A Comprehensive Look at the Empirical Performance of Equity Premium Prediction II

Amit Goyal Ivo Welch Athanasse Zafirov

May 2022

Dmitri turned out to be a great coauthor and his last name is after mine!

Why This Presentation?

- I opened my mouth too wide on the lunch table, trying to get Gregor or Shohini to give a lunch talk.
- When no one else stepped up.
  - Val got me.
  - So I had to eat my own dogfood.
  - 24 hours from zero to talk. no complaints.

Why This Paper?

Goyal-Welch (RFS 2008) made a simple point:

- As an investor, could I really predict the equity premium better with some other variable?
- (Should I be in the stock market?)

Answer: No and Don't-Know.

- Many IS forecasting regressions.
- Typical Interpretations:
  - Variables are compensation for bearing risk,
  - or they are blatant violations of market efficiency,
  - ... never as *ex-post* happenstance or luck.
- But many predictions turned out unstable.
  $\Longrightarrow$ Interpretation??
  - *ex-post* happenstance or luck?

Why This Paper II?

Goyal-Welch-Zafirov (2022) makes an even simpler point:

- Many papers have appeared after 2008.
- A good number claim OOS performance.
- Many papers claim *economic theory* as bulwark against spurious inference.
  - Lucas critique?
  - RAD ("Referees Against Data-Fishing") insist on it.
  - Based on solid economic theory, how could they go wrong?

Back To Original Question

- As an investor, can I *now* really predict the equity premium better?
  - Novel variables are usually characterized as compensation for investors bearing risk,
  - ...and sometimes as blatant violations of market efficiency,
  - ... but never as *ex-post* happenstance or luck.

What's New Since 2008?

- We included all papers from 1A journals and well-cited papers that we found and that we could replicate.
- About 30 variables from about 25 papers. (Wow!) More than GW (2008). Literature is lively!
- We had some amazing years since 2008 — a Great Recession and a bull market. (Covid "non-event.")
- ...but only about 10 new years on average since 25 original papers were published. Too early?

Very Low Bar

- We never do a full OOS experiment on the paper. We always reuse their own sample and just extend their samples by a few years.
- Question: *Including their own samples*, which papers remain reliable as performers in 2022:
  - in-sample (to believe in them now),
  - and out-of-sample (to have been investor implementable in-time).

(Enough if papers do not screw up badly and just keep their heads low.)

Problem 1: Novelty

Problem 1: Novelty

- Anonymous opinion from anonymous reviewer

Problem 1: Novelty

- Anonymous opinion from anonymous reviewer

  Useful to have these questions revisited but by its very nature the paper is rather derivative.

Problem 1: Novelty

- Anonymous opinion from anonymous reviewer

  Useful to have these questions revisited but by its very nature the paper is rather derivative.

- ...and that's correct. The original papers were more novel.
  - Our own paper is clearly less "innovative" than the big 1A journal ideas, findings, and theories that we look at.
  - Frankly, I don't trust them. Do I really want to invent more?
  - Publication process reminds me a little of ...
  - Our paper happens to be correct about the predictive variables *as of 2021*. Not all the published papers *still* are.

Problem 2: Pushing the Research Frontier With Technique

- Mike C. will probably state that any student could have written this paper.

Problem 2: Pushing the Research Frontier With Technique

- Mike C. will probably state that any student could have written this paper.
- ...and that's correct, too.

Problem 2: Pushing the Research Frontier With Technique

- Mike C. will probably state that any student could have written this paper.
- ...and that's correct, too.
  - *Should we believe these 1,000 pages of top-1A journal space? Should we believe the next ones? Is large stock-prediction a viable and credible research direction?*
  - *Are we interested in new variables and techniques, or in understanding the behavior of financial markets?*

Problem 2: Pushing the Research Frontier With Technique

- Mike C. will probably state that any student could have written this paper.
- ...and that's correct, too.
  - *Should we believe these 1,000 pages of top-1A journal space? Should we believe the next ones? Is large stock-prediction a viable and credible research direction?*
  - *Are we interested in new variables and techniques, or in understanding the behavior of financial markets?*
  - *(Probably both interesting.)*

Implementation Challenges

- We (GWZ) have about 40 pages to summarize 25 papers and 30 variables.
- This limits what we can do. It has to be systematic and the same. We are not writing a book where we can dedicate one paper each to one paper each.
- Also limits what my presentation can do. For more details, see original and our paper.
  https://dx.doi.org/10.2139/ssrn.3929119

Plan of Action

Use authors' variable definition and frequency of prediction to predict equity-premium. Log and plain. 1948-. Small sanity improvements (e.g., use also monthly for stock variables).

1. Does X predict IS in author's sample?

   not multivariate regressions, etc. we need comp perf.
2. Does X predict IS to 2020?
3. Does X predict OOS to 2020?
4. Compare investment performance of strategies based on X to unconditional buy-and-hold performance.
   - base tilt more or less towards equities?
   - $\pm$ \$1 or Z weight towards signal?
   - Look at non-risk-neutral performance, too.

Categories

- 9 Macroeconomic Variables, often annual.
- 4+1 Sentiment Variables.
- 2 Variance Variables.
- 7 X-sect stock variables.
- 3 other stock variables.
- 1 commodity variable.

(mea poor memory: papers named by author abbreviations and via mnemonic variables, soon to be described.)

Tally I: 24 out of 26

We can confirm authors' IS predictive performance in univariate regressions in their samples, usually with their own frequencies.

2 papers failed: (AMS (sbdlev) and CGMS (skew)) are not reproducible to us.

Tally II: 20 out of 26

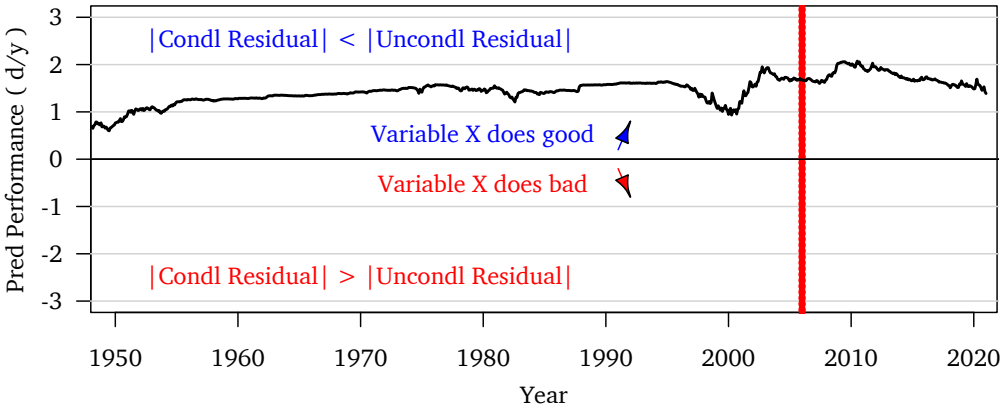
No sign reversal in prediction (or statistically significant decline to irrelevance) in first half vs. second half of author's own sample.

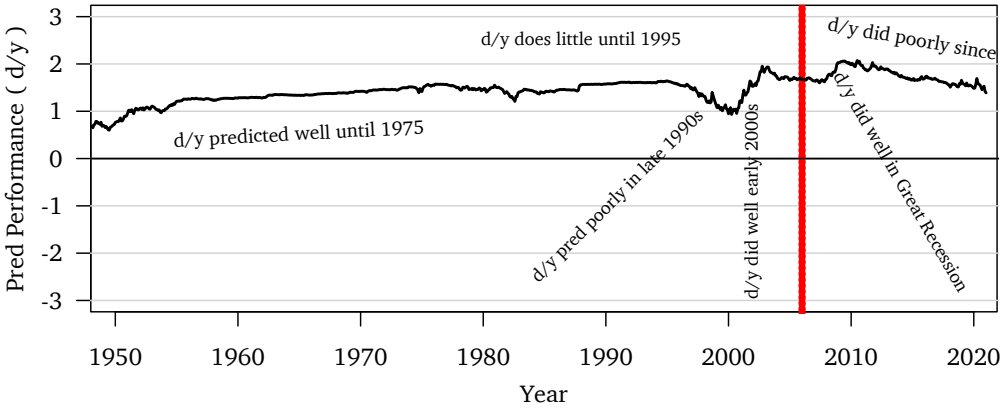
6 papers failed: (AMS (sbdlev), BPS (impvar), BY (govik), CGMS (skew), CP (ogap), KP (fbm), PST (house), Y (disag)) had significant (internal) unstable halves.

