008	0.001	0.009	0.002
Ann Skewness	0.000	0.153	0.386	0.990
Ann Kurtosis	2.977	0.300	8.833	6.600

Panel B: Drawn Presumed-True Market Beta and Sigma

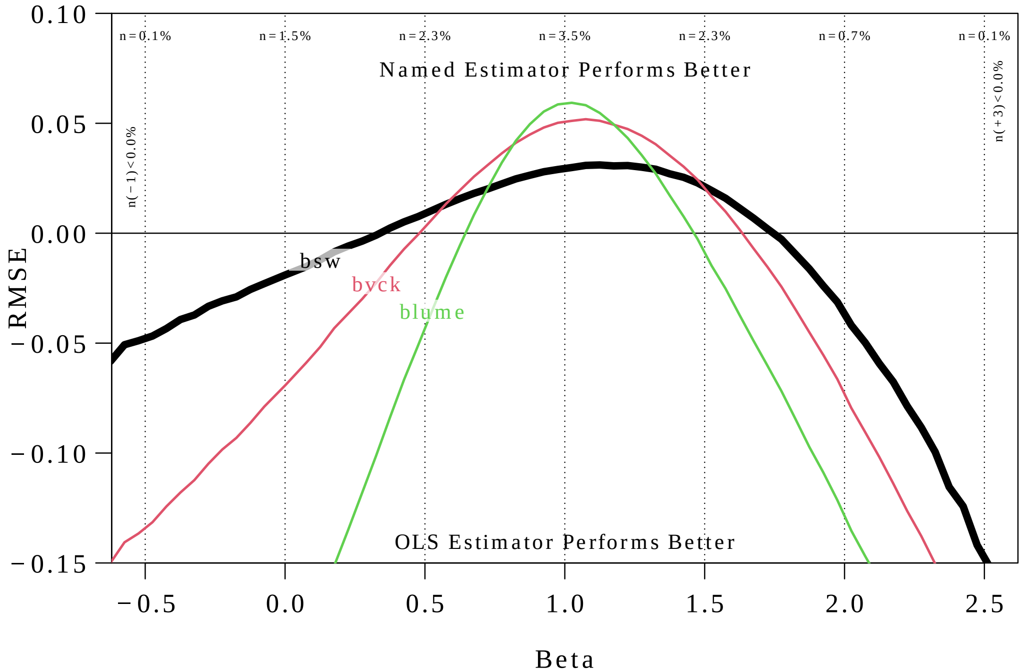
	Gaussian		Empirical	
	Mean	SD	Mean	SD
beta	1.00	0.58	1.00	0.58
sigma	0.025	0.014	0.025	0.016

Explanations: These are the distributions used in the simulations.

Interpretation: The skewness and kurtosis of market-model residuals is higher under the empirical resampling simulations.

Figure G.1: Performance Under (Unknown) True Betas

Panel A: Performance Under Normal Scenario



Panel B: Performance Under Empirical Scenario