Market-Beta

Ivo Welch

Nov 2019

Motivation

Q: Why still bother with “boring” old market-beta?

A: Market-beta is interesting even w/o CAPM (ER)

- Measures risk contribution to diversified pfio (m).
- Measures hedging against bear markets
- Down-Beta Theories (as in Ang+ or Lettau+)
- Betting against Beta (as in Frazzini-Pedersen)
- Pragmatic: used in regulation, etc.

Estimation

Q: Does estimation make a difference?

A: Only for individual stocks.

- Matters little for portfolios.
- Any method is roughly equally good.
- Errors average out
- Extreme: value-weighted stock beta is 1.0.

Performance Metric

Q: How to assess beta estimates?

A: Prediction

- of future ols(/other) 1-mo or 1-yr market-beta **estimates**
- never of future average returns.

Unknown True Beta

Q: Proxy Estimate vs True Beta?

A: Wait just a little.

- I will tell you exactly how good my proxies correlate with the true unknown market-beta, not just with the future market-beta.

Unknown True Beta

Need a good benchmark for comparing my estimator:

1. OLS — obvious (self-) estimator
2. Vasicek — best performing estimator known.

Vasicek

- random-effects estimator = bayesian shrinkage
- Run OLS Regressions
- Calculate x-sect means and sds of betas
- For each stock i,

$$b_{i,vck} = w_i \cdot b_{xs} + (1 - w_i) \cdot b_{i,ts},$$

where $w_i = \sigma_{i,ts} / (\sigma_{i,ts} + \sigma_{xs})$.

- **Blume shrinkage $\neq$ Vasicek shrinkage, as claimed by FP**

Other Important Choices

- Always use daily stock returns, never monthly.
- Use about 1-3 years of data.
- Never use industry beta for individual stocks.
  - Indeed, they are less noisy;
  - ...just like using “1” — low predictive power.
- vasicek has derivatives
  - (random-effects and/or bayesian justification if no drift.)
  - Levi-Welch linear de-bias.

more alternatives below: Dimson, Frazzini-Pedersen, Levi-Welch, Ait-Sahalia-Kalnina-Xiu, Martin-Simin, etc.

Vasicek Disadvantages

- Ad-Hoc (i.e., wrong claim of optimal design)
  - "optimal estimator design" was never suited to problem:
  - vasicek is designed for measurement error,
  - not for underlying beta drift
  - (ergo 12–24 months windows)
- Vasicek has good R^2, but is badly biased
  - levi-welch (2017) suggests empirical de-biasing
  - requires another linear debiasing stage
- spooky entangled estimates
- requires multi-step ts and xs procedure

known, but rarely used.

We Could Use a Simpler Estimator

...and if it is better, all the better!

The New Estimator

Standard Bayesian Use of Prior

Data

Posterior

Priors

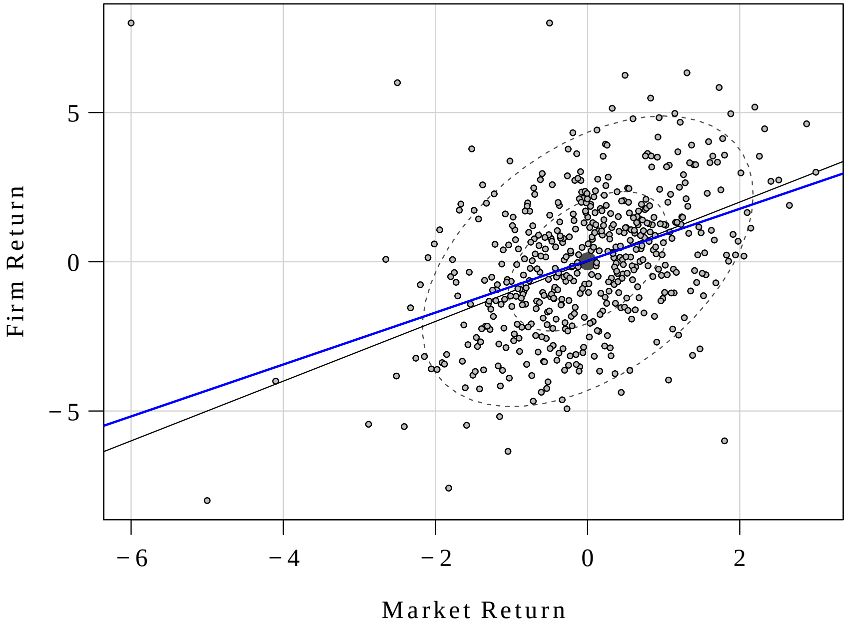
- Involves arguments about reasonable priors
- Often painful—days babysitting, not minutes.
- Usually primarily in dedicated estimation papers

(Ab-)Use of Prior

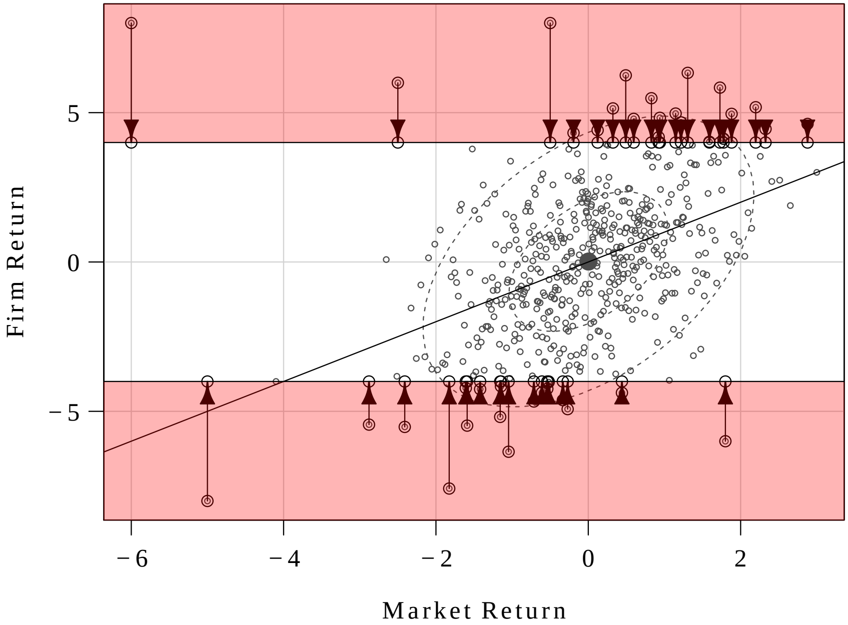
$$\begin{tikzcd}
\text{Data X} \arrow[dr] & \\
& \text{Posterior} \\
\text{Priors} \arrow[uu] &
\end{tikzcd}$$