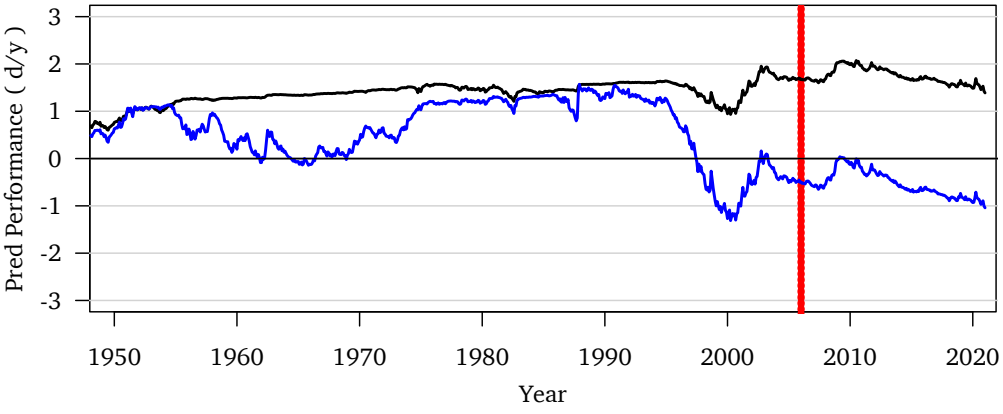
Mostly graphical presentation, one plot per paper.

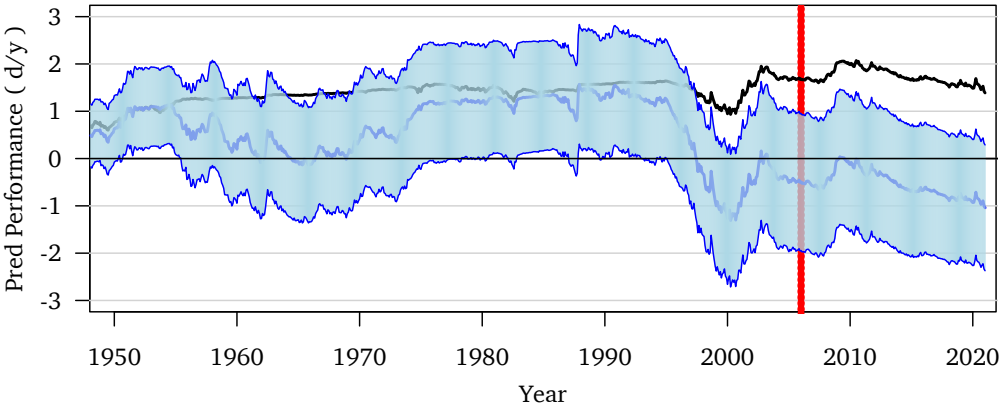
Illustration next.

Dividend Yield




Pred Performance (d/y)
3
2
1
0
-1
-2
-3
1950
1960
1970
1980
1990
2000
2010
2020
Year
d/y predicted well until 1975
d/y does little until 1995
d/y pred poorly in late 1990s
d/y did well early 2000s
d/y did poorly since
d/y did well in Great Recession

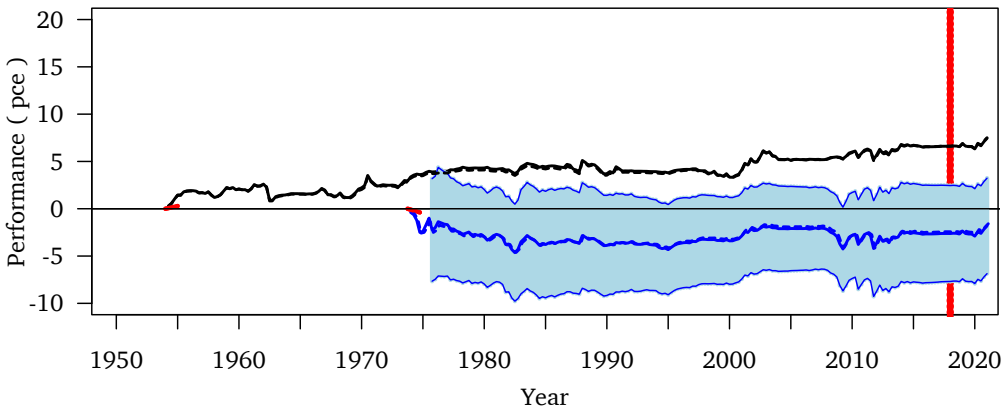

Pred Performance (d/y)
3
2
1
0
-1
-2
-3
1950
1960
1970
1980
1990
2000
2010
2020
Year


Pred Performance (d/y)
3
2
1
0
-1
-2
-3
1950
1960
1970
1980
1990
2000
2010
2020
Year

AMP (pce)

Atanasov, Møller, Priestley (JF 2020): *Consumption Fluctuations and Expected Returns*.

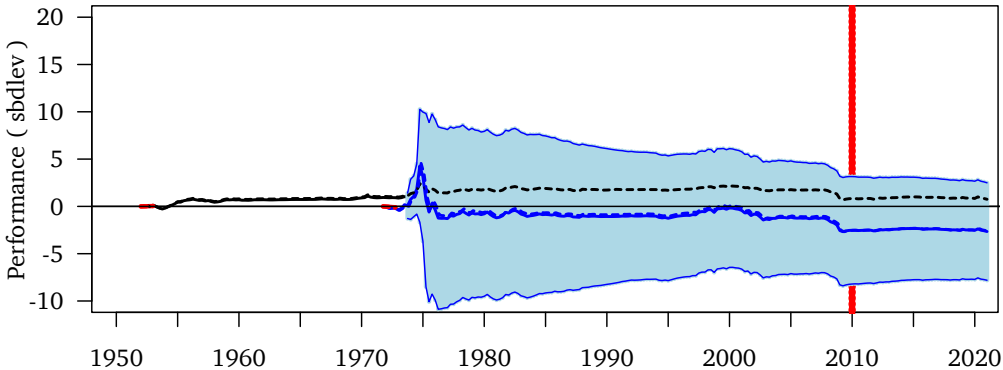
aggregate consumption to its trend, 1953:1-2020:4.



AMS (sbdlev)

Adrian, Mönch, Shin (FRBNY 2010): *Financial intermediation, asset prices, and macroeconomic dynamics*.

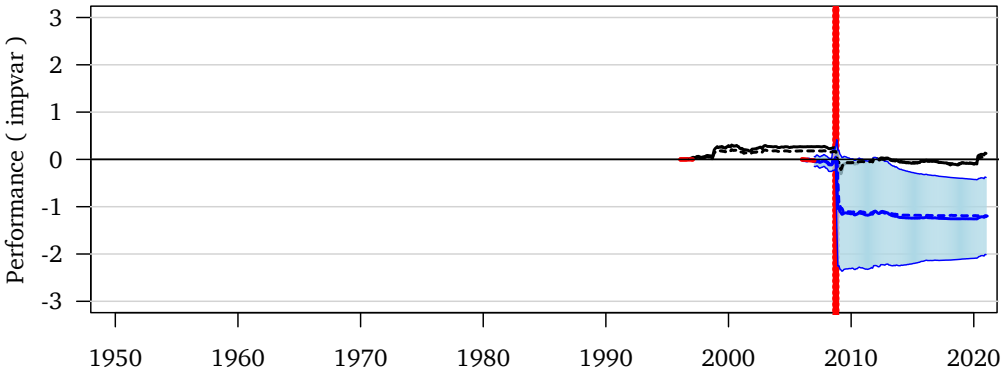
growth rate of security broker-dealer leverage (our own version to avoid division by zero), 1951:4-2020:4.



BPS (impvar)

Bakshi, Panayotov, Skoulakis (JFE 2011): *Improving the predictability of real economic activity and asset returns with forward variances inferred from option portfolios*.

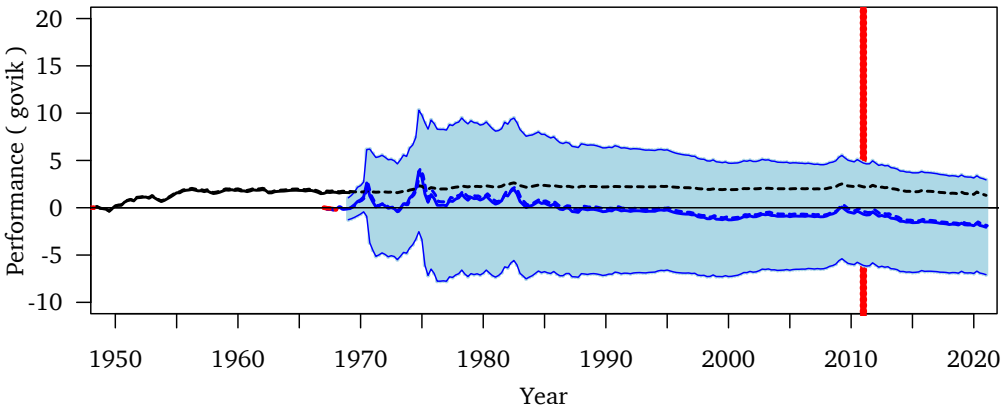
forward implied variances, 1996:01-2020:12.



BY (govik)

Belo and Yu (JME 2013): *Household & government investment and the stock market*.

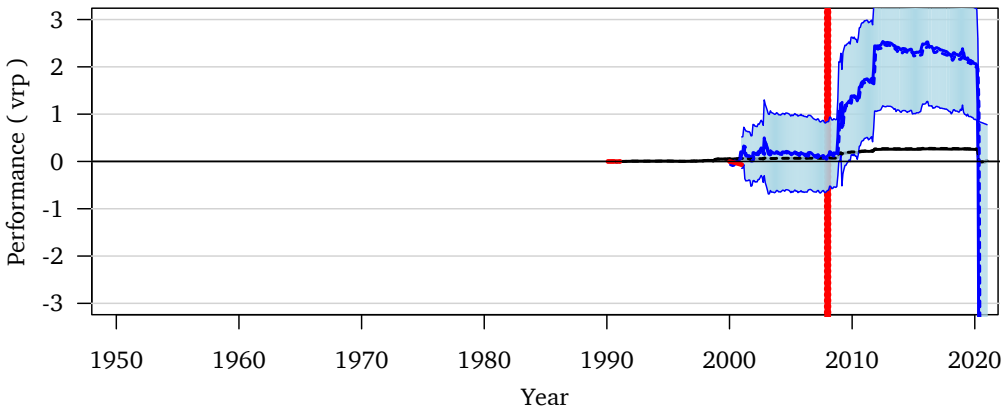
public-sector investment, 1947:1-2020:4.



BTZ (vrp)

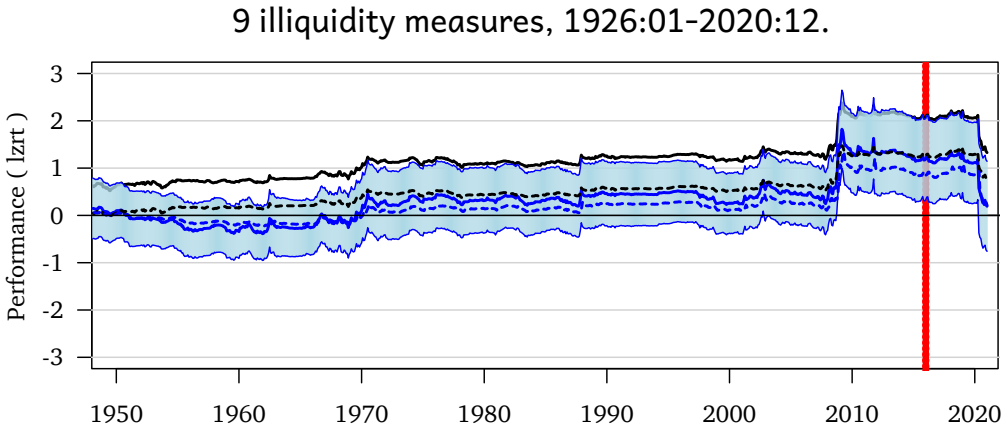
Bollerslev, Tauchen, Zhou (RFS 2009): *Expected Stock Returns and Variance Risk Premia*.

variance risk premium, 1990:01–2020:12.



CEP (lzrt)

Chen, Eaton, Paye (JFE 2018): *Micro(structure) before macro? The predictive power of aggregate illiquidity for stock returns and economic activity.*



CGMS (skew)

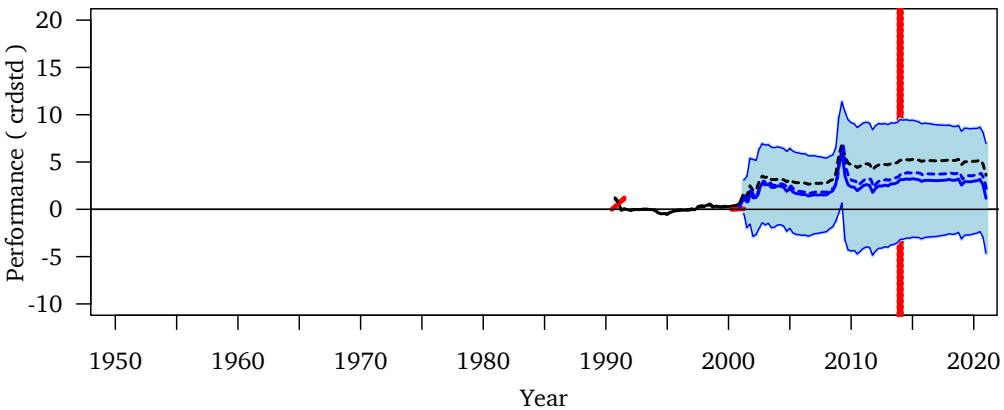
Colacito, Ghysels, Meng, Siwasarit (RFS 2016): *Skewness in Expected Macro Fundamentals and the Predictability of Equity Returns: Evidence and Theory*.

skewness of GDP growth forecasts, 1951:2-2019:2.

CGP (crdstd)

Chava, Gallmeyer, Park (JME 2015): *Credit conditions and stock return predictability*.

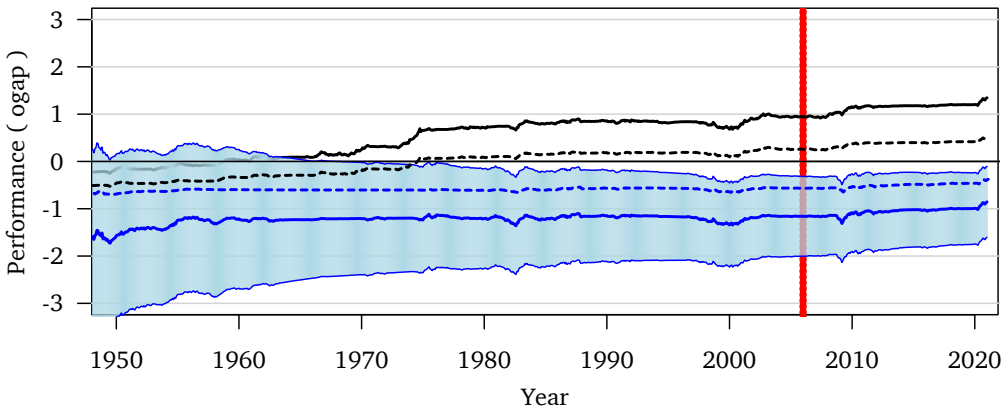
loan officer credit standards, 1990:2–2020:4.



CP (ogap)

Cooper and Priestley (RFS 2009): *Time-Varying Risk Premiums and the Output Gap*.