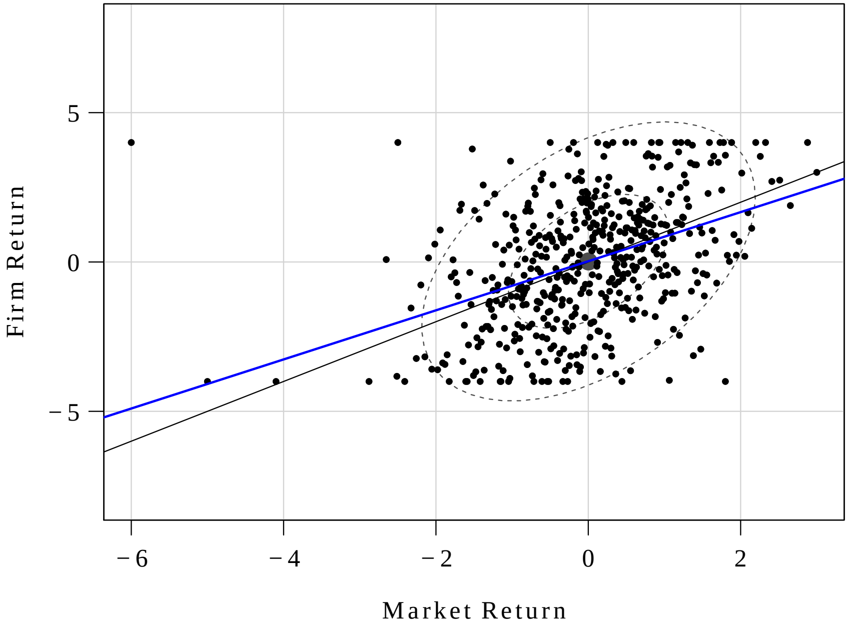
- Still involves arguments about reasonable priors
- Easy to use. minutes, not days.
- It's just a robust = winsorizing method.

- Likely novel method.



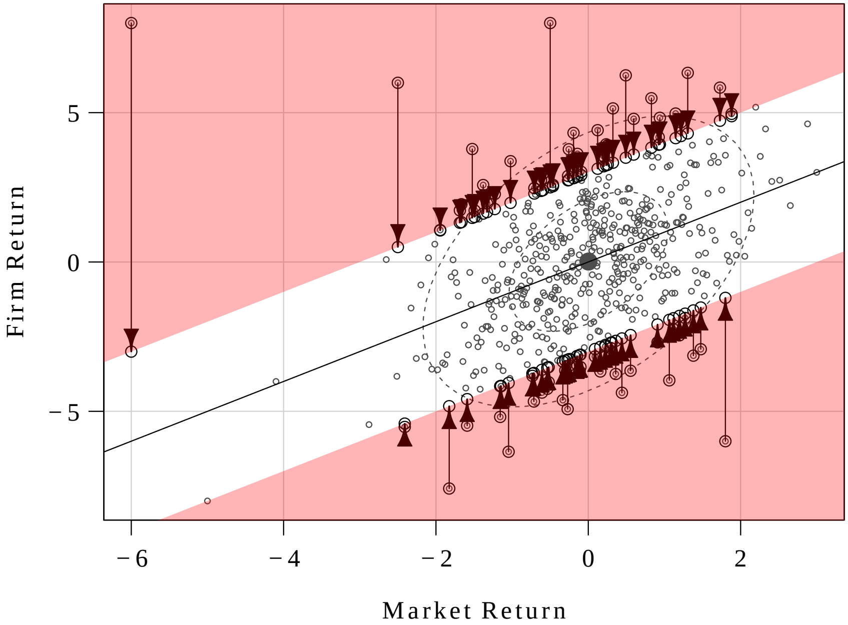
5
0
−5
−6
−4
−2
0
2
Firm Return
Market Return

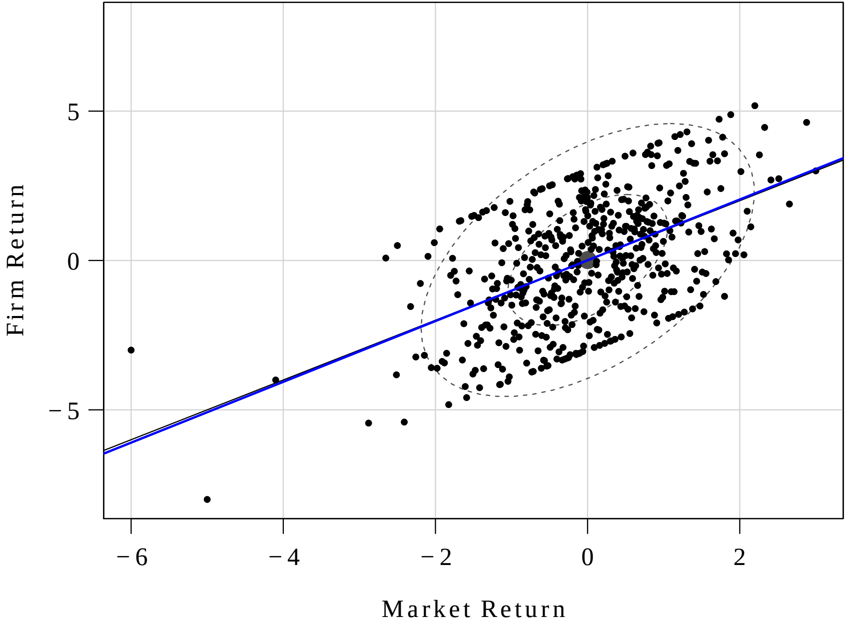

5
0
−5
−6
−4
−2
0
2
Firm Return
Market Return


Firm Return
5
0
−5
−6
−4
−2
0
2
Market Return

Bad Idea
Biases Estimator Down

(commonly used)

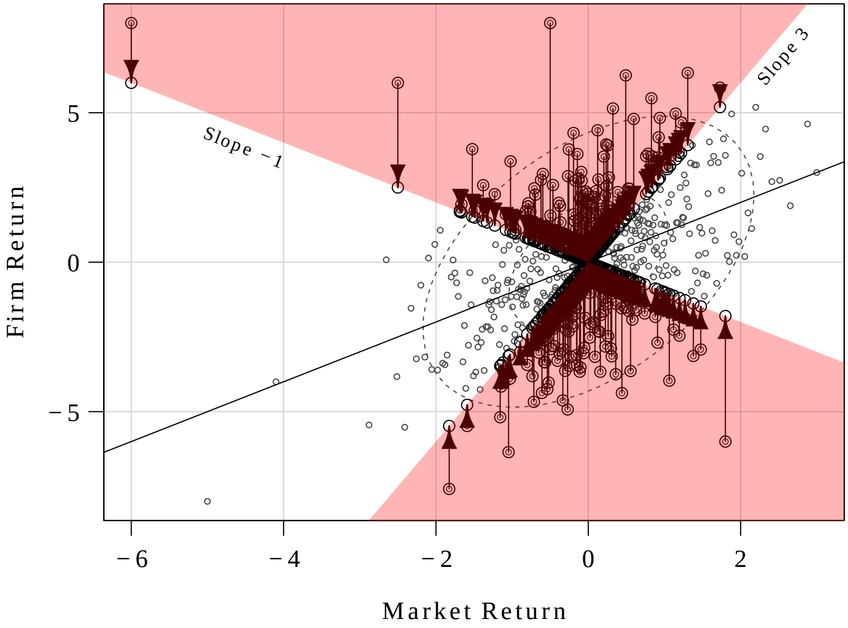

Firm Return
Market Return
5
0
−5
−6
−4
−2
0
2

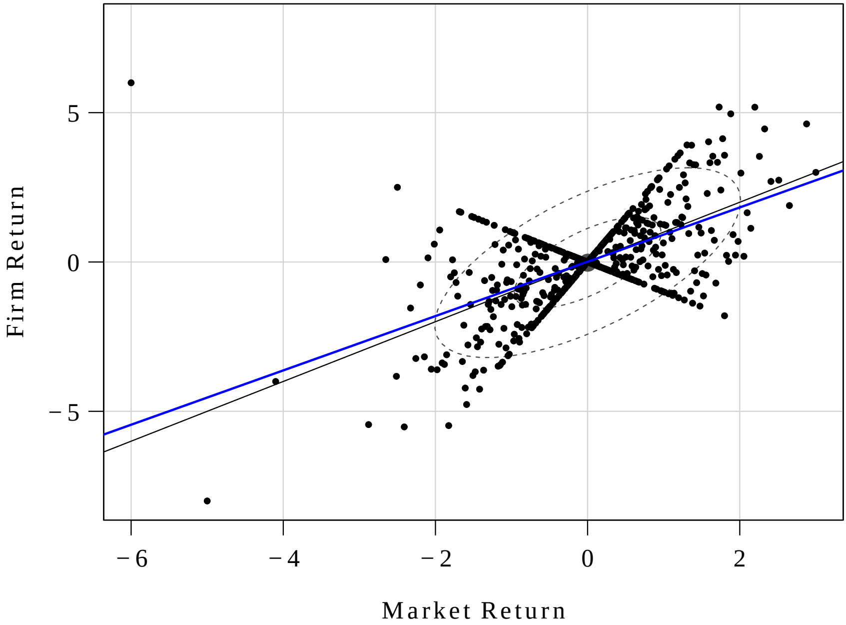


5
0
−5
−6
−4
−2
0
2
Firm Return
Market Return

Good Idea

(never used afaik)


Slope 3
Slope −1
5
0
−5
−6
−4
−2
0
2
Firm Return
Market Return



Firm Return
Market Return
5
0
−5
−6
−4
−2
0
2

Good Idea

(never used afaik)
(happens to work a little better)

- non-Bayesian use of prior
- With wide priors, like –2 to +4, this use should not be very costly, **even if** the panel is true OLS w/o outliers.

How Different From Bayesian Prior Use?

Very Different! If OLS estimate is $\hat{b} = 0.8$:

- Bayesian use of prior of –2 to +4 would do almost nothing to the resulting beta.
  - Bayesian OLS-type prior would work on overall $\hat{b}$ estimate.
  - If final is near 1.0, Bayesian method says “just fine.”
- My use of prior of –2 to +4 could still do a lot.
  - Here, a prior(–2,4) still influences almost all points, and thus can drastically change estimate, even if estimate is close to 1.0. $\rightarrow$ can move a $\hat{b}$ away from 1.0.

  - PS: could use Bayesian with priors on mixed distributions, plain + outliers. Would be painful and rely on distributional priors. No one would use this.

Progress Plan

- Typically, we will predict $^{**}b_{i,y}$ with $^{**}b_{i,y-1}$:
- Apples to apples: Predicted OLS beta:

$$b^{**}{}_{i,y-1} \to \text{bols}_{i,y}$$

1. Direct Proxy Use: .... $-\text{RMSE}(\text{bols}_{i,y} - b^{**}{}_{i,y-1})$
2. Rebiase (Best Prediction): .... $R^2(\text{bols}_{i,y}, b^{**}{}_{i,y-1})$

1. w/ undecayed 1-year betas: $b^{**}{}_{i,y-1} \equiv \text{bsw}_{i,y-1}$.
2. w/ decayed long-history rets: $b^{**}{}_{i,y-1} \equiv \text{bswa}_{i,y-1}$.

Plan

Undecayed Slope Winsorized

$\mathrm{bsw}_{i,y-1}$

Recipe: beta slope winsorized (bsw)

Will use:

1. 12(–24) mos of **daily** stock returns
2. winsorize all returns ($\Delta_s = 3$):

$$\mathrm{rsw}_{i,t} \in \left(1.0 + \left[-\Delta_s, \Delta_s\right]\right) \cdot r_{m,t} \,.$$

3. estimate ols market-model

$$\mathrm{rsw}_{i,t} = a_i + \mathrm{bsw}_i \cdot r_{m,t} + e_{i,t}$$

$$\Rightarrow \qquad \mathrm{bsw}_i = \frac{\mathrm{cov}(\mathrm{rsw}_{i,t}, r_{m,t})}{\mathrm{var}(r_{m,t})}$$