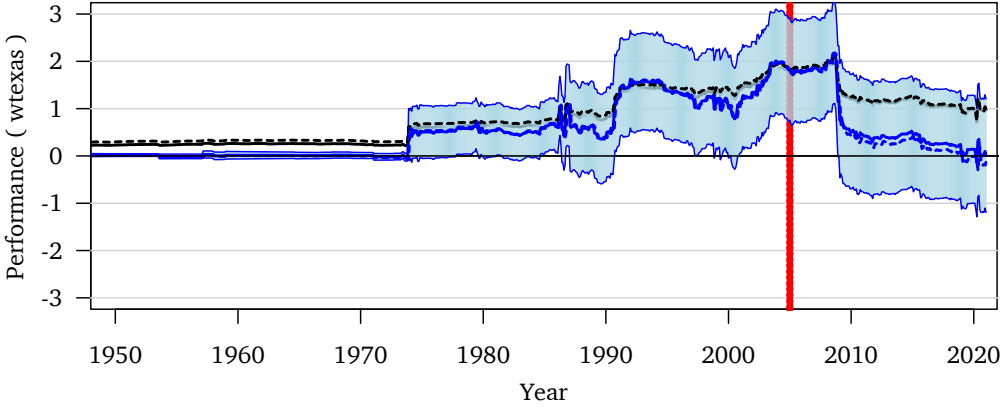
output gap of industrial production, 1926:01-2020:12.



DJM (wtexas)

Driesprong, Jacobsen, Maat (JFE 2008): *Striking oil: Another puzzle?*.

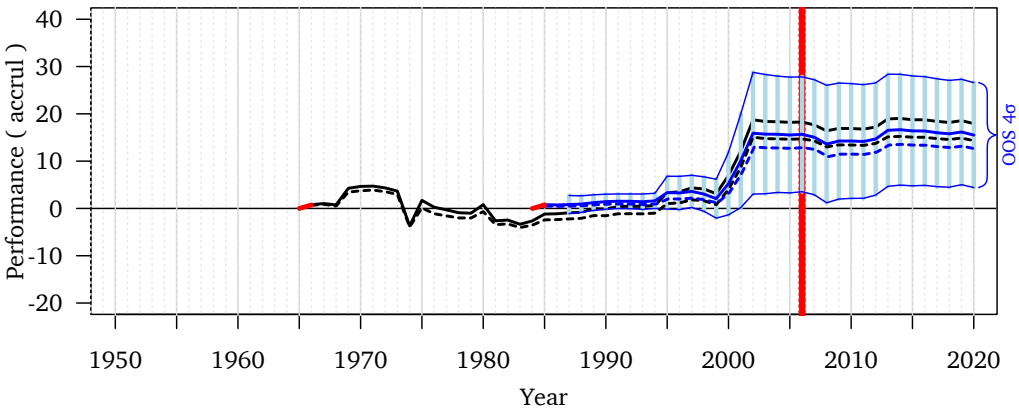
oil price changes, 1926:01–2020:12.



HHT (accrul)

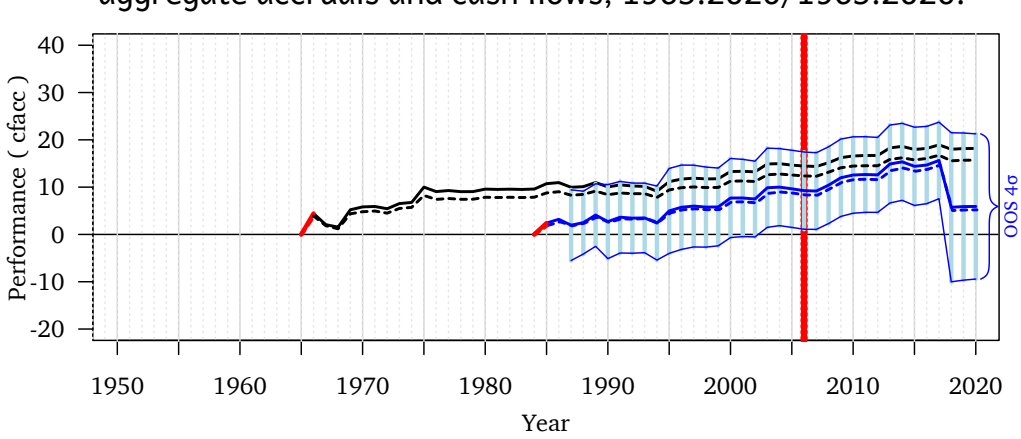
Hirshleifer, Hou, Teoh (JFE 2008): *Accruals, cash flows, and aggregate stock returns*.

aggregate accruals and cash flows, 1965:2020/1965:2020.



HHT (cfacc)

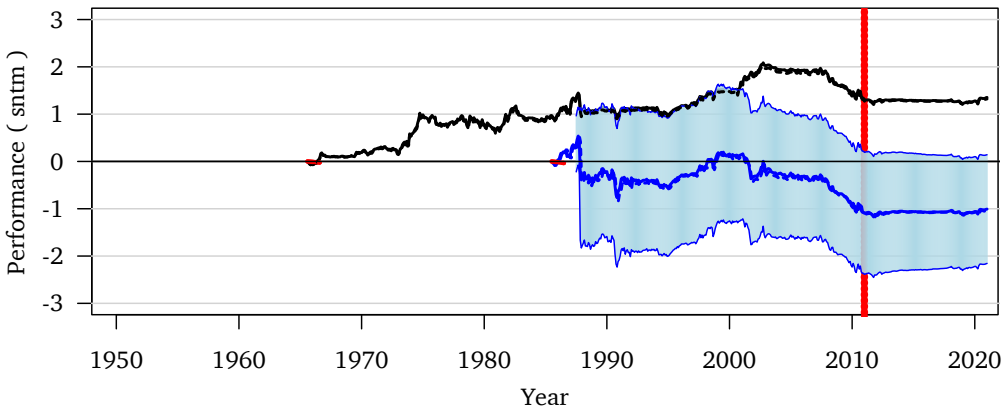
Hirshleifer, Hou, Teoh (JFE 2008): *Accruals, cash flows, and aggregate stock returns*.



HJTZ (sntm)

Huang, Jiang, Tu, Zhou (RFS 2015): *Investor Sentiment Aligned: A Powerful Predictor of Stock Returns*.

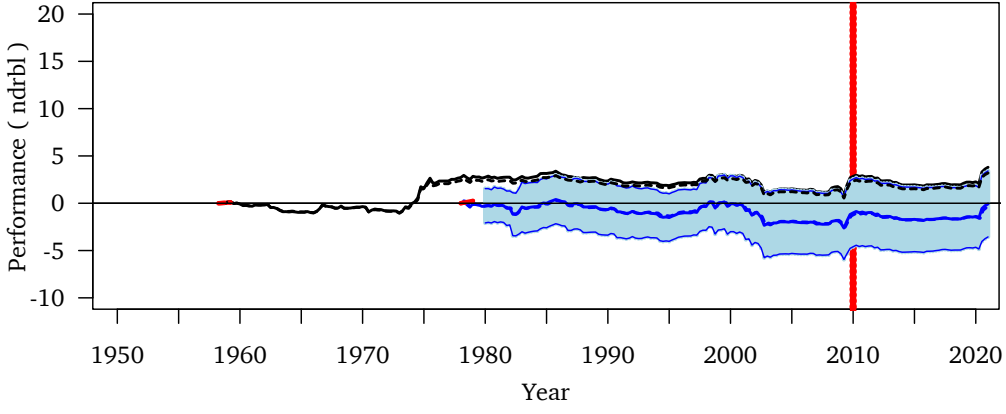
optimized investor sentiment index, 1965:07-2018:12.



JT (ndrbl)

Jones and Tuzel (RFS 2013): *New Orders and Asset Prices*.