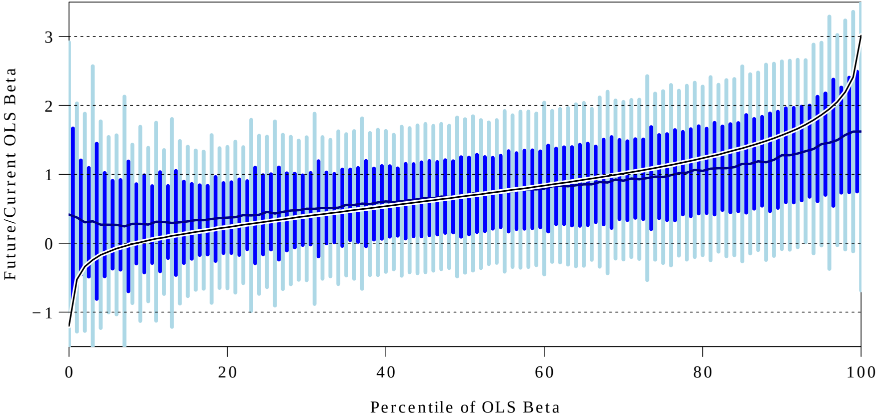
Holla? Why $\Delta_S = 3$?

1. because we are not doing philosophy or math;
2. any time you use a utility function or empirical functional form, you introduce equivalent assumptions;
3. we are analyzing empirical data;
4. we want parsimony and robustness.

Holla? Why $\Delta_S = 3$?

1. because we are not doing philosophy or math;
2. any time you use a utility function or empirical functional form, you introduce equivalent assumptions;
3. we are analyzing empirical data;
4. we want parsimony and robustness.

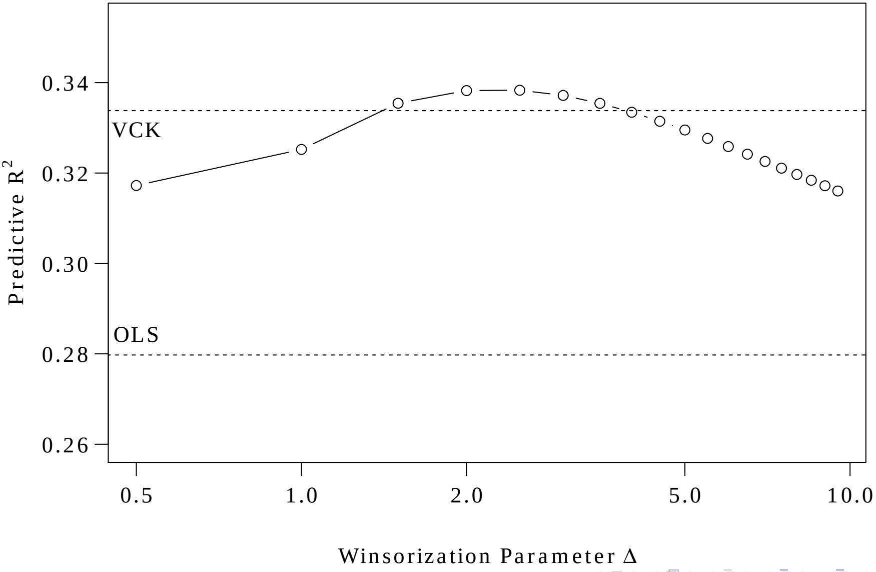
F2: why $\Delta_S = 3$?



$\Delta_S = 3$ Seems Sensible

1. $\Delta_S = 3$ is in top and bottom percentile of bols.
2. No more monotonicity between b_t and $E(b_{t+1})$.
3. Not independent, but also not much dependence.
    - fewer than 1% of betas exceed –1 and +3
    - fewer than 0.03% repeat in consecutive years
    - (yes, greater than $1\% \cdot 1\%$, but not by much.)
    - suggests most such extreme betas are more outlier based, than representative.

F4: Sensitivity to Δ_S, Full Sample



Reasonable Assessment for $\Delta_S = 3$

- not philosophical, but also not highly searched:
  - Basecase: $\Delta_S = 3$, i.e., from rsw($b \in [-2, 4]$)
  - Reasonable Range: $\Delta_s \in (1.5, 4.0)$.
    i.e., from $[-0.5, 2.5]$ or $[-3, +5]$.
  - lower Δ_s forces too much towards 1.
  - higher Δ_s forces too little.
  - Market-beta has an intuitive economic meaning...use it. Different from band winsorization, firm-specific?

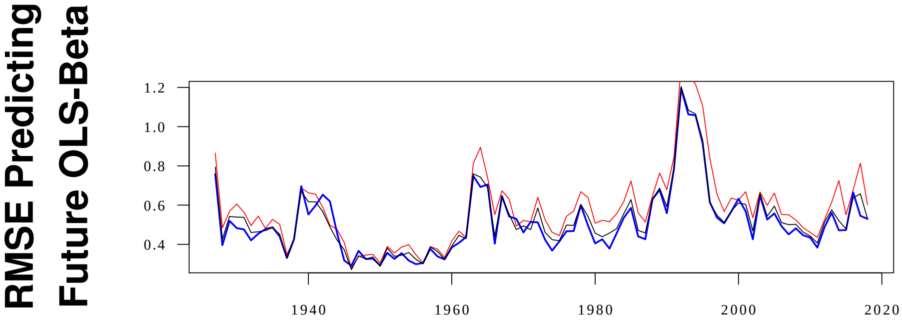
T2: Descriptive Stats

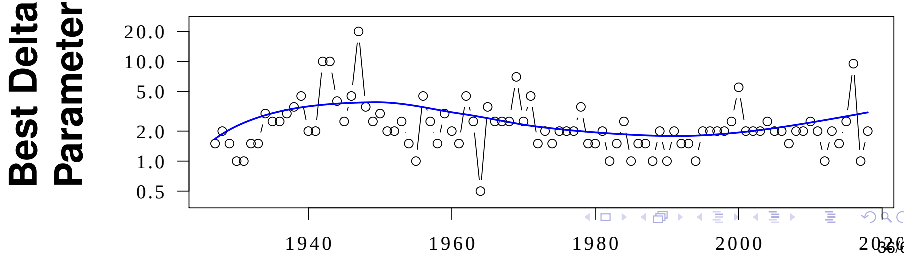
	Mean	SD	Abbrev	Predictor $b_{i,t}$
A	0.80	0.21	$\overline{\text{bols}}$	Past Year Firm-Average OLS
B	0.79	0.68	bols	(Own) OLS Market-Beta
C	0.79	0.55	bVCK	Vasicek Market-Beta
D	0.79	0.41	bLW	... Levi-Welch (0.75)
E	0.71	0.56	blw	Level-Winsorized (Δ_l=7%)
F	0.79	0.44	bbw	Band-Winsorized (Δ_b=3%)
G	0.79	0.43	bsw	Slope-Winsorized (Δ_s=3)
H	0.79	0.42		Slope-Wins Then Vasicek
I				Multivariate, bsw and bVCK
J				Multivariate A to G