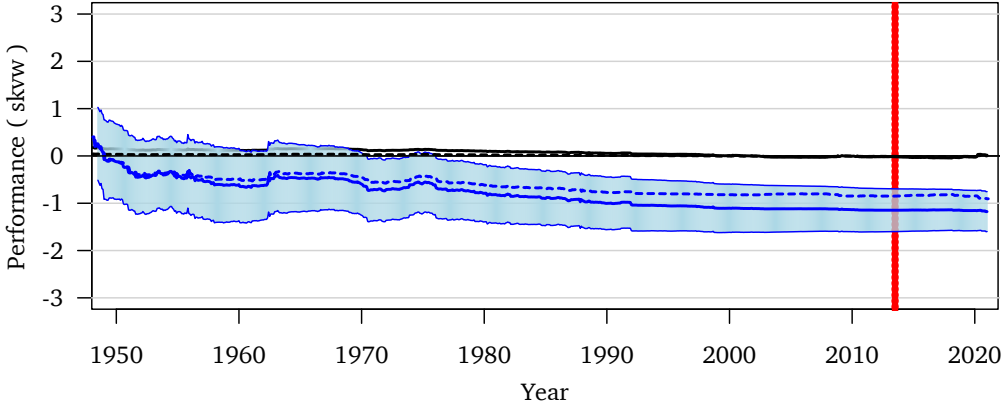
new orders to shipments of durable goods, 1958:02-2020:12.



JZZ (skvw)

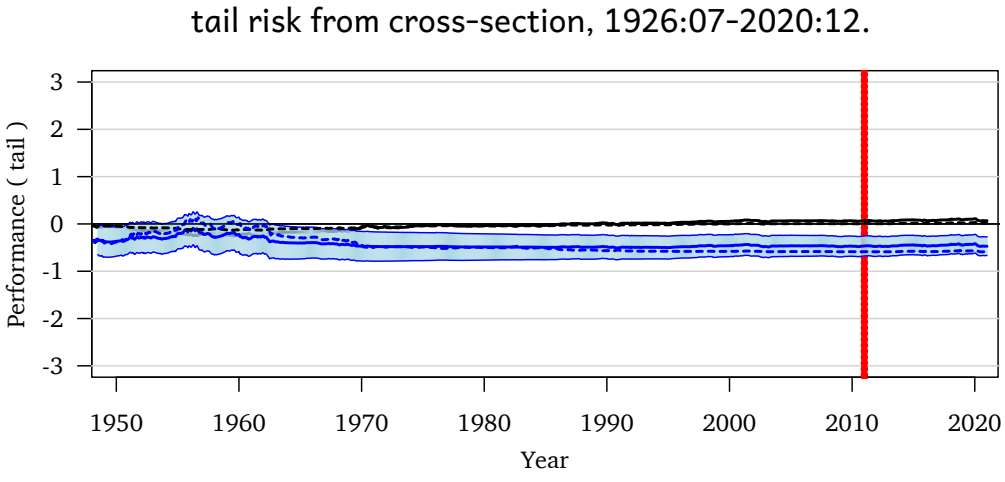
Jondeau, Zhang, Zhu (JFE 2019): *Average Skewness Matters*.

average stock skewness, 1926:07-2020:12.



KJ (tail)

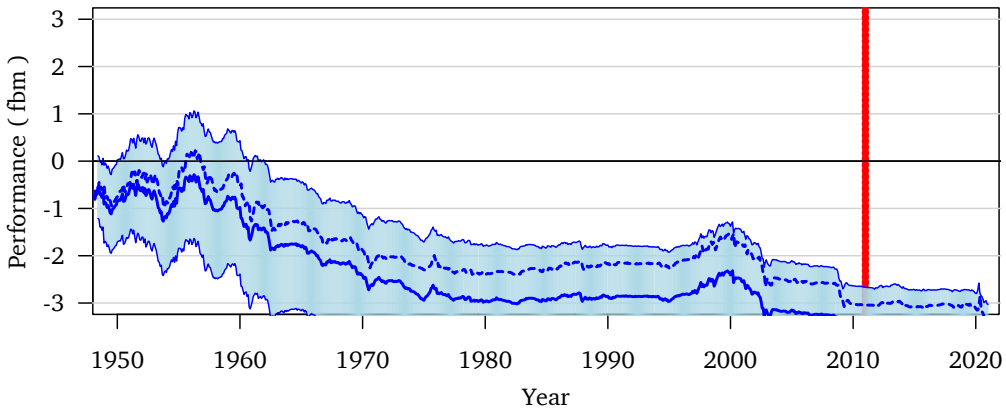
Kelly and Jiang (RFS 2014): *Tail Risk and Asset Prices*.



KP (fbm)

Kelly and Pruitt (JF 2013): *Market Expectations in the Cross-Section of Present Values*.

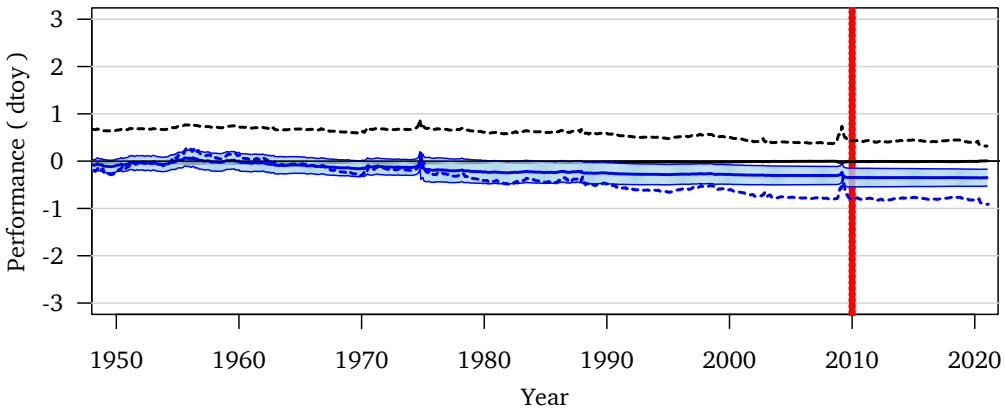
single factor from B/M cross-section, 1926:06-2020:12.



LY (dtoy)

Li and Yu (JFE 2012): *Investor attention, psychological anchors, and stock return predictability*.

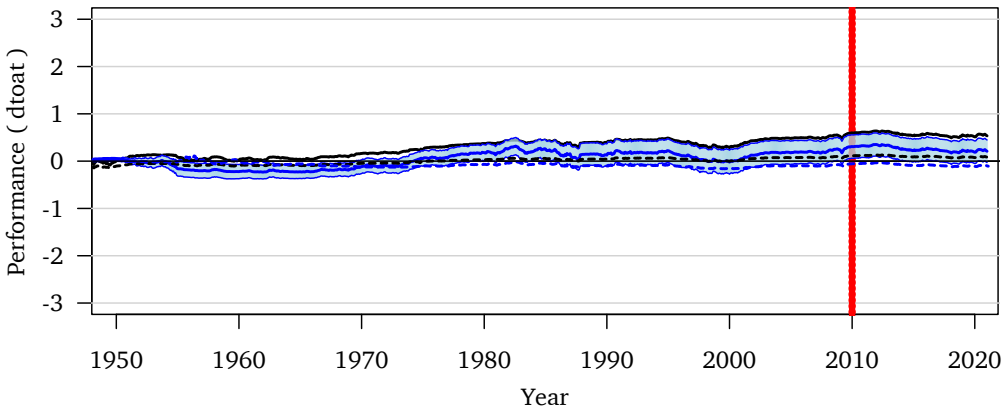
nearness to Dow 52-week high, 1926:01-2020:12/1926:01-2020:12.



LY (dtoat)

Li and Yu (JFE 2012): *Investor attention, psychological anchors, and stock return predictability*.

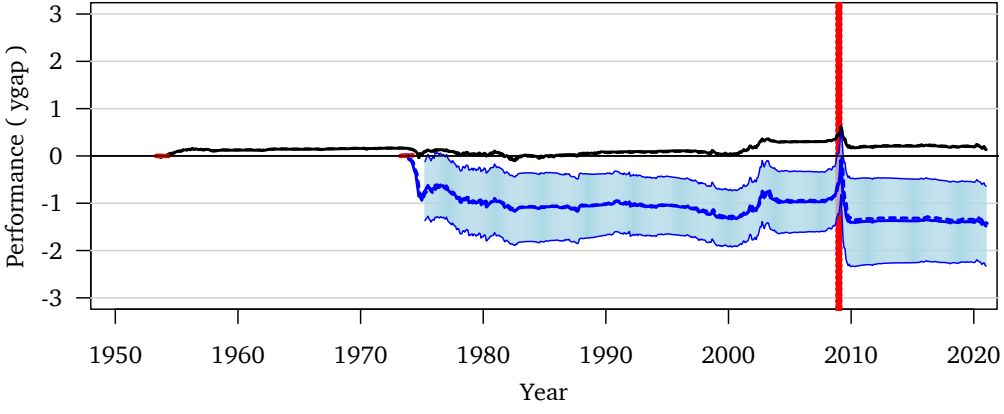
nearness to Dow 52-week high, 1926:01-2020:12/1926:01-2020:12.



Maio$_{(13)}$ (ygap)

Maio (RF 2013): *The Fed Model and the Predictability of Stock Returns*.

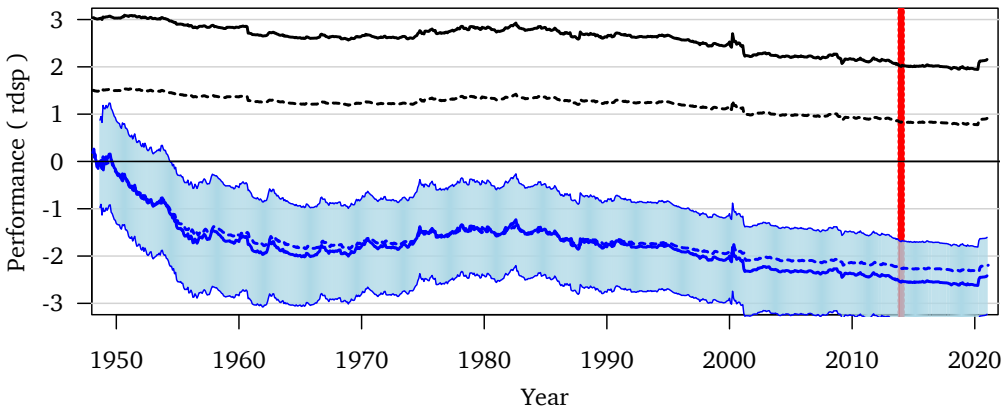
stock-bond yield gap, 1953:04-2020:12.



$\text{Maio}_{(16)}$ (rdsp)

Maio (JFM 2016): *Cross-sectional return dispersion and the equity premium*.

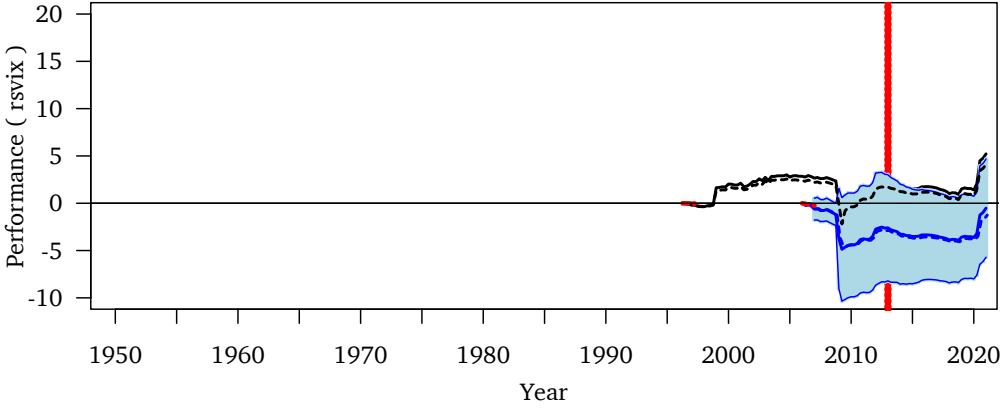
stock-return dispersion, 1926:09-2020:12.



Mrtn (rsvix)

MA (QJE 2017): *Expected Return on the market.*

scaled risk-neutral vix, 1996:01-2020:12.



Huh?

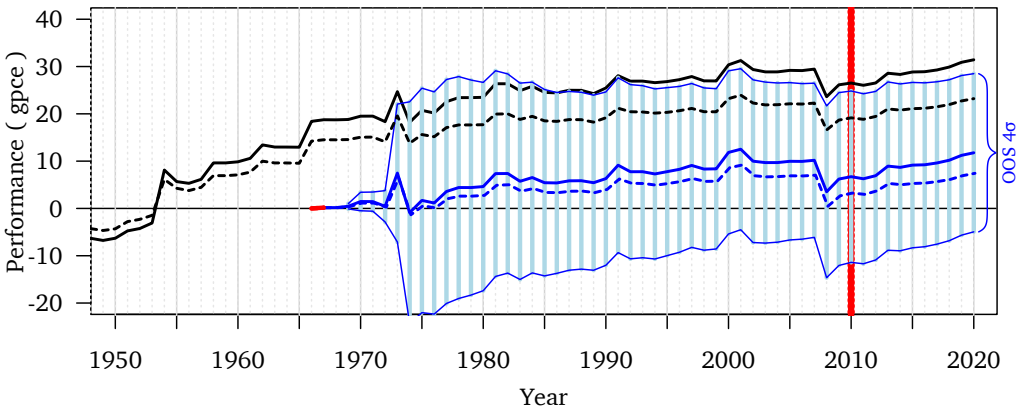
- 3-mo fails, 12-mo fails; 6-mo works *a* little, but only March/Sep ends.
- Performs very poorly — cannot outperform, but fails test whether it significantly underperforms.

  (Very large standard errors.)

MR (gpce)

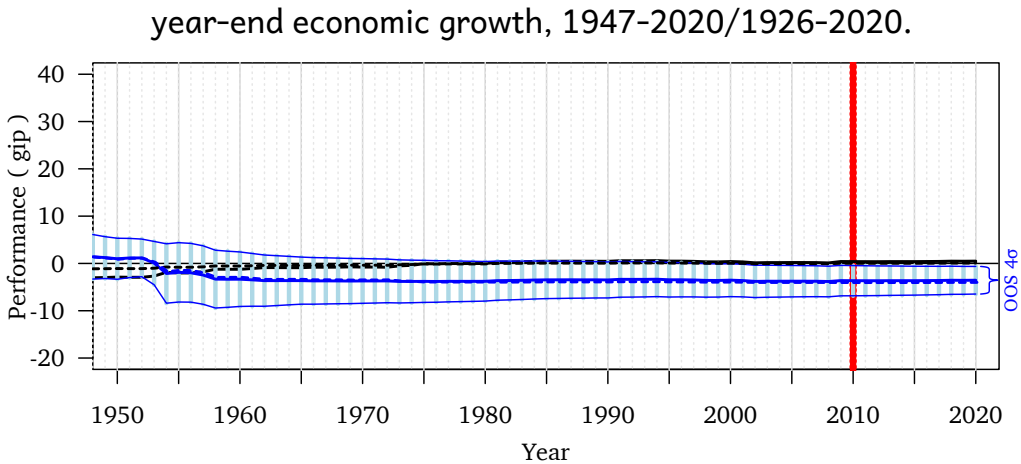
Møller and Rangvid (JFE 2015): *End-of-the-year economic growth and time-varying expected returns*.

year-end economic growth, 1947-2020/1926-2020.



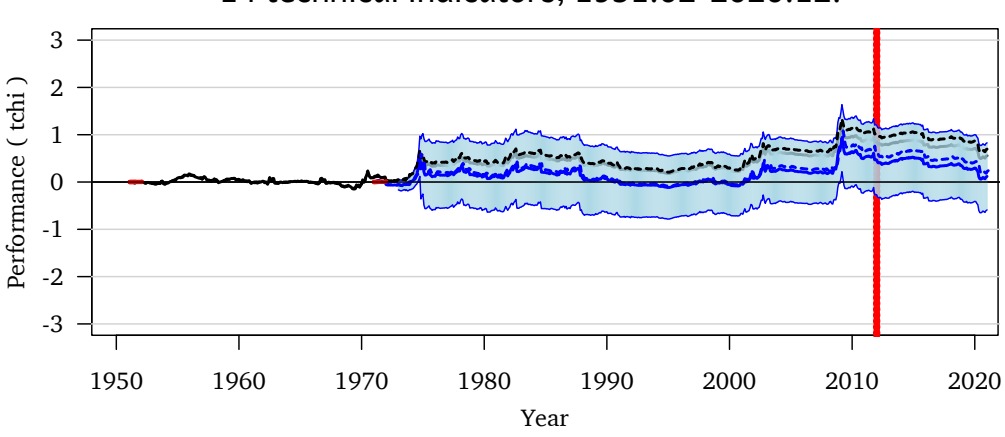
MR (gip)

Møller and Rangvid (JFE 2015): *End-of-the-year economic growth and time-varying expected returns*.