T2: Performance ($\text{bols}_{i,+1}$)

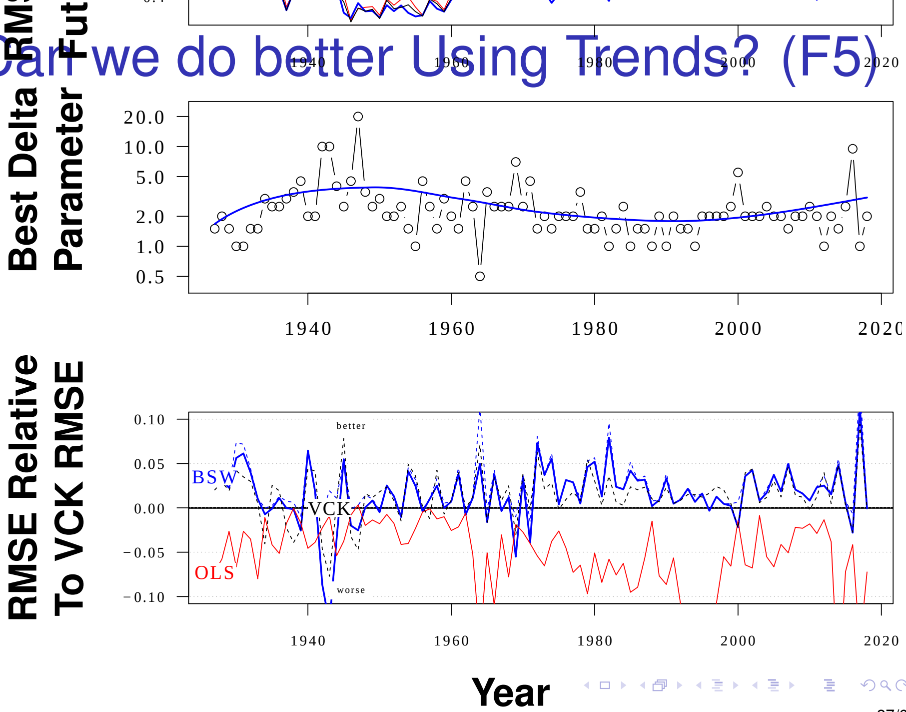
	Abbrev	RMSE	γ_0	γ_1	R^2
A	$\overline{\text{bols}}$	0.700	0.111	0.842	6.09%
B	bols	0.680	0.332	0.565	27.97%
C	bVCK	0.604	0.184	0.756	33.38%
D	bLW	0.589	–0.017	1.008	–"–
E	blw	0.621	0.271	0.721	31.84%
F	bbw	0.590	0.033	0.943	33.27%
G	bsw	0.587	0.008	0.977	33.82%
H		0.586	–0.014	1.008	33.97%
I					34.51%
J					34.77%

Can we do better Using Trends? (F5)





Can we do better Using Trends? (F5)



Did I Peek?

Yeah, but it would have made no difference.

Will show you soon.

Plan

(Infinitely but)
Decayed Slope Winsorized

$\mathrm{bswa}_{i,y-1}$

Decay

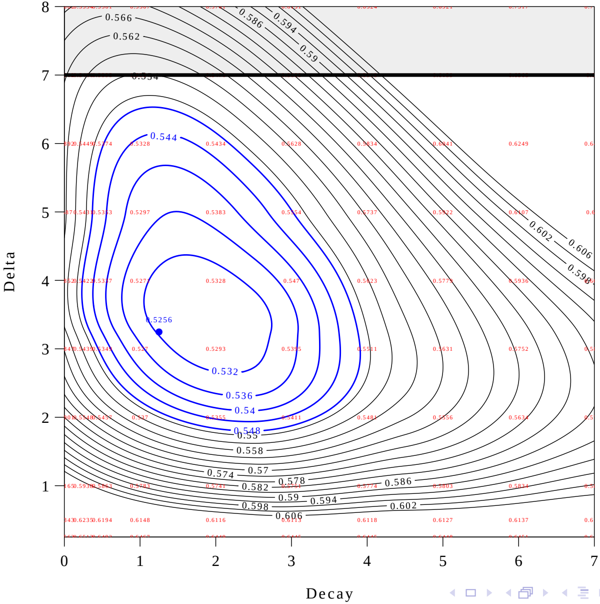
- Older stock returns are probably less relevant
- No good reason to use (common 1-year) cutoff.

Measure decay as $\rho/256$ per trading day:

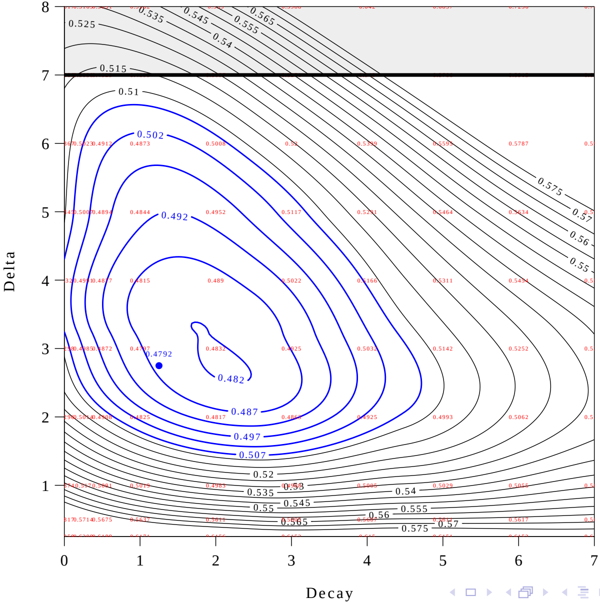
ρ	Decline	Halflife
1.0	0.4%/day	180 trading days
2.0	0.8%/day	90 trading days
3.0	1.2%/day	60 trading days