NRTZ (tchi)

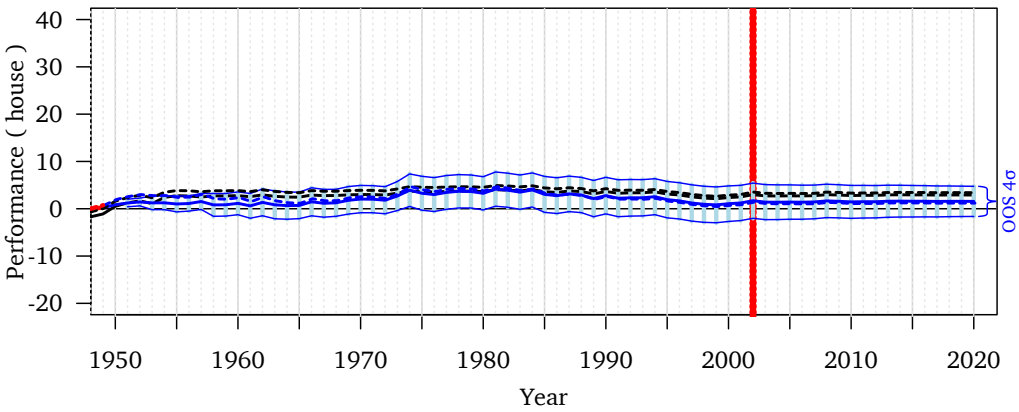
Neely, Rapach, Tu, Zhou (MS 2014): *Forecasting the Equity Risk Premium: The Role of Technical Indicators*.



PST (house)

Piazzesi, Schneider, Tuzel (JFE 2007): *Housing, consumption, and asset pricing*.

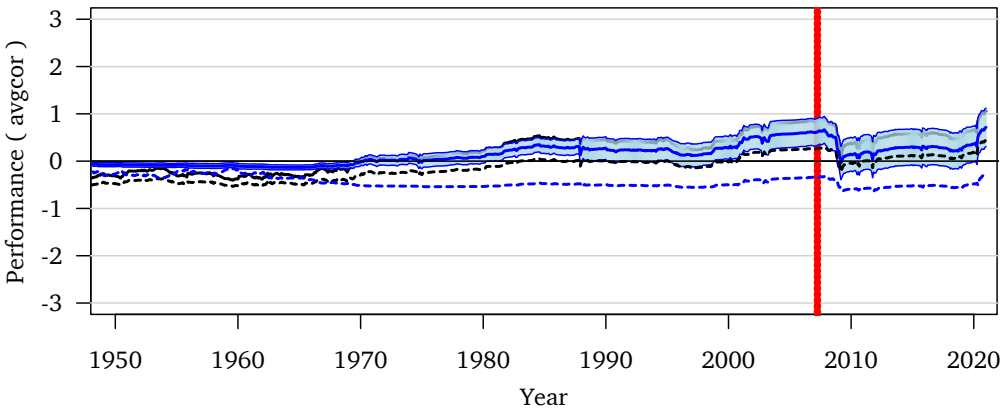
share of housing in consumption, 1929-2020.



PW (avgcor)

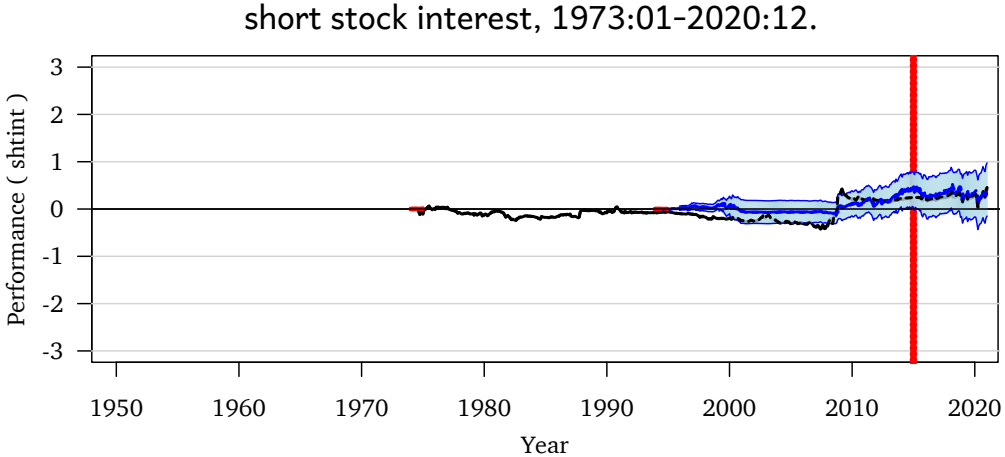
Pollett and Wilson (JFE 2010): *Average correlation and stock market returns*.

average correlation of daily stock returns, 1926:03-2020:12.



RRZ (shtint)

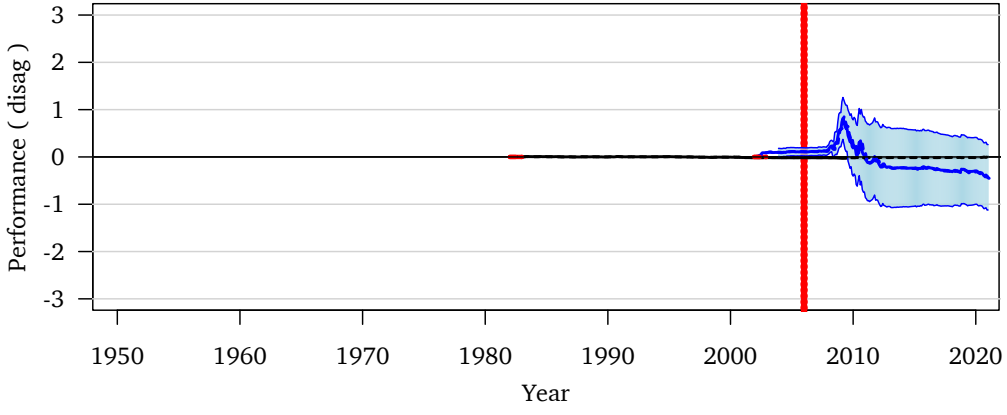
Rapach, Ringgenberg, Zhou (JFE 2016): *Short interest and aggregate stock returns*.



Y (disag)

Yu (JFE 2011): *Disagreement and return predictability of stock portfolios*.

analyst forecast disagreements, 1981:12-2020:12.



Tally III: ≈10 out of 26

OOS performance described and confirmed in author's own sample:

Good: BTZ (vrp), CEP (lzrt), CGP (crdstd), DJM (wtexas), HHT (accrul), HHT (cfacc), HJTZ (sntm), MR (gpce), NRTZ (tchi), RRZ (shtint).

Another 10 variables had discrepancies between our OOS analysis and the authors' (some rather creative).

For 8 variables, the original paper had no OOS predictions.

Tally IV (-2020): 3 out of 17 Monthly Freq

IS Performance still by 2020.

HJTZ (sntm), JT (ndrbl), KP (fbm).

Tally V (-2020): 0 out of 17 Monthly Freq

IS Performance still by 2020.

and

OOS Performance still by 2020.

Tally VI (-2020): Lower-Frequency

4/6 Q only IS by 2020.
1/6 Q both IS and OOS by 2020.
(Plus Cochrane i/k from original variables now.)
(crdstd and i/k would have helped mildly risk-averse investor.)