$(\mathbf{1.0} : 1 - 1/(1 + \mathbf{1.0}/252) \approx 0.004)$

F6: 1-Yr Pred **bols**, 1963–1973



F6: 1-Yr Pred bols, 1973–2018

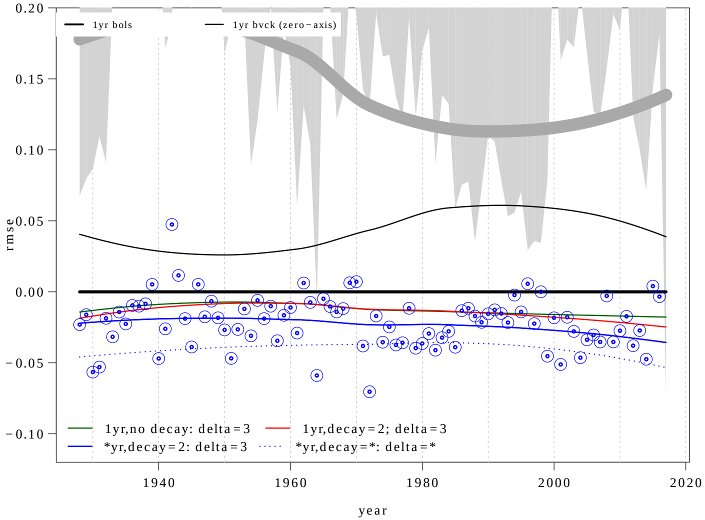


Refinements?

None greatly useful.

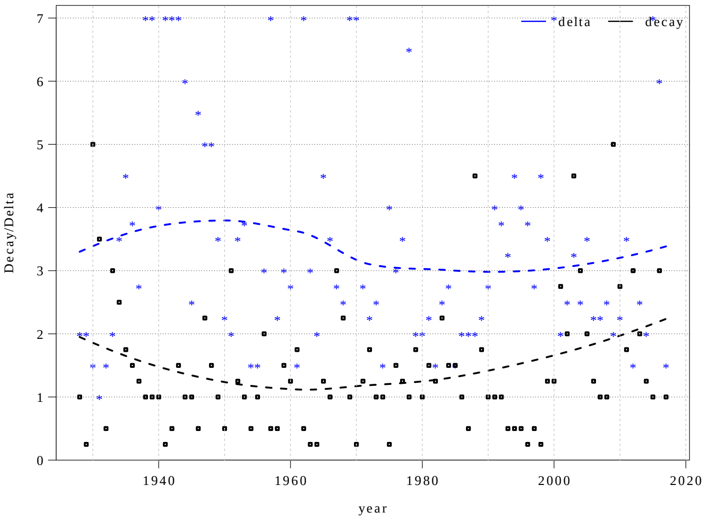
we are really just capturing and winsorizing extremes

F9: By Year?

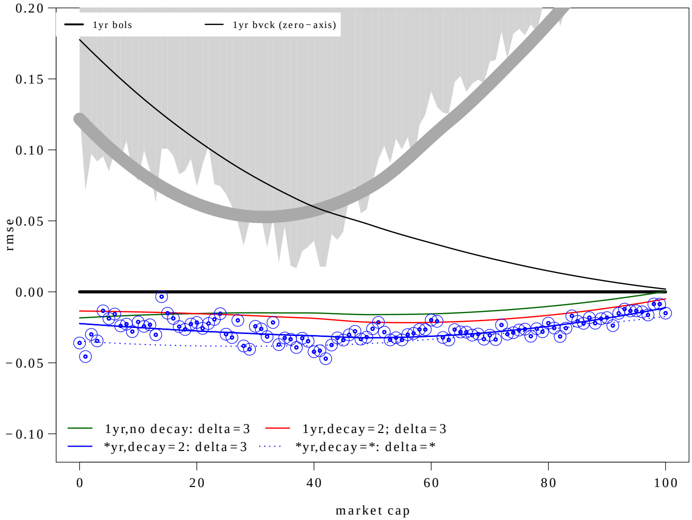


0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
rmse
1yr bols
1yr bvck (zero−axis)
1yr,no decay: delta=3
1yr,decay=2; delta=3
*yr,decay=2: delta=3
yr,decay=: delta=*
1940
1960
1980
2000
2020
year

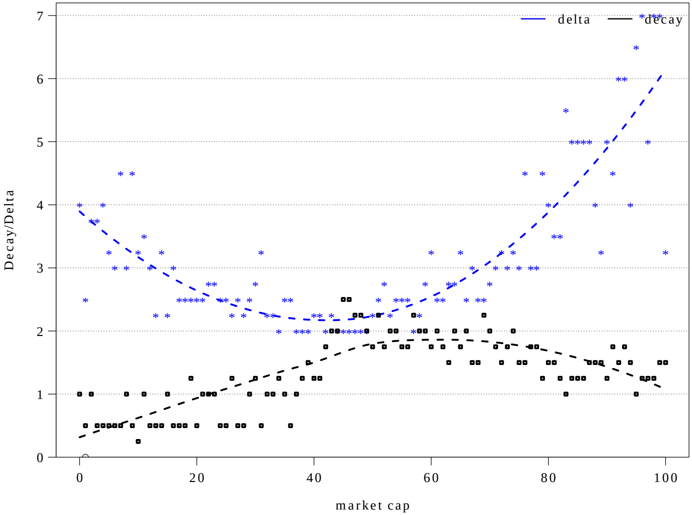
F9: By Year? — Ex-Post Δ_S^*



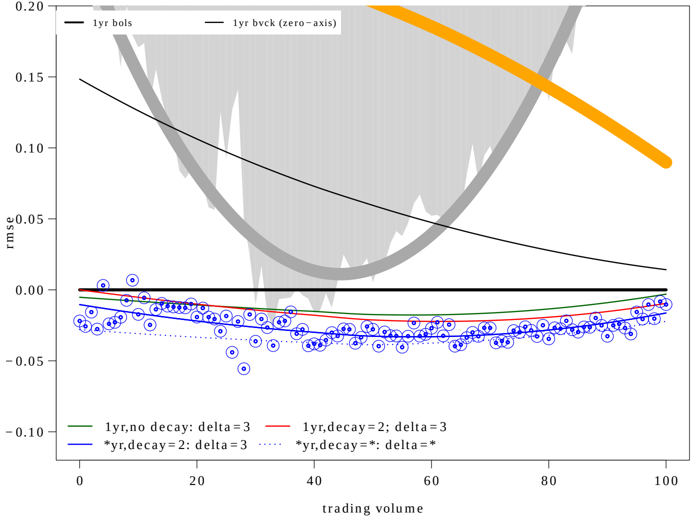
F10: By MarketCap?



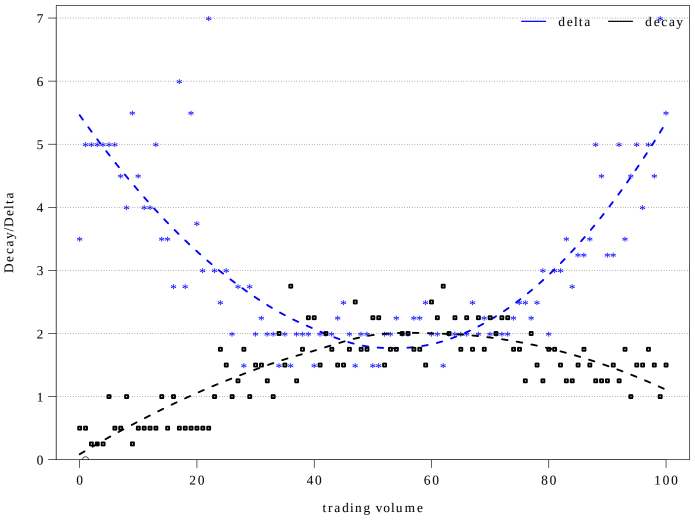
F10: By MarketCap? — Ex-Post Δ_S^*



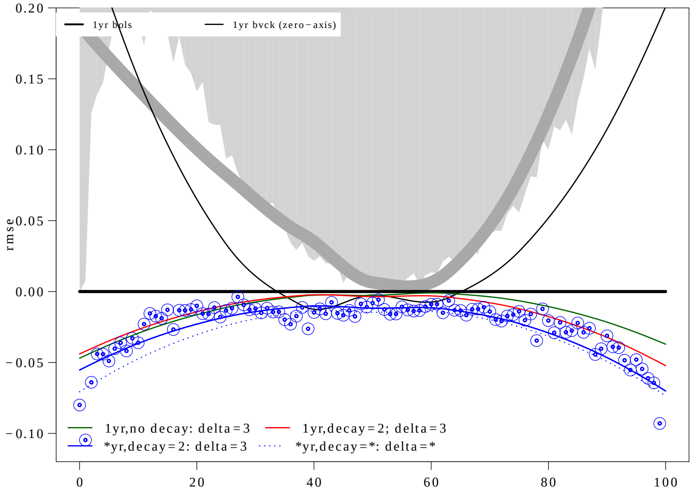
F11: By TradeVol?



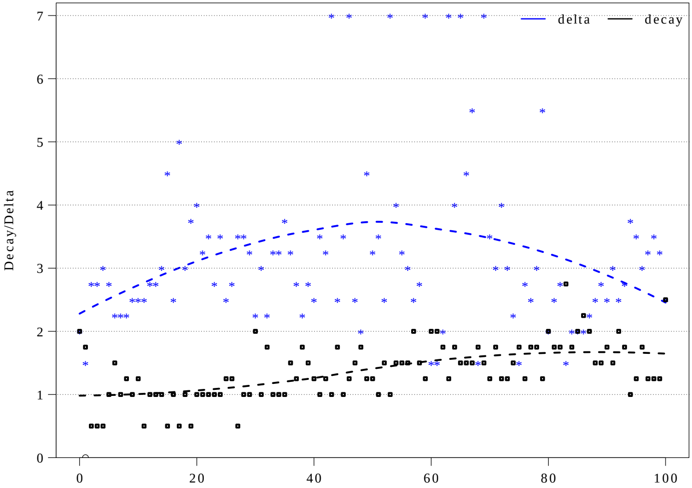
F11: By TradeVol? — Ex-Post Δ_S^*



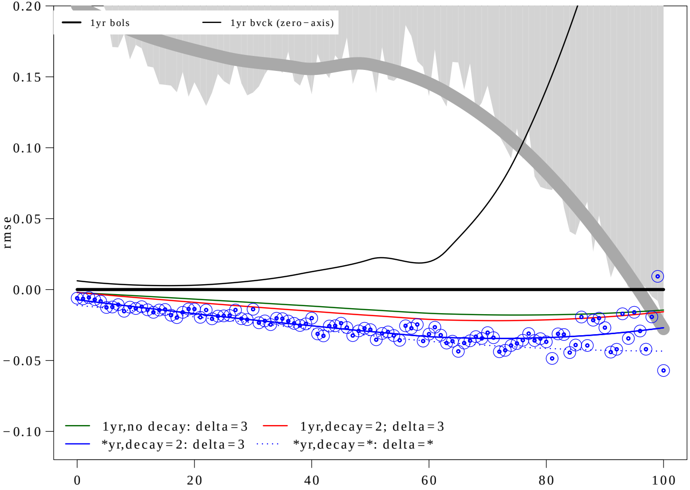
F12: By $\text{bols}_{y,t-1}$?



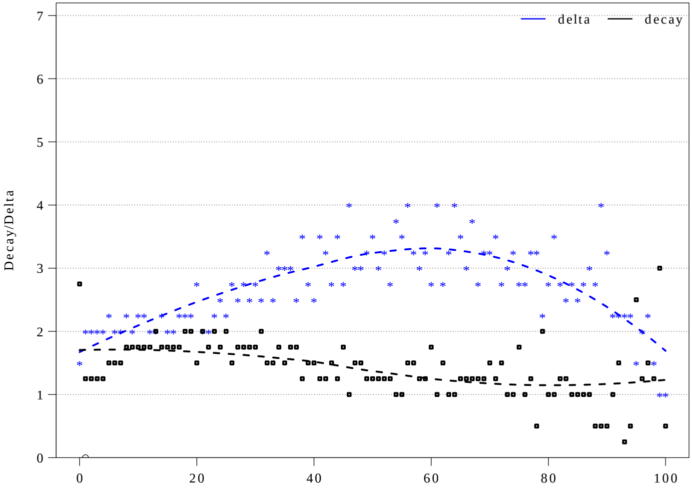
F12: By $\text{bols}_{y,t-1}$? — Ex-Post Δ_S^*



F13: By se($bols_{y,t-1}$)?