3/6 A IS by 2020 and OOS.
2/6 also on Jun annual starts. see next
(None would have helped mildly risk-averse investor.)

Best Predictors

- MR (Møller and Rangvid), gpce, growth rate of personal consumption expenditures.
- HHT (Hirshleifer, Hou, Teoh), accrul, accruals.

  really worked only predicting the 2001-2 Tech bubble collapse. indescript otherwise.
- Cochrane's i/k.

Investment Performance

Buy if > Past Medium. Sell Otherwise.

Variable (V)			Conditional R (V)			#Obs		Unconditional R (U)			Δ V – U	
Fq	Ppr	Var	Long	Short	L–S	Bull	Bear	L (Eq)	S (TB)	L–S	Mean	SR
M	BPS	impvar	2.6	9.2	–6.6	40	140	10.7	1.1	9.6	–16.2	–0.22
M	BTZ	vrp	6.7	2.6	4.1	86	166	7.7	1.6	6.2	–2.1	–0.03
M	CEP	lzrt	6.5	9.0	–2.5	263	637					

	Variable (V)		Conditional R (V)			#Obs		Unconditional R (U)			Δ V − U	
Eq	Ppr	Var	Long	Short	L–S	Bull	Bear	L (Eq)	S (TB)	L–S	Mean	SR
M	BPS	impvar	2.6	9.2	−6.6	40	140	10.7	1.1	9.6	−16.2	−0.22
M	BTZ	vrp	6.7	2.6	4.1	86	166	7.7	1.6	6.2	−2.1	−0.03
M	CEP	lzrt	6.5	9.0	−2.5	263	637	11.7	3.8	7.9	−10.4	−0.12
M	CP	ogap	10.4	5.1	5.3	685	215	11.7	3.8	7.9	−2.6	−0.05
M	DJM	wtexas	10.6	4.9	5.8	597	303	11.7	3.8	7.9	−2.1	−0.04
M	HJTZ	sntm	9.5	5.6	4.0	309	117	12.0	3.1	8.9	−4.9	−0.08
M	JT	ndrbl	9.8	7.2	2.6	358	157	12.8	4.3	8.5	−5.9	−0.12
M	JZZ	skvw	5.6	9.9	−4.3	399	495	11.7	3.9	7.8	−12.2	−0.16
M	KJ	tail	8.9	6.6	2.3	531	363	11.7	3.9	7.8	−5.5	−0.09
M	KP	fbm	8.6	6.9	1.7	478	417	11.6	3.9	7.8	−6.1	−0.08
M	LY_{dtoat}	dtoat	5.8	9.7	−3.9	121	779	11.7	3.8	7.9	−11.8	−0.14
M	LY_{dtoy}	dtoy	7.5	8.0	−0.5	364	536	11.7	3.8	7.9	−8.4	−0.12
M	$Maio_{(16)}$	rdsp	4.1	11.6	−7.5	179	713	11.8	3.9	8.0	−15.4	−0.18
M	$Maio_{(13)}$	ygap	6.7	9.4	−2.7	203	370	11.6	4.4	7.2	−9.9	−0.12
M	NRTZ	tchi	9.3	6.7	2.6	333	267	11.6	4.4	7.2	−4.6	−0.06
M	PW	avgcor	9.4	6.1	3.3	412	486	11.7	3.8	7.9	−4.6	−0.07
M	RRZ	shtint	10.2	3.1	7.2	219	106	11.0	2.3	8.7	−1.6	−0.03
M	Y	disag	9.0	1.8	7.2	215	14	9.5	1.2	8.3	−1.1	−0.03
M	BMRR	ntis	10.2	5.7	4.6	560	329	12.0	3.9	8.1	−3.6	−0.06
M	CS_{de}	d/e	4.8	10.7	−5.9	125	775	11.7	3.8	7.9	−13.7	−0.15
M	CS_{dp}	d/p	6.2	9.3	−3.0	179	721	11.7	3.8	7.9	−10.9	−0.12
M	CS_{dy}	d/y	6.4	9.1	−2.7	177	722	11.6	3.8	7.8	−10.5	−0.12
M	CS_{ep}	e/p	7.6	8.0	−0.4	298	602	11.7	3.8	7.9	−8.2	−0.10
M	Cmpl	tbl	6.0	9.6	−3.6	245	655	11.7	3.8	7.9	−11.5	−0.13
M	FF_{dfr}	dfr	9.0	6.6	2.4	469	431	11.7	3.8	7.9	−5.4	−0.07
M	FF_{dfy}	dfy	7.6	7.9	−0.4	382	518	11.7	3.8	7.9	−8.2	−0.12
M	FF_{ltr}	ltr	9.7	5.8	3.9	441	459	11.7	3.8	7.9	−4.0	−0.06
M	FF_{lty}	lty	6.7	8.8	−2.1	224	676	11.7	3.8	7.9	−10.0	−0.12
M	FF_{tms}	tms	8.4	7.2	1.2	403	497	11.7	3.8	7.9	−6.7	−0.09
M	FS	infl	10.4	5.1	5.3	411	489	11.7	3.8	7.9	−2.6	−0.04
M	G	svar	7.4	8.1	−0.7	373	527	11.7	3.8	7.9	−8.6	−0.14
M	KS	b/m	5.9	9.6	−3.6	265	635	11.7	3.8	7.9	−11.5	−0.14
Q	AMP	pce	10.7	5.8	5.0	134	55	12.1	4.4	7.6	−2.6	−0.06
Q	AMS	sbdlev	9.1	7.3	1.8	111	86	11.9	4.5	7.4	−5.6	−0.12
Q	BY	govik	4.5	11.6	−7.1	12	204	11.6	4.5	7.1	−14.2	−0.23
Q	CGP	crdstd	8.1	1.4	6.7	46	37	8.0	1.5	6.5	0.2	0.00
Q	Mrtn	rsvix	7.8	4.3	3.4	19	41	11.0	1.1	9.9	−6.5	−0.15
Q	Crn	i/k	9.6	6.6	3.0	88	128	11.6	4.5	7.1	−4.1	−0.08
Q	LL	cay	7.4	8.9	−1.6	60	136	11.9	4.5	7.4	−9.0	−0.16
D	MR	gip	11.8	4.7	7.1	52	23	12.5	3.9	8.6	−1.4	−0.08
D	MR	gpce	10.8	5.8	5.0	31	23	12.0	4.7	7.4	−2.4	−0.12
D	PST	house	7.0	10.2	−3.2	23	49	13.0	4.1	8.9	−12.2	−0.40
D	BW	eqis	9.8	7.0	2.8	43	31	12.8	4.0	8.8	−6.0	−0.23
J	HHT_{acc}	accrul	7.5	7.6	−0.1	15	20	11.9	3.2	8.7	−8.8	−0.39
J	HHT_{cf}	cfacc	10.9	4.2	6.7	23	12	11.9	3.2	8.7	−2.0	−0.10

Tally VII: 0 out of Untilted $1-Unscaled I

20/45 even lost money in absolute terms.

"Perfect" Track Record: 45/45 underperformed unconditional buy-and-hold.

9/45 outperformed UC SR when signal is z-scaled
9/45 outperformed UC SR when tilted towards equity.
14/45 when both.

Conclusion

- “Strong support of theory”

Conclusion

- “Strong support of theory”
- Occasional absurd individual claims; absurd *collective* claims.

Conclusion

- "Strong support of theory"
- Occasional absurd individual claims; absurd *collective* claims.
- **Hypothesis and Opinion:** We cannot reliably predict stock returns *forward-looking* for top-1000 stocks. (Nagel-Martin?)

Conclusion

- "Strong support of theory"
- Occasional absurd individual claims; absurd *collective* claims.
- **Hypothesis and Opinion:** We cannot reliably predict stock returns *forward-looking* for top-1000 stocks. (Nagel-Martin?)
- I doubt the intellectual integrity of our *collective* research enterprise here.
  - Our collective bias towards non-mundane research weeds out skeptics.
  - Why are we still publishing papers explaining book-to-market or stock momentum as risk compensation? What is the HML premium today for risk-bearing? What is UMD's?

(I have little doubt about the integrity of individual authors, though.)