F13: By se($bols_{y,t-1}$)? — Ex-Post Δ^*_S



T4: Statistically

- Same Insights in regression format: Minor.
- Maybe a little marketcap or tradving vol
  - Larger firms have larger market-betas
- Basic Prediction:
  - bswa only: $R^2 = 34.74\%$.
  - Add log dolvol and cross: $R^2 = 35.89\%$.
  - Add log mcap and cross: $R^2 = 35.67\%$.

  - Then explain residuals on log-marketcap model

T4: Ala VCK by stderr(beta)?

- R^2 with adding all previous (CRSP) variables and x-variables: 0.01% to 0.46% (dollar trading volume).
- R^2 with adding tons of Compustat ratios: 0.01 to 0.22% (cash/at).

Estimator Benchmarking

Careful to use the same aset!

Y-Variable and Observations !

T6: 2.9 million obs

	(One-Period-Ahead)			(Lagged)				
	Mean	SD	Dependent	Independent	g_0	g_1	$R^2_{(\%)}$	rmse
A[1]	0.81	0.655	bols	bols	0.30	0.62	38.8	0.562
				$\overline{\text{bols}}$	0.15	0.82	7.6	0.622
				bVCK	0.17	0.79	43.7	0.498
				bdim	0.38	0.46	27.9	0.685†
				bsw	0.10	0.88	44.2	0.486
				bswa	0.07	0.92	**46.2**	**0.475**
B	0.80	0.539	bVCK	bVCK	0.23	0.71	50.6	0.411
				bswa	0.14	0.83	**53.4**	**0.377**
C	0.91	0.731	bdim	bdim	0.48	0.47	22.0	0.756
				bswa	0.21	0.86	**31.6**	**0.617**
D	0.80	0.485	bsw	bsw	0.21	0.73	53.9	0.355
				bswa	0.19	0.76	**56.3**	**0.340**
E	0.80	0.474	bswa	bswa	0.17	0.78	62.4	0.308

T6: 2.9 million obs

	Dep	Indep	g_0	g_1	$R^2_{(\%)}$	rmse
A[1]	bols	bols	0.30	0.62	38.8	0.562
		$\overline{\text{bols}}$	0.15	0.82	7.6	0.622
		bVCK	0.17	0.79	43.7	0.498
		bdim	0.38	0.46	27.9	0.685†
		bsw	0.10	0.88	44.2	0.486
		bswa	0.07	0.92	**46.2**	**0.475**

T6: 2.9 million obs

	Dep	Indep	g_0	g_1	$R^2_{(\%)}$	rmse
C	bdim	bdim	0.48	0.47	22.0	0.756
		bswa	0.21	0.86	**31.6**	**0.617**
D	bsw	bsw	0.21	0.73	53.9	0.355
		bswa	0.19	0.76	**56.3**	**0.340**
E	bswa	bswa	0.17	0.78	62.4	0.308

Unknown True Beta

Can Assess!

- If two proxies are drawn with noise from true value, the expected R^2 of each proxy with the true value is the squareroot of the R^2 of one proxy with the other proxy.
- If underlying beta is constant, and the R^2 of last year's beta estimate (proxy) with this year's beta estimate is 49%, then the association of one-year beta estimates with underlying true unknown betas is $\sqrt{56\%} = 75\%$ (cor > 87%).
- Conservative: If beta is moving, then bsw should be $R^2 > 75\%$.
- Conservative: If beta is moving, then bswa should be $R^2 > 79\%$.

Side Note

- bsw on bsw: 53.9%
  bswa on bswa: 62.4%
- $\Rightarrow$ bswa on true β: > 79% R^2, 89% correlation.
- Higher if time-varying beta
- This was equal-weighted, many small stocks.
  higher if we excluded noisiest stocks.

T7: Martin-Simin Robust (2.0M)

	Dep	Indep	g_0	g_1	$R^2_{(\%)}$	rmse
A^2	bols	bols	0.29	0.62	38.8	0.542
		bsw	0.09	0.88	44.0	0.472
		bswa	0.06	0.92	**46.0**	**0.461**
		bmm	0.30	0.69	42.5	0.514
		blts	0.33	0.68	40.5	0.533
F	bmm	bmm	0.21	0.70	49.7	0.453
		bswa	–0.02	0.92	**52.3**	**0.417**
G	blts	blts	0.21	0.68	45.7	0.472
		bswa	–0.04	0.89	**49.7**	**0.438**

T8: Frazzini-Pedersen (1.4M)

	Dep	Indep	g_0	g_1	$R^2_{(\%)}$	rmse
A^3	bols	bols	0.28	0.65	42.8	0.512
		bfp	–0.10	0.92	29.9	0.547
		bswa	0.07	0.93	**49.2**	**0.439**
H	bfp	bfp	0.54	0.46	20.6	**0.385**
		bols	0.74	0.31	27.1	0.564
		bswa	0.64	0.44	**31.1**	0.449

T9: Ait-Sahalia, Kalnina, Xiu (940k)

	(Dep)	(Indep)	g_0	g_1	$R^2_{(\%)}$
A^4	bols (1 mo)	btaq1 (1 mo)	0.67	0.33	7.4
		bswa (1 yr)	–0.04	1.08	**17.1**
I	btaq1 (1 mo)	btaq1 (1 mo)	0.63	0.31	9.7
		bswa (1 yr)	0.01	0.97	**20.6**

Does it matter?

Are betas different? Mean RMSE between bswa and:

$\overline{\text{bols}}$	0.47				
bols	0.20	bmm	0.17	bmols	0.46
bVCK	0.15	blts	0.21	bmvck	0.44
blw	0.19	bdim	0.40	btaq1	0.64
bbw	0.15	bfp	0.29	btaq12	0.25
bsw	0.10				

Simple Code:

```
_bswa <- function( ri, rm, Delta, rho ) {
   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 )
   beta <- function(...) coef(lm(...))[2]

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

bsw <- function( ... ) _bswa( ... , Delta=3.0, rho=0.0 )
bswa <- function( ... ) _bswa( ..., Delta=3.0, rho=2.0/256.
```

CFR Commercial

- Liquidity Issue Coming Out Soon. Acharya-Pederson. Amihud. Pastor-Stambaugh.
- Specialty: Provocative papers. Critiques. But others, too. Less Theory.
- PhD Students: Updates (cannot possibly upset authors—just newer data).
- Per paper CFR recursive 10-year impact is now between JFE and JFQA/RF.